Die Paradoxe Rationalität Einfacher Heuristiken

Inaugural - Dissertation

zur
Erlangung des Grades eines Doktors der Philosophie

in der

Fakultät für Psychologie

der

RUHR - UNIVERSITÄT BOCHUM

vorgelegt von:

Jan K. Woike

Gedruckt mit Genehmigung der Fakultät für Psychologie der
RUHR-UNIVERSITÄT BOCHUM

Referent: Prof. Dr. Heinrich Wottawa
Korreferentin: Prof Dr. Laura Martignon

Datum der mündlichen Prüfung: 04.02.2008

[...] we often need to imagine ourselves in conversation with those who helped make us what we are in order to help us decide whether we have gone, or are going, in the right direction. I see such conversations, and the stories of progress or regress that result from imagining them, as giving substance and concreteness to the otherwise thin and useless notion of „rationality.“ (Rorty, 2000)

● ● ●

And when we finally get to Cinderella's castle, itself a Disney malappropriation of many Eastern European myths (the bloody feet and poked-out eyes have been repressed), I am surprised it is nothing more than an archway which serves as a rite of passage: Main Street before it, the rest of Magic Kingdom beyond it. This is literally metonymy of colossal proportions, for the castle is meant to be viewed from a distance, a directional icon for entry and exit, never actually entered or thought about. Part of the whole, the hollow archway fills and empties, fills and empties, while the gigantic castle itself lies largely vacant. (van Wert, 1995-1996)

Danksagung

Danksagungen sind wie die Dankesreden bei den *Academy Awards*, in gewisser Weise irrelevant für die Rezeption der vorgestellten Werke, aber oft der eigentliche Grund warum der Fernseher überhaupt eingeschaltet und manches Buch überhaupt aufgeschlagen wird[1]. Unabhängig davon ist es mir ein Anliegen, mich bei den vielen Menschen zu bedanken, die mich in den letzten Jahren bei der Abfassung dieser Arbeit unterstützt haben. Da fast alle erwähnten Personen inzwischen einen oder mehrere Titel aufweisen[2], möchte ich allein zugunsten der Lesbarkeit und nicht als Zeichen fehlenden Respekts außerhalb der Fußnote auf die Nennung selbiger Namensvorsätze verzichten.

Ich möchte zunächst meinem Doktorvater Heinrich Wottawa für die Möglichkeit danken, mich einem Thema widmen zu können ohne den Phänomenbereich auf eine einzelne Lehrmeinung reduzieren zu müssen. Gleichzeitig danke ich für die erfreulich differenzierte und kritische Einführung in die psychologische Methodologie durch ihn und auch Raphael Diepgen, dem ich für zahlreiche interessante Diskussionen über die psychologische Statistik danke. Diese Lehrveranstaltungen gehören verbunden mit bereichernden Diskussionen zu den wesentlichen Inspirationen und Grundsteinen dieser Arbeit. In meiner Zeit als Mitarbeiter am Bochumer Methodenlehre-Lehrstuhl konnte ich insbesondere auch von der Praxiserfahrung meines Doktorvaters im Bereich der Personaldiagnostik profitieren.

In diesem Kontext möchte ich auch Christian Montel (mittlerweile in Berlin) danken, sowohl für die Bereitstellung von drei Datensätzen als auch für einige anregende Diskussionen. Danken möchte in Bochum auch Jürgen Stränger für viele fruchtbare Reflexionen über die Theorie und Philosophie der Psychologie, sowie Onur Güntürkün, Axel Schölmerich und Rüdiger Hossiep für studentische Hilfskraftstellen, die mir die psychologische Forschung aus verschiedenen Perspektiven näher gebracht haben.

Aus der Zeit an der *Harvard University* sind mir insbesondere die *Brown Bag*-Diskussionen mit Robert Rosenthal und Dan Rubin in positiver Erinnerung geblieben, in denen auch einige frühe Ideen dieser Arbeit geformt werden konnten. Richard Hackman danke ich für die Erfahrung, dass akademische Diskussionen begeistert, engagiert und direkt geführt werden können. Gleichzeitig riskierte ich in dieser Zeit als Nichtschwimmer den Sprung ins kalte Wasser der Philosophie, der dank der Unterstützung durch kompetente „Schwimmlehrer“ wie Hilary Putnam und den inzwischen leider verstorbenen Robert Nozick zu einer anhaltenden Begeisterung für diese Disziplin geführt hat. Vor allem an die Graduiertenseminarstunden zur *Philosophy of Mind* in Hilary Putnams Büro denke ich mit großer Faszination und menschlicher Hochachtung zurück, erst in den letzten Jahren haben sich Facetten seiner klaren Argumentation und seines subtilen Humors für mich eröffnet.

Die Explorationen in die Philosophie wurden unterstützt durch eine Einführung in die Logik und Metamathematik durch Warren Goldfarb, von der ich in Bochum in einer Hilfskraftstelle am Sprachphilosophie-Lehrstuhl profitieren konnte. Hier danke ich Ulrich Pardey für eine stete kritische Diskussionsbereitschaft und lebendige Seminare.

Ein großer Teil dieser Arbeit ist in den letzten beiden Jahren in Berlin entstanden, wo es für mich eine große Freude war, an den Forschungstätigkeiten der Gruppe *Adaptive Behavior and Cognition* am Max-Planck-Institut für Bildungsforschung teilnehmen zu können. Hier möchte ich vor allem Gerd Gigerenzer aber auch allen Mitgliedern der ABC-Gruppe für die Bereitstellung eines fantastischen Arbeitsumfeldes danken und faszinierende Unterhaltungen während der täglichen gemeinschaftlichen *Coffe and Tea Break* und die Teilnahme an einem sehr anregenden *Retreat*. Diese Zeit hat auch im Gesicht dieser Arbeit deutliche Spuren hinterlassen.

Insbesondere gilt mein Dank Laura Martignon und Konstantinos Katsikopoulos für eine fachlich aber vor allem persönlich sehr bereichernde Zusammenarbeit. Es sind die seltenen Momente, in denen ich die Begeisterung für ein Themengebiet in der Zusammenarbeit gespiegelt finden konnte, die zu einer entscheidenden Motivationsquelle werden konnten. Laura Martignon danke ich für ihre wunderbare Unterstützung, für zahlreiche persönliche und reichhaltige Gespräche über Bäume, die Mathematik, Gott und die Welt sowie ihre Bereitschaft als Korreferentin dieser Arbeit zu fungieren.

Aber auch den übrigen Mitgliedern der ABC-Gruppe und den Gästen des Instituts danke ich für Rückmeldungen, Diskussionsbereitschaft und eine mehr als angenehme Zeit in Berlin, insbesondere Adrien Barton, Arndt Bröder, Juliet Conlin, Uwe Czienskowski, Wolfgang Gaissmaier, Mirta Galesic, Floris Heukelom, Linnea Karlsson, Monika Keller, Julian Marewski, Shabnam Mousavi, Henrik Ollson, Thorten Pachur, Magnus Persson, Jing Qian,

[1]

> Es ist so schwer, den *Anfang* zu finden. Oder besser: Es ist schwer, am Anfang anzufangen. Und nicht zu versuchen, weiter zurückzugehen. (Wittgenstein, 1984b, §471)

[2] Prof., Dr. phil, Dr. med., Dr. rer. nat., Dipl.-Psych., Dipl.-Inf., Dipl.-Biol., Dipl-Ökon., M.A., Dipl.-Ing., Dipl-Math., PhD u.a.

Jörg Rieskamp, Lael Schooler, Nils Straubinger, Bettina von Helversen, Peter Todd, Odette Wegwarth und Ulrich Hoffrage, dem ich zudem auch für die Unterstützung in der Endphase meiner Dissertation in Lausanne danken möchte. Christian Elsner war eine immense technische und vor allem emotionale Unterstützung in der monatelangen Durchführung der Simulationen am Institut.

Dem Cusanuswerk bin ich dankbar für die finanzielle und intellektuelle Unterstützung dieser Arbeit durch ein Graduiertenstipendium. Die Graduiertentagungen haben es mir ermöglicht, Ideen an einer engagierten Zuhörerschaft zu erproben und nebenbei wunderschöne Tage in zumeist merkwürdigen Orten Deutschlands zu verbringen. Dem DAAD danke ich für die Finanzierung einer Kurzzeitdozentur in Arizona.

In einer vierseitigen Danksagung - die zu meinen Lieblingstexten des Genres gehört - schildert Jones (1995) eine zentrale Erfahrung seiner Dissertationszeit:

> At Indiana University, I was fortunate to be supervised by Douglas Hofstadter. Doug has both wide-ranging interests and the ability to think (and write!) clearly about his interests, in a highly original and probing way. Doug was very supportive of my extracurricular activities. When I began spending six hours a day juggling and unicycling rather than telling me to do some work Doug became the faculty advisor for the IU juggling club. (Jones, 1995, S. v)

In gleicher Weise möchte ich die während der Arbeit erfahrene Hilfe nicht auf eine akademische beschränkt sehen. Meinen Eltern danke ich in einer sehr bewegten Zeit für ihre unkonditionale Unterstützung, die mir immens viel bedeutet hat und ohne die dieses Projekt wohl weder angefangen noch abgeschlossen worden wäre. Patricia Kanngießer danke ich für ihre Toleranz hinsichtlich meiner Dissertationspersönlichkeit insbesondere im letzten Jahr und nicht zuletzt die Bereitschaft, sich durch den Text dieser Arbeit zu kämpfen - alles weitere sage ich persönlich. Claus Noppeney danke ich für eine integrale Freundschaft in den letzten „Berliner“ Jahren, für seine Unterstützung und auch für seine inhaltliche und logistische Sorge um meine Arbeit.

Zur Gewährleistung des körperlichen Ausgleichs in einer einseitig gedankenzentrierten Lebensphase danke ich Oliver Brandner, meinem Ninjutsu-Lehrer in Bochum und Jochen Knau, meinem Berliner Aikido-Lehrer. In dieser Zeit habe ich gelernt, dass die Arbeit an Ideen für mich nicht von einer begleitenden körperlichen Lebenspraxis getrennt werden kann. Das Training hat mein Arbeiten und meinen Alltag in vielfältiger und mittlerweile unverzichtbarer Weise bereichert.

Für Unterstützungen in den letzten Jahren in vielfältiger Art und Weise danke ich weiterhin Frederik Ahlgrimm, Sven Arndt, Shana Bell, Pamela Copei, Lars Düesmann, Timo Ehrig, Christian Eisert, Luka Fineisen, Nina Flamm, Patrizia Flores, Susan Gniechwitz, Cordula Grüner, Markus Herbert, Maren Hiltmann, Stephan Huber, Quenby Hughes, Mathias Imöhl, Hilmar Klink, Katharina Kluge, Stephan Lanzinger, Frank Martens, Gunther Meindlschmidt, Arne Moritz, Stefanie Neubeck, Tina Schäfer, Jan Scheitza, Heiko Schwarzburger, Johannes Stein, Halina Tegetmeyer, TJ, Volker, Sebastian Voß, Denise Wagner, Martin Wawro, Julia Woike sowie den Teilnehmern meiner Seminare, Theaterworkshops und Filmprojekte.

Die Textgestaltung dieser Arbeit verdanke ich der Arbeit der Entwickler von LaTeXund seinen Komponenten. Eingesetzt habe ich *mikTeX* und das *TeXnic Center*, sowie hauptsächlich die Erweiterungen *ae*, *amsfonts*, *amsmath*, *apager*, *calc*, *caption*, *color*, *enumitem*, *epigraph*, *eurofont*, *fontenc*, *geometry*, *graphicx*, *hhline*, *ifthen*, *inputenc*, *listings*, *makeidx*, *minitoc*, *multido*, *newalg*, *ngerman*, *pst-3Dplot*, *pst-all*, *pst-gr3d*, *pst-uml*, *pst-xkey*, *quotchap2*, *rotating*, *scrpage2*, *subfigure*, *supertabular*, *thumbpdf*, *titling* und *xypic*.

Während bei den *Academy Awards* and dieser Stelle meine Worte längst von der Themenmusik übertönt würden, sehe ich hier das zweite Seitenende auf mich zukommen und beende die Danksagungen in der Hoffnung, nicht zu viele Menschen vergessen zu haben. Die letzten Jahre waren eine spannende Zeit. Danke.

Lausanne, 04.10.2007
Jan K. Woike

Zusammenfassung

Hierarchische Regellisten für binäre Klassifikationen auf der Basis von Testprofilen werden durch drei unterschiedliche Klassen von Heuristiken konstruiert: durch einen gierigen Algorithmus aus dem Data Mining, durch genetische Algorithmen und Schwarmalgorithmen aus dem Bereich des Künstlichen Lebens und zuletzt durch ein Verfahren, das eine Brücke zwischen diagnostischen Problemstellungen und dem kognitiven Forschungsprogramm der adaptiven Werkzeugkiste spannt: Fast and Frugal Trees sind minimale binäre Entscheidungsbäume, die Informationen ignorieren, Abhängigkeiten zwischen Prädiktoren ignorieren und als lexikografische Klassifikatoren beschrieben werden können. In einer umfangreichen Simulationsstudie erweist sich dieses Verfahren als konkurrenzfähig mit klassischen Verfahren wie CART oder der logistischen Regression. Die Frage der Rationalität simpler Heuristiken wird diskutiert, und es werden Implikationen für die Praxis der psychologischen Diagnostik abgeleitet.

Inhaltsverzeichnis

1 Einführung in Thema und Struktur der Arbeit **1**
1.1 Klassifikation und Regelinduktion . . . 1
1.2 Rationalität und Heuristiken . . . 1
1.3 Paradoxien und Heuristiken . . . 3
1.4 Methodische Innovationen . . . 3
1.5 Die Struktur der Arbeit . . . 4
1.6 Programmierte Softwaremodule . . . 6

2 Ein heuristischer Ansatz zur Regelinduktion **7**
2.1 *Data Mining*: Der Datensatz als Goldmine . . . 7
2.1.1 Data Mining und traditionelle Statistik . . . 7
2.1.2 Heuristiken und Algorithmen . . . 11
2.1.3 Data Mining und Bias . . . 12
2.2 Die Relevanz von Regelinduktion für die psychologische Diagnostik . . . 15
2.2.1 Experten und einfache lineare Modelle . . . 15
2.2.2 Computergestütztes Recruiting und Regelinduktion . . . 20
2.2.3 DIN 33430 und die Form von Entscheidungsregeln . . . 21
2.2.4 Beispiele für Entscheidungsregeln in computergestützten Testverfahren . . . 25
2.2.4.1 Anforderungsprofile und regelgestützte Auswahl mit ELIGO . . . 25
2.2.4.2 Potenzialanalyse mit „pro facts“ . . . 26
2.2.4.3 Internet-Recruiting mit PERLS . . . 26
2.3 Charakterisierung des statistischen Regelinduktions-Problems . . . 27
2.3.1 Die erste Form des statistischen Problems . . . 27
2.3.2 Modifikation des statistischen Problems . . . 28
2.3.3 Abgrenzung des statistischen Problems von verwandten Problemen . . . 30
2.3.3.1 Segmentierung des Prädiktorprofilraums . . . 30
2.3.3.2 Klassifikationsgütekriterien . . . 31
2.3.4 Vorschlag für eine zweite Modifikation der Zielaufgabe . . . 36
2.3.4.1 Definition des U_c-Kriteriums und des KNMK . . . 36
2.3.4.2 Vergleichbarkeit zu anderen Kriterien . . . 37
2.3.5 Fitting, Vorhersage und Kreuzvalidierung . . . 38
2.3.5.1 Unterscheidung zwischen Aufgabentypen . . . 38
2.3.5.2 Beispiele für *Overfitting* und Vergleich mit eigenen Simulationsergebnissen 40
2.4 Visualisierung von Klassifikationsergebnissen . . . 42
2.4.1 Das ROCS-Diagramm . . . 42
2.4.2 Iso-Klassifikationsgenauigkeitslinien für $n_1 = n_0$ . . . 43
2.4.3 Iso-Klassifikationsgenauigkeitslinien für $n_1 \neq n_0$ . . . 44
2.4.4 Beispieldiagramme . . . 47
2.5 Die Ergebnisse des Algorithmenvergleichs bei Montel (2005) . . . 50
2.5.1 Auswahl der Klassifikationsaufgaben und Algorithmen bei Montel . . . 51
2.5.2 Auswahl der Datensätze . . . 51
2.5.2.1 Der Datensatz *Hochschule* . . . 51
2.5.2.2 Der Datensatz *Beratungsunternehmen* . . . 52
2.5.2.3 Der Datensatz *Schüler* . . . 52
2.5.3 Klassifikationsergebnisse . . . 53
2.5.4 Visualisierung der Ergebnisse von Montel (2005) . . . 54

3 Heuristiken des Künstlichen Lebens 57
3.1 Grundstruktur von genetischen Algorithmen . . . 57
3.2 Komponenten genetischer Algorithmen . . . 58
3.2.1 Repräsentation . . . 59
3.2.1.1 Lösungsraum und Enkodierung . . . 59
3.2.1.2 Kodierung mit binären Strings . . . 59
3.2.1.3 Gray-Kodierung . . . 61
3.2.1.4 Direkte Kodierung mit reellen Zahlen . . . 61
3.2.2 Erstellung der Anfangspopulation . . . 62
3.2.2.1 Randomisierte Erzeugung . . . 62
3.2.2.2 Systematische Auswahl der ersten Individuen . . . 62
3.2.2.3 Empiriegestützte Auswahl der ersten Individuen . . . 63
3.2.2.4 Erstellung der ersten Population durch einen Hilfsalgorithmus . . . 63
3.2.3 Fitnessfunktion . . . 63
3.2.3.1 Objective functions und Fitnessfunktionen . . . 63
3.2.3.2 Fitnessfunktionen für Klassifikationsprobleme . . . 64
3.2.4 Selektionsoperatoren . . . 64
3.2.4.1 *Roulette-Wheel Selection* . . . 64
3.2.4.2 *Tournament Selection* . . . 66
3.2.4.3 *Probabilistic Tournament Selection* . . . 66
3.2.4.4 (μ, λ)- Selection . . . 68
3.2.4.5 $(\mu + \lambda)$- Selection . . . 69
3.2.4.6 Vergleich der Einzelverfahren . . . 69
3.2.4.7 Varianten und Ergänzungen . . . 70
3.2.4.8 Interaktive Selektion . . . 72
3.2.5 Fitnessskalierung . . . 72
3.2.5.1 Additive Skalierung . . . 72
3.2.5.2 Lineare Skalierung . . . 73
3.2.5.3 Sigma Truncation . . . 76
3.2.5.4 *Power Scaling* . . . 79
3.2.5.5 Transformation in Rangwerte . . . 79
3.2.6 Abbruchkriterien . . . 79
3.2.6.1 Generationenzahl . . . 79
3.2.6.2 Zielkriterium für die Modellgüte . . . 82
3.2.6.3 Stagnation des Fortschritts . . . 82
3.2.6.4 Kreuzvalidierung . . . 82
3.3 Genetische Operatoren . . . 83
3.3.1 Mutation . . . 83
3.3.1.1 Funktion der Mutation in genetischen Algorithmen . . . 84
3.3.1.2 Implementierung der einfachen Mutation . . . 84
3.3.2 Rekombination . . . 85
3.3.2.1 Funktion der Rekombination in genetischen Algorithmen . . . 85
3.3.2.2 Single Point Cross-Over . . . 85
3.3.2.3 Two Point Cross-Over . . . 86
3.3.2.4 Uniform Cross-Over . . . 87
3.3.2.5 Mehr als zwei Elternteile . . . 87
3.4 Erstellung von hierarchischen Regelmengen mit genetischen Algorithmen . . . 88
3.4.1 Pittsburgh vs. Michigan: Ein Repräsentationsproblem . . . 88
3.4.2 Kodierung von Antezedenzbedingungen für Regeln . . . 89
3.4.3 Kodierung von Prädiktionswerten für Regeln . . . 90
3.4.4 Spezielle Operatoren für die Regelinduktion . . . 90
3.4.5 Auswahl der Prädiktoren . . . 91
3.5 Genetische Algorithmen und Einschränkungen im Lösungsraum . . . 91
3.5.1 Das Problem unzulässiger Werte und Restriktionen . . . 91
3.5.2 Bestrafung und Belohnung als Lösungsansatz . . . 92

3.5.3 Reparaturmechanismen als Lösungsansatz . . . 94
3.6 Der Partikelschwarm und Optimierungsaufgaben . . . 95
3.6.1 Komponenten des Partikelschwarmalgorithmus . . . 95
3.6.1.1 Der Suchraum . . . 95
3.6.1.2 Die Partikel . . . 95
3.6.1.3 Nachbarschaft und Kommunikation . . . 96
3.6.1.4 Schematischer Ablauf der Partikelschwarmoptimierung . . . 97
3.6.2 Initialisierung des Partikelschwarms . . . 97
3.6.3 Wahl der Kommunikationstopologie . . . 98
3.6.3.1 Totalvernetzung . . . 98
3.6.3.2 Räder . . . 98
3.6.3.3 Ringe . . . 99
3.6.3.4 Randomisierte Topologien . . . 99
3.6.4 Beschleunigung und Bewegung . . . 99
3.6.4.1 Bewegungsprinzip . . . 99
3.6.4.2 Ermittlung von Bewegung und Geschwindigkeit . . . 100
3.6.4.3 Die Begrenzung der Geschwindigkeit . . . 101
3.6.5 Die Evaluation der Lösungen . . . 101
3.6.6 Das Abbruchkriterium . . . 102
3.6.7 Der Partikelschwarm in Pseudocode . . . 103
3.6.8 Individuelle Exploration und Kommunikation . . . 103
3.7 Praktische Anwendungen der vorgestellten Heuristiken . . . 104
3.7.1 Anwendungen genetischer Algorithmen . . . 104
3.7.2 Anwendungen von Partikelschwärmen . . . 108
3.7.2.1 Statistische Klassifikation . . . 108
3.7.2.2 Stromversorgung . . . 109
3.7.2.3 Projektplanung . . . 110
3.7.2.4 Strukturelle Reliabilität . . . 110
3.7.2.5 Weitere Anwendungen . . . 110
3.7.2.6 Überblick und Diskussion . . . 112

4 Grenzen der Rationalität und das Forschungsprogramm der Adaptiven Werkzeugkiste 115
4.1 Arten der Rationalität . . . 116
4.1.1 Einige begriffliche Unterscheidungen . . . 116
4.1.2 Die Problematik von Normen der Rationalität . . . 119
4.1.2.1 Gigerenzers Kritik am Fehlerkonzept von Kahneman und Tversky . . . 120
4.1.2.2 Der Streit . . . 121
4.1.2.3 Wacklige Prüfsteine menschlicher Rationalität . . . 124
4.2 Bounded Rationality bei Herbert Simon . . . 125
4.2.1 Heuristiken und Bounded Rationality . . . 127
4.2.2 Ameisen, Gelatine und Scherenklingen . . . 128
4.3 Ökologische Rationalität und die adaptive Werkzeugkiste . . . 131
4.3.1 Bounded Rationality im Forschungsprogramm der ABC-Gruppe . . . 131
4.3.2 Die Ökologische Rationalität der ABC-Gruppe . . . 133
4.3.3 Die *Adaptive Toolbox* . . . 135
4.3.4 Das Forschungsprogramm der ABC-Gruppe . . . 138
4.4 Die Rekognitions-Heuristik . . . 139
4.4.1 Bedingungen für die ökologische Rationalität der Heuristik . . . 140
4.4.2 Analytische Ergebnisse . . . 141
4.4.2.1 Analyse der Vorhersagefähigkeit . . . 141
4.4.2.2 Der Less-is-More Effekt . . . 143
4.4.3 Empirische Studien zur Rekognitionsheuristik . . . 145
4.4.3.1 Städtevergleiche . . . 145
4.4.3.2 Aktienmarkt . . . 147
4.4.3.3 Kritische Studien zur Rekognitionsheuristik . . . 148

4.5 Die Take-The-Best-Heuristik INTRO . . . 149
4.5.1 Darstellung des Algorithmus . . . 149
4.5.1.1 Der Standardalgorithmus . . . 149
4.5.1.2 Alternative Bestimmung der Cuereihenfolge . . . 152
4.5.1.3 Variationen von Take The Best: Take The Last und Minimalist . . . 152
4.5.2 Simulationsstudien zur TTB-Heuristik INTRO . . . 153
4.5.2.1 Testvergleich in zwanzig Umgebungen . . . 153
4.5.2.2 Kriterien zur Bestimmung der Cuereihenfolge für TTB . . . 155
4.5.3 Empirische Studien zu Take The Best . . . 156
4.5.3.1 Kompensatorische vs. nicht-kompensatorische Strategien . . . 157
4.5.3.2 Die empirischen Studien von Bröder . . . 158
4.5.3.3 Weitere Studien . . . 159

5 Regelinduktion mit Fast and Frugal Trees 161
5.1 Definition von *Fast and Frugal Trees* . . . 161
5.2 Eigenschaften von Fast and Frugal Trees . . . 163
5.3 Empirische Studien zur Anwendung von Fast and Frugal Trees . . . 165
5.4 Konstruktionsregeln für Fast and Frugal Trees . . . 169
5.4.1 Das Konstruktionsproblem . . . 169
5.4.2 Fünf Vorschläge für Konstruktionsregeln . . . 170
5.4.3 Beispielergebnisse für die Konstruktionsverfahren . . . 174
5.4.4 Einige Unterschiede zwischen den Strategien . . . 179
5.4.5 Herausforderungen für die Strategien . . . 180
5.4.6 Konstruktion neuer Konstruktionsverfahren . . . 184
5.5 Analytische Ergebnisse zu *Fast and Frugal Trees* . . . 186
5.5.1 Das *Splitting Profile* . . . 186
5.5.2 Fast and Frugal Trees und lineare Klassifikatoren . . . 187
5.5.3 Fast and Frugal Trees und *Take The Best* . . . 188

6 Simulationen zur Regelinduktion mit empirischen Daten 190
6.1 Untersuchung der Klassifikationsleistung des gier.downhill-Algorithmus an drei Datensätzen 190
6.1.1 Darstellung von gier.downhill . . . 190
6.1.1.1 Grundform des Algorithmus . . . 190
6.1.1.2 Die Bedeutung der Parameter . . . 191
6.1.2 Simulationsergebnisse hinsichtlich der Klassifikationsgenauigkeit . . . 192
6.1.2.1 Ergebnisse für den Datensatz *Hochschule* . . . 193
6.1.2.2 Ergebnisse für den Datensatz *Beratungsunternehmen* . . . 196
6.1.3 Ergebnisse hinsichtlich Δ_{UAPV}-Score und Regelanzahl . . . 199
6.1.3.1 Ergebnisse für den Datensatz *Hochschule* . . . 199
6.1.3.2 Ergebnisse für den Datensatz *Beratungsunternehmen* . . . 202
6.1.4 ROCS-Diagramme . . . 202
6.1.4.1 Ergebnisse für den Datensatz *Hochschule* . . . 202
6.1.4.2 Ergebnisse für den Datensatz *Beratungsunternehmen* . . . 206
6.1.5 Diskussion der Ergebnisse und Empfehlungen für die Parameterwahl . . . 208
6.2 Untersuchung der Klassifikationsleistung von Fast and Frugal Trees an dreißig Datensätzen 208
6.2.1 Darstellung der Wettbewerbsteilnehmer . . . 209
6.2.1.1 *Context Model* und PROBEX als Beispiele für *Exemplar Models* . . . 209
6.2.1.2 *Exemplar Models* im Wettbewerb . . . 210
6.2.1.3 CART und Logistische Regression . . . 210
6.2.1.4 Eingesetzte Varianten der *Fast and Frugal Trees* . . . 211
6.2.2 Datensätze und Wettbewerbsmodus . . . 211
6.2.2.1 Überblick über die verwendeten Datensätze . . . 211
6.2.2.2 Turniermodus . . . 214
6.2.3 Ergebnisse des Wettbewerbs . . . 215
6.2.3.1 Ergebnisse hinsichtlich der Wahl des Klassifikationsverfahrens . . . 215

6.2.3.2 Ergebnisse hinsichtlich des Vergleichs mit konkurrierenden Algorithmen . 217
6.2.4 Hinweise auf ökologische Rationalität von *Fast and Frugal Trees* 220
6.2.4.1 Differenzierte Darstellung der Klassifikationsleistungen 220
6.2.4.2 Medizinische Datensätze 220
6.2.4.3 Einzelne Datensätze 226
6.2.4.4 Ergebnisse für die drei Ausgangsdatensätze 226
6.3 Untersuchung der Klassifikationsleistung von gier.ga und gier.swarm 228
6.3.1 Präzisierung der Aufgabenstellung 228
6.3.1.1 Verwendete Algorithmen 228
6.3.1.2 Datensätze und Vorgabemodus 232
6.3.2 Ergebnisse für den Datensatz *Hochschule* 232
6.3.2.1 Ergebnisse für gier.ga 233
6.3.2.2 Ergebnisse für gier.swarm 233
6.3.3 Vergleich der Algorithmen 234
6.3.3.1 Vergleich von gier.ga und gier.swarm 234
6.3.3.2 ROCS-Diagramme für die Algorithmen 235
6.3.3.3 Streudiagramme der Trials für den Datensatz *Hochschule* 237
6.3.3.4 Vergleich mit konkurrierenden Algorithmen 239
6.4 Diskussion der Ergebnisse 240
6.4.1 Fazit des Methodenvergleichs 240
6.4.2 Anwendungsempfehlungen und Modifikationsmöglichkeiten für *Fast and Frugal Trees* 241
6.4.2.1 Erste Variation: Binarisierung und mögliche Erweiterungen 242
6.4.2.2 Zweite Variation: Cuereihenfolge versus Entscheidungsreihenfolge 243

7 Die paradoxe Rationalität einfacher Klassifikationsverfahren 244
7.1 Einfache Heuristiken und das *Frame*-Problem 244
7.1.1 Roboter, die Kognitionswissenschaft und das *Frame*-Problem 244
7.1.2 Forsters Komplexitätserhaltungssatz 246
7.1.3 Der Mechaniker in der Werkzeugkiste 250
7.2 Ein Rückblick auf das Ausgangsproblem 253
7.2.1 Ein Modellierungsvorschlag für die Regelinduktion 253
7.2.1.1 Modellierung von Klassifikation als q-Morphismus 254
7.2.1.2 Modell und Regelinduktionsproblem 261
7.2.2 Ausgeklammerte Aspekte der Aufgabenstellung 262
7.2.2.1 Detektoren und Messinstrumente 262
7.2.2.2 Das Kriterienproblem 264
7.2.2.3 Prädiktoren und Verrechnungsmodelle 267
7.2.2.4 Institutionelle Rahmenbedingungen 269
7.2.2.5 Gesellschaftliche Perspektiven auf Eignungsdiagnostik 272
7.3 Über Einfachheit, Komplexität und die Welt hinter den Heuristiken 274
7.3.1 Einfache Heuristiken und epistemisch verantwortbare Entscheidungen 274
7.3.2 Einfachheit und Komplexität 278
7.3.3 Hinter den Heuristiken 283

Literaturverzeichnis 287

Anhang 315

A Ergänzungen zu Partikelschwärmen: Analyse der Dynamik und Erweiterungen 316
A.1 Die Dynamik von Partikelschwärmen in eindimensionalen Suchräumen 316
A.1.1 Das vereinfachte System 316
A.1.2 Auswirkungen der Wahl von φ_{max} und V_{max} bei randomisierter Beschleunigung . . 318
A.1.3 Auswirkungen der Wahl von φ_{max} und V_{max} bei nichtrandomisierter Beschleunigung 320
A.2 Surfende Partikel 323
A.3 Die Dynamik von Partikelschwärmen in zweidimensionalen Suchräumen 323
A.3.1 Das erweiterte vereinfachte System 323

A.3.2 Dynamisches Verhalten bei einer gespeicherten Optimalposition . . . 326
A.3.3 Untersuchungen mit zwei gespeicherten Optimalpositionen . . . 328
A.4 Mögliche Erweiterungen des Partikelschwarmalgorithmus . . . 334
A.4.1 Constriction . . . 334
A.4.2 Inertia . . . 334
A.4.3 Zusätzliche Entwicklungsideen für Schwarmalgorithmen . . . 338

B Ergänzungen zu genetischen Algorithmen: TSP und Inselalgorithmen 339
B.1 Genetische Algorithmen für das TSP-Problem . . . 339
B.1.1 Kodierung von Permutationen und Sequenzen . . . 339
B.1.2 Fitnessfunktionen für das TSP . . . 341
B.1.3 Mutation für das TSP-Problem . . . 341
B.1.4 Rekombination für das TSP-Problem . . . 343
B.2 Inselalgorithmen . . . 344
B.2.1 Inselalgorithmen zur Klassifikation . . . 344
B.2.2 Grundalgorithmus . . . 345
B.2.3 Einzelalgorithmen . . . 348
B.2.4 Migrationsformen . . . 348
B.2.5 Abbruchkriterium . . . 351
B.2.6 Inseltopologien . . . 351
B.2.7 Anwendungen von Inselalgorithmen . . . 356

C Entwicklung und Struktur der begleitenden Software 357
C.1 Design- und Implementierungsentscheidungen . . . 358
C.1.1 Modularisierung der Arbeitsschritte . . . 358
C.1.2 Die Script-Sprache gier.script . . . 358
C.1.3 Dateiformate . . . 361
C.1.4 Beschreibung der Oberfläche und Steuerung . . . 362
C.2 Objektorientierte Programmierung von Algorithmen . . . 364
C.2.1 Relevante Konzepte Objektorientierter Programmierung . . . 364
C.2.2 Objektorientierte Programmierung und visuelle Methodenbaukästen . . . 369
C.3 Darstellung der Algorithmensammlung in $\text{GIER}_{[1.0]}$ . . . 375
C.3.1 Algorithmenübergreifende Eigenschaften und Parameter . . . 375
C.3.2 Charakterisierung des gier.ga-Algorithmus . . . 378
C.3.3 Charakterisierung des gier.gaset-Algorithmus . . . 386
C.3.4 Charakterisierung des gier.downhill-Algorithmus . . . 393
C.3.5 Charakterisierung des gier.swarm-Algorithmus . . . 395
C.3.6 Charakterisierung des gier.swarmset-Algorithmus . . . 397
C.3.7 Charakterisierung des gier.island-Algorithmus . . . 398
C.3.8 Die Verschachtelung von Algorithmen . . . 402
C.3.9 Visualisierungen . . . 403
C.4 Beschreibung zusätzlicher Komponenten . . . 404
C.4.1 Charakterisierung des gier.tsp-Algorithmus . . . 404
C.4.2 Die TSP-Demonstration . . . 406
C.4.3 Darstellung von $\text{GIER}_{[1.0]}$.typag . . . 408
C.4.4 Integration von r2-f2 in die Toolbox . . . 411
C.4.5 Komponenten zur Simulation künstlichen Lebens . . . 418
C.5 Programmierung einfacher Heuristiken: Überblick über den zweiten Werkzeugkasten . . . 420
C.5.1 Komponenten von $\text{GIER}_{[1.0]}$.simple_heuristics . . . 420
C.5.2 Einige visuelle Demonstrationen . . . 422

D Vollständige Ergebnisse der Fast and Frugal Tree-Simulation 428
D.1 Vergleichsalgorithmen . . . 428
D.2 Ursprüngliche Baumalgorithmen . . . 436
D.3 Neue Baumalgorithmen . . . 446

Abbildungsverzeichnis

2.1 Günstige Funktion für die Anwendung der *Hill-Climbing*-Heuristik . . . 13
2.2 Ungünstige Funktion für die Anwendung der *Hill-Climbing*-Heuristik . . . 14
2.3 Beispielsegmentierung durch vier Regeln . . . 31
2.4 Ergebnisäquivalenter Entscheidungsbaum zum Beispielregelsystem . . . 32
2.5 ROC-Shading-Diagramm für $n_1 = n_0$ . . . 43
2.6 ROCS-Diagramm mit Iso-Genauigkeitslinien, $n_1 = n_0$ . . . 45
2.7 ROCS-Diagramm mit Iso-Genauigkeitslinien, $n_1 = 2n_0$ . . . 46
2.8 ROC-Diagramm mit Standard-Iso-Genauigkeitslinien, $n_1 = 2n_0$ . . . 46
2.9 ROCS-Diagramm für sechs fiktive Klassifikationsalgorithmen ($n_1 = 300$, $n_0 = 100$) . . . 47
2.10 ROCS-Diagramm für sechs fiktive Klassifikationsalgorithmen ($n_1 = 100$, $n_0 = 300$) . . . 49
2.11 Beispiele für ROCS-Diagramme bei unterschiedlichen Klassengrößen . . . 50
2.12 ROCS-Diagramm für die Ergebnisse von Montel (2005) . . . 56

3.1 Schritte eines einfachen genetischen Algorithmus . . . 58
3.2 Beispielpopulation . . . 64
3.3 Roulette-Wheel Selection . . . 65
3.4 Tournament Selection . . . 67
3.5 Probabilistic Tournament Selection . . . 67
3.6 (μ, λ)-Selektion . . . 68
3.7 $(\mu + \lambda)$-Selektion . . . 69
3.8 Boltzmann Trial 1 . . . 71
3.9 Boltzmann Trial 2 . . . 71
3.10 Lineare Skalierung 1 . . . 74
3.11 Lineare Skalierung 2 . . . 74
3.12 Lineare Skalierung 3 . . . 75
3.13 Lineare Skalierung für $A_1(0)$ . . . 77
3.14 Sigma Truncation für $A_1(0)$ . . . 78
3.15 Power Scaling für $A_1(0)$ . . . 80
3.16 Rangtransformation für $A_1(0)$ . . . 81
3.17 Abbruch über Kreuzvalidierung . . . 83
3.18 Mutationsoperator . . . 84
3.19 Beispiel für die Anwendung des Mutationsoperators . . . 85
3.20 Single Point Cross-Over . . . 86
3.21 Two Point Cross-Over . . . 87
3.22 Uniform Cross-Over . . . 87
3.23 Uniform P-Cross-Over . . . 88
3.24 Cross-Over mit vier Eltern . . . 89
3.25 Beispiel für einen Suchraum mit unzulässigen Lösungen . . . 92
3.26 Interne und externe Variablen der Schwarmpartikel . . . 96
3.27 Repräsentation der Variablenwerte eines Partikels im zweidimensionalen Raum . . . 96
3.28 Abfolge der Partikelschwarmoptimierung . . . 97
3.29 Darstellung einer Gruppe von Partikeln ohne Kommunikationsrelationen . . . 98
3.30 Darstellung von drei Schwarmkommunikationstopologien . . . 99
3.31 Darstellung von randomisierten Kommunikationstopologien . . . 100
3.32 Beschleunigungsbeispiel für Partikel . . . 101
3.33 Partikelschwärme und lokale Maxima im Suchraum . . . 104

4.1 Entscheidungsbaum der Rekognitionsheuristik . . . 139
4.2 Ökologische Rationalität der Recognition-Heuristik . . . 140
4.3 Beispielobjekte für die Rekognitionsheuristik . . . 141
4.4 Aufteilung der Paarvergleiche nach Rekognition . . . 142
4.5 Illustration des Less-Is-More Effekts . . . 144
4.6 Schritte des TTB-Algorithmus . . . 149
4.7 Der TTB-Algorithmus als Entscheidungsbaum . . . 151

5.1 Abgrenzung von FaF Trees gegenüber anderen Bäumen . . . 162
5.2 Sonderformen von FaF Trees . . . 164
5.3 Entscheidungsbaum für die Einweisung von Patienten mit Verdacht auf Herzkrankheit . . 165
5.4 Entscheidungsbaum für Umschulungsentscheidungen . . . 168
5.5 Entscheidungsbaum für die Vorgesetztenbeurteilung . . . 169
5.6 Anwendung der Konstruktionsregeln Faf MaxSens, Faf MaxVal, FaF ZigSens und FaF ZigVal . . . 171
5.7 Anwendung der Konstruktionsregel Faf MaxAcc . . . 172
5.8 Hierarchischer Cueverteilungsbaum . . . 175
5.9 Hierarchischer Cueverteilungsbaum mit Profilentscheidungen für FaF MaxSens . . . 176
5.10 ROCS-Diagramm für Verteilung 1 . . . 177
5.11 ROCS-Diagramm für Verteilung 2 . . . 178
5.12 ROCS-Diagramm für Verteilung 3 . . . 178
5.13 Anwendung der Konstruktionsregeln Faf MaxSens, Faf MaxVal, FaF ZigSens und FaF ZigVal 2 . . . 181
5.14 Anwendung der Konstruktionsregel Faf MaxAcc 2 . . . 182
5.15 ROCS-Diagramm für Verteilung 4 . . . 184
5.16 ROCS-Diagramm für die neuen Bäume (Verteilung 4) . . . 185
5.17 TTB als *Fast and Frugal Tree* . . . 189

6.1 Iteration im $\textsc{Gier}_{[1.0]}$.downhill-Algorithmus . . . 191
6.2 Intervallanpassung nach einem Downhill-Schritt . . . 192
6.3 VBk-Darstellung des eingesetzten Algorithmus $\textsc{Gier}_{[1.0]}$.downhill . . . 193
6.4 Kastendiagramm Klassifikationsgenauigkeit Fitting für die Downhillvarianten (Hochschule) 195
6.5 Kastendiagramm Klassifikationsgenauigkeit Prediction für die Downhillvarianten (Hochschule) . . . 195
6.6 Kastendiagramm zur Verschlechterung der Klassifikationsgenauigkeit für die Downhillvarianten (Hochschule) . . . 196
6.7 Kastendiagramm Klassifikationsgenauigkeit Fitting für die Downhillvarianten (Beratung) 198
6.8 Kastendiagramm Klassifikationsgenauigkeit Prediction für die Downhillvarianten (Beratung) 198
6.9 Kastendiagramm zur Verschlechterung der Klassifikationsgenauigkeit für die Downhillvarianten (Beratungsunternehmen) . . . 199
6.10 Kastendiagramm Δ_{UAPV}-Score Fitting für die Downhillvarianten (Hochschule) . . . 201
6.11 Kastendiagramm Δ_{UAPV}-Score Prediction für die Downhillvarianten (Hochschule) . . . 201
6.12 Kastendiagramm Regelanzahl für die Downhillvarianten (Hochschule) . . . 202
6.13 Kastendiagramm Δ_{UAPV}-Score Fitting für die Downhillvarianten (Beratungsunternehmen) 204
6.14 Kastendiagramm Δ_{UAPV}-Score Prediction für die Downhillvarianten (Beratungsunternehmen) . . . 204
6.15 Kastendiagramm Regelanzahl für die Downhillvarianten (Beratungsunternehmen) . . . 205
6.16 ROCS-Diagramm für die Fitting-Ergebnisse der Downhill-Varianten (Hochschule) . . . 205
6.17 ROCS-Diagramm für die Prediction-Ergebnisse der Downhill-Varianten (Hochschule) . . . 206
6.18 ROCS-Diagramm für die Fitting-Ergebnisse der Downhill-Varianten (Beratungsunternehmen) . . . 207
6.19 ROCS-Diagramm für die Prediction-Ergebnisse der Downhill-Varianten (Beratungsunternehmen) . . . 207
6.20 Ergebnisse der *Fast and Frugal Trees* mit den fünf ursprünglichen Konstruktionsverfahren (30 Datensätze) . . . 216

6.21 Ergebnisse der *Fast and Frugal Trees* mit den sechs modifizierten Konstruktionsverfahren (30 Datensätze) . . . 218
6.22 Ergebnisse der konkurrierenden Algorithmen (30 Datensätze) . . . 219
6.23 Vergleich der besten Prediction-Leistungen . . . 221
6.24 Balkendiagramm für den Algorithmenvergleich (30 Datensätze) . . . 222
6.25 Balkendiagramm für den Algorithmenvergleich (11 medizinische Datensätze) . . . 222
6.26 Balkendiagramm für den Algorithmenvergleich (Datensatz *Broadway*) . . . 226
6.27 Balkendiagramm für den Algorithmenvergleich (Datensatz *Yeast*) . . . 227
6.28 Balkendiagramm für den Algorithmenvergleich (Datensätze *Hochschule*, *Beratungsunternehmen*, *Schule*) . . . 227
6.29 VBk-Darstellung des verwendeten Algorithmus $\text{GIER}_{[1.0]}$.ga . . . 230
6.30 VBk-Darstellung des verwendeten Algorithmus $\text{GIER}_{[1.0]}$.swarm . . . 231
6.31 Vergleich der Regelzahlen für gier.ga und gier.swarm . . . 235
6.32 Vergleich der Fitting-Ergebnisse für gier.ga und gier.swarm . . . 235
6.33 Vergleich der Prediction-Ergebnisse für gier.ga und gier.swarm . . . 236
6.34 ROCS-Diagramm für die Ergebnisse von gier.ga und gier.swarm für den Datensatz *Hochschule* . . . 236
6.35 ROCS-Diagramm für die Ergebnisse von gier.ga und gier.swarm für den Datensatz *Beratungsunternehmen* . . . 237
6.36 Streudiagramm der Ergebnisse von gier.ga für den Datensatz *Hochschule* (50%) . . . 238
6.37 Streudiagramm der Ergebnisse von gier.swarm für den Datensatz *Hochschule* (50%) . . . 238
6.38 Streudiagramm der Ergebnisse von gier.swarm für den Datensatz *Hochschule* (80%) . . . 239

7.1 Kategorisierung im Modell . . . 255
7.2 Klassifikation im Modell . . . 255
7.3 Diagnose im Modell . . . 256
7.4 Das Ausgangsproblem als Homomorphismus . . . 258
7.5 Das Ausgangsproblem als q-Morphismus . . . 260
7.6 Komplexitäterhaltung bei Weick . . . 280
7.7 Bertrands Problem . . . 282

A.1 Darstellung des eindimensionalen Simulationssystems . . . 317
A.2 Schwarmsimulation im eindimensionalen Suchraum . . . 317
A.3 Schwarmsimulation im eindimensionalen Suchraum, Variante 1 . . . 318
A.4 Schwarmsimulation im eindimensionalen Suchraum, Variante 2 . . . 319
A.5 Schwarmsimulation im eindimensionalen Suchraum, Variante 3 . . . 319
A.6 Schwarmsimulation im eindimensionalen Suchraum, Variante 4 . . . 320
A.7 Schwarmsimulation im eindimensionalen Suchraum (Überblick) . . . 321
A.8 Nichtrandomisierte Schwarmsimulation im eindimensionalen Suchraum: Beispiel . . . 322
A.9 Nichtrandomisierte Schwarmsimulation 1 . . . 324
A.10 Nichtrandomisierte Schwarmsimulation 2 . . . 325
A.11 Darstellung eines Partikels im zweidimensionalen Suchraum . . . 326
A.12 Ausgangsbedingung im zweidimensionalen Suchraum mit einem Optimum . . . 326
A.13 Schwarmsimulation im zweidimensionalen Suchraum . . . 327
A.14 Schwarmsimulation im zweidimensionalen Suchraum, Variante2 . . . 327
A.15 Nichtrandomisierte Schwarmsimulation im zweidimensionalen Raum . . . 329
A.16 Ausgangsbedingung im zweidimensionalen Suchraum mit zwei Speicherstellen . . . 330
A.17 Schwarmsimulation im zweidimensionalen Suchraum . . . 332
A.18 Nichtrandomisierte Schwarmsimulation im zweidimensionalen Suchraum mit zwei Speicherstellen . . . 333
A.19 Schwarmsimulation im zweidimensionalen Suchraum mit Simplest Constriction 1 . . . 334
A.20 Schwarmsimulation im zweidimensionalen Suchraum mit Simplest Constriction 2 . . . 335
A.21 Schwarmsimulation im zweidimensionalen Suchraum mit Simplest Constriction . . . 335
A.22 Schwarmsimulation im zweidimensionalen Suchraum mit Inertia Weight . . . 336
A.23 Funktionen für die Bestimmung des Inertia-Faktors . . . 336

A.24 Funktionen für die Bestimmung des Inertia-Faktors 2 . . . 337
A.25 Funktionen für die Bestimmung des Inertia-Faktors 3 . . . 337

B.1 Mutation für eine TSP-Lösung . . . 342
B.2 Alternative Mutation für die TSP-Lösung . . . 343
B.3 OX-Cross-Over . . . 344
B.4 OX-Cross-Over für das TSP-Problem . . . 345
B.5 Darstellung einer Inselgruppe . . . 346
B.6 Schritte eines Inselalgorithmus . . . 346
B.7 Darstellung von Evolution und Migration am Beispiel einer Inselgruppe . . . 347
B.8 Integration der Ergebnisse . . . 348
B.9 Einzelinsel vor Migration . . . 349
B.10 Migrationsbeispiel 1 . . . 350
B.11 Migrationsbeispiel 2 . . . 350
B.12 Einseitige Stepping-Stones-Topologie . . . 351
B.13 Reziproke Stepping-Stones-Topologie . . . 352
B.14 Reziproke Kreistopologie . . . 352
B.15 Einseitige Kreistopologie . . . 352
B.16 Mehrfachkreistopologie . . . 353
B.17 2D-Gitter-Topologie . . . 354
B.18 Torustopologie . . . 355
B.19 Pyramidentopologie . . . 355

C.1 Verzeichnisstruktur von GIER$_{[1.0]}$ . . . 364
C.2 Unterverzeichnisstruktur für Auswertungen . . . 365
C.3 Unterverzeichnisstruktur für Scripte . . . 366
C.4 UML-Darstellung von Klassen und Objekten . . . 366
C.5 UML-Darstellung einer Aggregationsbeziehung . . . 367
C.6 UML-Darstellung einer Vererbungsbeziehung . . . 368
C.7 Modifizierte UML-Darstellung von Objekten mit vererbten Attributen . . . 369
C.8 Beispiel für polymorphe Methoden in UML-Darstellung . . . 370
C.9 UML-Darstellung eines fiktiven Softwareprojektes zur Klassifikation . . . 371
C.10 Beispiel für Algorithmenklassen als Aggregatklassen . . . 372
C.11 UML-Darstellung einer Klasse aus GIER$_{[1.0]}$ . . . 373
C.12 UML-Darstellung eines Moduls von t_swarm . . . 374
C.13 Auszug der VBk-Darstellung von Schwarmalgorithmen . . . 375
C.14 VBk-Darstellung des Algorithmus GIER$_{[1.0]}$.ga . . . 378
C.15 VBk-Darstellung der Komponente GIER$_{[1.0]}$.ga.creation . . . 379
C.16 VBk-Darstellung der Komponente GIER$_{[1.0]}$.ga.bewertung . . . 380
C.17 VBk-Darstellung der Komponente GIER$_{[1.0]}$.ga.fitness . . . 381
C.18 VBk-Darstellung der Komponente GIER$_{[1.0]}$.ga.skalierung . . . 381
C.19 VBk-Darstellung der Komponente GIER$_{[1.0]}$.ga.penalty . . . 382
C.20 VBk-Darstellung der Komponente GIER$_{[1.0]}$.ga.bonus . . . 382
C.21 VBk-Darstellung der Komponente GIER$_{[1.0]}$.ga.selektion . . . 383
C.22 VBk-Darstellung der genetischen Operatoren von GIER$_{[1.0]}$.ga . . . 384
C.23 VBk-Darstellung der Komponente GIER$_{[1.0]}$.ga.reparatur . . . 385
C.24 VBk-Darstellung der Komponente GIER$_{[1.0]}$.ga.abbruch . . . 386
C.25 VBk-Darstellung des Algorithmus GIER$_{[1.0]}$.gaset . . . 387
C.26 VBk-Darstellung des Prinzips der modularen Wiederverwendung . . . 388
C.27 VBk-Darstellung der nicht in GIER$_{[1.0]}$.ga enthaltenen Komponenten . . . 389
C.28 VBk-Darstellung der Komponente GIER$_{[1.0]}$.gaset.creation . . . 390
C.29 VBk-Darstellung der genetischen Operatoren von GIER$_{[1.0]}$.gaset . . . 392
C.30 VBk-Darstellung des Algorithmus GIER$_{[1.0]}$.downhill . . . 393
C.31 VBk-Darstellung der Komponente GIER$_{[1.0]}$.downhill.entscheidung . . . 394
C.32 VBk-Darstellung der Komponente GIER$_{[1.0]}$.downhill.optimierung . . . 394

C.33 VBk-Darstellung von GIER$_{[1.0]}$.downhill.variablenreduktion 395
C.34 VBk-Darstellung des Algorithmus GIER$_{[1.0]}$.swarm 396
C.35 VBk-Darstellung der Komponente GIER$_{[1.0]}$.swarm.topologie 396
C.36 VBk-Darstellung der Komponente GIER$_{[1.0]}$.swarm.movement 397
C.37 VBk-Darstellung der Komponente GIER$_{[1.0]}$.swarm.inertia 397
C.38 VBk-Darstellung des Algorithmus GIER$_{[1.0]}$.swarmset 398
C.39 VBk-Darstellung des Algorithmus GIER$_{[1.0]}$.island 399
C.40 VBk-Darstellung der Komponente GIER$_{[1.0]}$.island.topologie 400
C.41 VBk-Darstellung der Komponente GIER$_{[1.0]}$.island.migration 400
C.42 VBk-Darstellung der Komponente GIER$_{[1.0]}$.island.epochenabbruch 402
C.43 Darstellung möglicher Klassenverschachtelung 402
C.44 Beispiel für ein Doppelhistogramm in GIER$_{[1.0]}$ 404
C.45 VBk-Darstellung des Algorithmus GIER$_{[1.0]}$.tsp 405
C.46 VBk-Darstellung der Komponente GIER$_{[1.0]}$.tsp.bewertung 406
C.47 VBk-Darstellung der genetischen Operatoren von GIER$_{[1.0]}$.tsp 406
C.48 VBk-Darstellung des Algorithmus GIER$_{[1.0]}$.typag 410
C.49 Beispielergebnis für den r2-f2-Algorithmus 411
C.50 Ergebnisäquivalenter Entscheidungsbaum zum Beispielergebnis des r2-f2-Algorithmus . . . 412
C.51 VBk-Darstellung des Algorithmus GIER$_{[1.0]}$.r2-f2 416
C.52 VBk-Darstellung der in GIER$_{[1.0]}$.r2-f2 enthaltenen Komponenten 417
C.53 VBk-Darstellung der Komponente GIER$_{[1.0]}$.r2-f2.ranking 417
C.54 VBk-Darstellung der Komponente GIER$_{[1.0]}$.r2-f2.grenzen 418
C.55 Vollständige Regelinduktionstoolbox in GIER$_{[1.0]}$ 419
C.56 Beispieldarstellung für einen Heuristikenvergleich 422
C.57 Beispieldarstellung für einen Vergleich zwischen verschiedenen Cueprozentwerten 423
C.58 Beispieldarstellung für die Analyse der Auswirkung unterschiedlicher bekannter Cuewertanteile 423
C.59 Beispieldarstellung für eine Analyse der Auswirkung der Anzahl bekannter Städte 424
C.60 Beispieldiagramm zur Manipulation der Rekognitionsvalidität 424
C.61 Beispiel für ein TTB-Pareto-Diagramm 425
C.62 Beispiel für ein TTB-Sequenzinversionsdiagramm 426
C.63 Beispiel für ein Stabilitästhistogramm für lexikografische Paarvergleiche 426

Tabellenverzeichnis

1.1 Module und Codezeilen der Begleitsoftware . . . 6

2.1 Konfusionsmatrix . . . 32
2.2 Beispielklassifikationsergebnis 1 . . . 33
2.3 Beispielklassifikationsergebnis 2 . . . 33
2.4 Beispielklassifikationsergebnis 3 . . . 35
2.5 Beispielklassifikationsergebnis 4 . . . 35
2.6 Nutzentabelle . . . 36
2.7 Generalisierungsergebnisse klassischer und konnektionistischer Klassifizierungsverfahren . 41
2.8 Ergebnisse für sechs fiktive Klassifikationsalgorithmen ($n_1 = 300$, $n_0 = 100$) . . . 47
2.9 Ergebnisse für sechs fiktive Klassifikationsalgorithmen ($n_1 = 100$, $n_0 = 300$) . . . 48
2.10 Algorithmenvergleich aus Montel (2005) - Genauigkeit . . . 55
2.11 Algorithmenvergleich aus Montel (2005) - Δ_{UAPV}-Score . . . 55

3.1 Binäre Kodierung . . . 59
3.2 Binäre Kodierung für kontinuierliche Zahlen . . . 60
3.3 Beispiel für Gray Code . . . 61
3.4 Lateinisches Quadrat für vier Elemente . . . 88
3.5 GA-Anwendungen in der Chemie und Physik . . . 105
3.6 GA-Anwendungen in der Biochemie . . . 105
3.7 Visuelle und statistische GA-Anwendungen . . . 105
3.8 GA-Anwendungen in den Ingenieurswissenschaften . . . 106
3.9 Wirtschaftsbezogene GA-Anwendungen . . . 106
3.10 GA-Anwendungen in der Medizin . . . 107
3.11 GA-Anwendungen in der Robotik . . . 107
3.12 GA-Anwendungen in Simulationen . . . 107
3.13 Zusammenfassende Angaben zu PSO-Anwendungen 1 . . . 113
3.14 Zusammenfassende Angaben zu PSO-Anwendungen 2 . . . 114

4.1 Algorithmenergebnisse über 20 Datensätze . . . 153
4.2 Ergebnisse der Varianten von TTB . . . 155

5.1 Anzahl möglicher FaF Trees . . . 170
5.2 Beispielergebnisse für vier Cues . . . 170
5.3 Gütekriterien für vier Cues . . . 170
5.4 Cuewertverteilungen . . . 173
5.5 Ergebnisse für Verteilung 1 . . . 174
5.6 Ergebnisse für Verteilung 2 . . . 177
5.7 Ergebnisse für Verteilung 3 . . . 178
5.8 Verteilung der Cueprofile auf die Entscheidungsschritte . . . 179
5.9 Beispielergebnisse für vier Cues bei stark unterschiedlichen Gruppengrößen . . . 180
5.10 Gütekriterien für vier Cues . . . 180
5.11 Cuewertverteilung 2 . . . 183
5.12 Ergebnisse für Verteilung 4 . . . 184
5.13 Ergebnisse der neuen Bäume für Verteilung 4 . . . 185

6.1 Klassifikationsgenauigkeit Fitting für die Downhillvarianten (Hochschule) . . . 194

6.2 Klassifikationsgenauigkeit Prediction für die Downhillvarianten (Hochschule) 194
6.3 Klassifikationsgenauigkeit Fitting für die Downhillvarianten (Beratungsunternehmen) . . . 197
6.4 Klassifikationsgenauigkeit Prediction für die Downhillvarianten (Beratungsunternehmen) . 197
6.5 Δ_{UAPV}-Score Prediction für die Downhillvarianten (Hochschule) 200
6.6 Regelzahl für die Downhillvarianten (Hochschule) 200
6.7 Δ_{UAPV}-Score Prediction für die Downhillvarianten (Beratungsunternehmen) 203
6.8 Regelzahl für die Downhillvarianten (Beratungsunternehmen) 203
6.9 Überblick über 30 Datensätze der Testsuite 212
6.10 Ausgewählte Wettbewerbsergebnisse für die fünf *Fast and Frugal Tree*-Algorithmen mit den ursprünglichen Konstruktionsverfahren 223
6.11 Ausgewählte Wettbewerbsergebnisse für die sechs modifizierten *Fast and Frugal Tree*-Algorithmen 224
6.12 Ausgewählte Wettbewerbsergebnisse für die *Context Models* sowie LR und CART 225
6.13 Ergebnisse von vier Verfahren (Ausgangsdatensätze) 229
6.14 Trialübersicht im Algorithmenvergleich von GIER$_{[1.0]}$.ga und GIER$_{[1.0]}$.swarm 232
6.15 Ergebnisse für die GIER$_{[1.0]}$.ga-Varianten für den Datensatz *Hochschule* (50%) 233
6.16 Ergebnisse für die GIER$_{[1.0]}$.ga-Varianten für den Datensatz *Beratungsunternehmen* (50%) 233
6.17 Ergebnisse für die GIER$_{[1.0]}$.swarm-Varianten für den Datensatz *Hochschule* (50%) 233
6.18 Ergebnisse für die GIER$_{[1.0]}$.swarm-Varianten für den Datensatz *Hochschule* (80%) 234
6.19 Ergebnisse für die GIER$_{[1.0]}$.swarm-Varianten für den Datensatz *Beratungsunternehmen* (50%) 234
6.20 Fitting-Ergebnisse aller Algorithmen 240
6.21 Prediction-Ergebnisse aller Algorithmen 241

C.1 Liste der Dateitypen in GIER$_{[1.0]}$ 362

D.1 Fitting-Ergebnisse für CART (33 Datensätze) 429
D.2 Prediction-Ergebnisse für CART (33 Datensätze) 430
D.3 Fitting-Ergebnisse für die logistische Regression (33 Datensätze) 431
D.4 Prediction-Ergebnisse für die logistische Regression (33 Datensätze) 432
D.5 Fitting-Ergebnisse für das Context Model Ex 0.3(33 Datensätze) 433
D.6 Prediction-Ergebnisse für das Context Model Ex 0.3(33 Datensätze) 434
D.7 Fitting-Ergebnisse für das Context Model Ex0.4 (33 Datensätze) 435
D.8 Prediction-Ergebnisse für das Context Model Ex0.4 (33 Datensätze) 437
D.9 Fitting-Ergebnisse für das Context Model Ex0.5 (33 Datensätze) 438
D.10 Prediction-Ergebnisse für das Context Model Ex0.5 (33 Datensätze) 439
D.11 Fitting-Ergebnisse für das FaF MaxSens (33 Datensätze) 440
D.12 Prediction-Ergebnisse für FaF MaxSens (33 Datensätze) 441
D.13 Fitting-Ergebnisse für das FaF ZigSens (33 Datensätze) 442
D.14 Prediction-Ergebnisse für FaF ZigSens (33 Datensätze) 443
D.15 Fitting-Ergebnisse für das FaF MaxVal (33 Datensätze) 444
D.16 Prediction-Ergebnisse für FaF MaxVal (33 Datensätze) 445
D.17 Fitting-Ergebnisse für das FaF ZigVal (33 Datensätze) 447
D.18 Prediction-Ergebnisse für FaF ZigVal (33 Datensätze) 448
D.19 Fitting-Ergebnisse für das FaF MaxAcc (33 Datensätze) 449
D.20 Prediction-Ergebnisse für FaF MaxAcc (33 Datensätze) 450
D.21 Fitting-Ergebnisse für das FaF ZigSens Plus (33 Datensätze) 451
D.22 Prediction-Ergebnisse für FaF ZigSens Plus (33 Datensätze) 452
D.23 Fitting-Ergebnisse für das FaF ZigSens Majority (33 Datensätze) 453
D.24 Prediction-Ergebnisse für FaF ZigSens Majority (33 Datensätze) 454
D.25 Fitting-Ergebnisse für das FaF ZigSens Fold (33 Datensätze) 455
D.26 Prediction-Ergebnisse für FaF ZigSens Fold (33 Datensätze) 456
D.27 Fitting-Ergebnisse für das FaF ZigVal Plus (33 Datensätze) 457
D.28 Prediction-Ergebnisse für FaF ZigVal Plus (33 Datensätze) 458
D.29 Fitting-Ergebnisse für das FaF ZigVal Majority (33 Datensätze) 459

D.30 Prediction-Ergebnisse für FaF ZigVal Majority (33 Datensätze) 460
D.31 Fitting-Ergebnisse für das FaF ZigVal Fold (33 Datensätze) 461
D.32 Prediction-Ergebnisse für FaF ZigVal Fold (33 Datensätze) 462

Listings

3.1 Pseudocode für die Generationenschleife mit drei Grenzwerten 82
3.2 Pseudocode für die Mutation 84
3.3 Pseudocode Single Point Cross-Over 86
3.4 Zeitschrittalgorithmus für den Partikelschwarm in Pseudocode 103
6.1 Matlab-Code für CART 210
6.2 Matlab-Code für die logistische Regression 211
6.3 Script-Code für die genetischen Algorithmen: An die Stelle von M wird jeweils die Mindestfallzahl eingesetzt 230
6.4 Script-Code für die Schwarmalgorithmen: An die Stelle von M wird jeweils die Mindestfallzahl eingesetzt 231
B.1 Pseudocode für die Reihenfolgemutation 342
B.2 Pseudo-Code für den Inselalgorithmus 347
C.1 Beispiel-Prototyp in $\text{GIER}_{[1.0]}$.script 360
C.2 Beispiel-Script in $\text{GIER}_{[1.0]}$.script 361
C.3 Beispiel-Meta-Script in $\text{GIER}_{[1.0]}$.script 361
C.4 Beispiel-Prototyp für $\text{GIER}_{[1.0]}$.r2-f2 in $\text{GIER}_{[1.0]}$.script 418

Wer das Denken zur Hauptsache macht, der kann es darin zwar weit bringen, aber er hat doch eben den Boden mit dem Wasser vertauscht, und einmal wird er ersaufen.

(Hesse, 1971, §278)

1 Einführung in Thema und Struktur der Arbeit

Im Titel dieser Arbeit treffen Begriffe aufeinander, die einzeln eine Vielzahl von Assoziationen in sich bergen und deren Zusammentreffen zunächst verwundern mag. Es soll zunächst das Ausgangsproblem dieser Arbeit skizziert werden, um den Stellenwert der Begriffe im Anschluss einrahmen zu können.

1.1 Klassifikation und Regelinduktion

Klassifikationen stellen einen allgegenwärtigen Teil des menschlichen Lebens dar und es ist nicht leicht, den ideellen und materiellen Auswirkungen von vollzogenen Klassifizierungen zu entkommen (Bowker & Star, 1999). Der Einfluss von Klassifikationen und Kategorien ist unvermeidbar, auch wenn er nicht in jedem Fall offensichtlich zu Tage tritt. Für Bowker und Star (1999) sind manche menschlichen Klassifikationen zu festen Bestandteilen menschlicher Umgebungen geworden wie die Standards der Wandfarben und Möbelstoffe, die Arten von elektrischen Kabeln.

Die Zielsetzung dieser Arbeit sieht die Konstruktion eines Verfahrens vor, das der Erstellung hierarchischer Regelmengen für binäre Klassifikationen dienen soll: Aufgrund der beobachteten Ausprägung von Prädiktorvariablen sollen Objekte durch dieses Verfahren in eine von zwei Klassen eingeteilt werden mit dem Ziel einer - am Kriterium gemessenen - möglichst fehlerfreien Aufteilung. Eine spezifische Form der Klassifikation, die hier untersucht werden soll, liegt im Bereich des *Internet-Recruiting*: es soll eine Software entwickelt werden, die es ermöglicht, Bewerber aufgrund von Testergebnissen zu klassifizieren. Für die Bearbeitung der Aufgabe werden zunächst strukturell einfache Verfahren aus dem Bereich des Künstlichen Lebens - insbesondere Genetische Algorithmen und Partikelschwarmoptimierung - an die Fragestellung angepasst und für den Diagnostiker nutzbar gemacht. Dies beinhaltet die Entwicklung einer umfangreichen Software. Die entwickelten Algorithmen werden beispielhaft auf Datensätze aus dem Anwendungsbereich angewandt und mit Standardverfahren verglichen. Die Komponenten, die dem Bereich des Künstlichen Lebens entnommen wurden, qualifizieren sich als einfache Heuristiken, die allerdings bisher nicht - mit wenigen Ausnahmen (vgl. Montel, 2005) - zum Gegenstand psychologischer Forschung geworden sind. Personalentscheidungen sind Entscheidungen, die auf Inferenzen beruhen (Landy & Shankster, 1994b). Klassische Verfahren der Kriteriumsschätzung nutzen kompensatorische Verrechnungsstrategien, die nicht in jedem Anwendungsfall indiziert erscheinen (Kubinger, 1999). In der vorliegenden Arbeit wird eine nicht-kompensatorische Form der Klassifikation durch hierarchische Regelmengen betrachtet. Dementsprechend kann die methodische Aufgabe auch als Regelinduktion bezeichnet werden.

1.2 Rationalität und Heuristiken

„Rationalität“ ist ein schillernder Begriff, dessen Gebrauch im Laufe der Zeit gravierenden Veränderungen unterworfen war und dessen Bestimmung auch synchron nicht durch Konsens geprägt ist:

> Die Formen der Rationalität sind Arten einer Gattung, sie lassen sich unterscheiden, aber sind doch erkennbar miteinander verwandt. Was sie gemeinsam haben, ist ein normativer Appell an die Vernunft, und die Kraft dieses normativen Appells hängt von einer breiten Palette an Stützen ab, die ihn plausibel machen - intuitiven, diskursiven, autoritativen, empirischen, experimentellen, technischen. (Daston, 2001, S. 13)

Die Vielfalt der Formen und Begründungen wird begleitet von einer Vielfalt der Forschungsansätze, die auch mit der Breite der Phänomene zusammenhängt, die in Verbindung mit dem Begriff diskutiert werden:

> Visions of rationality do not respect disciplinary boundaries. Economics, psychology, animal biology, artificial intelligence, anthropology and philosophy struggle with models of sound judgment, inference, and decision making. These models evolve over time, just as the idea of rationality has a history, a present, and a future [...]. (Gigerenzer & Selten, 2001, S.1)

Damit stehen die Zeichen schlecht für eine monodisziplinäre Analyse oder den Versuch, Rationalität als Ganzes mit irgendeiner Vorstellung von Vollständigkeit zum Gegenstand der Bearbeitung zu erklären. Rationalität wird in dieser Arbeit an einem praktischen Beispiel und im Hinblick auf Theorien der psychologischen Forschung diskutiert, in einer fundierten kritischen Aufarbeitung der Beziehungen würden allerdings massive Exkurse über die Grenzen von Nachbardisziplinen hinaus notwendig sein. Dementsprechend muss der Anspruch schon zu Beginn zurückgenommen werden, eine definitive Begriffsklärung leisten zu können.

Innerhalb der Psychologie soll ein Forschungsprogramm näher beleuchtet werden, das sich zentral mit Aspekten menschlicher Rationalität befasst und bei dem Heuristiken im Zentrum des Interesses stehen. Es grenzt sich ab von einem älteren Programm, dem Forschungsprogramm *Heuristics and Biases*, das maßgeblich von Daniel Kahneman und Amos Tversky initiiert und geprägt wurde (Kahneman, Slovic & Tversky, 1982; Kahneman & Tversky, 2000). Das neuere Programm formierte sich in Berlin am Max-Planck-Institut für Bildungsforschung mit der Gruppe *Adaptive Behavior and Cognition (ABC)* unter der Leitung von Gerd Gigerenzer und soll als das Programm der Adaptiven Werkzeugkiste bezeichnet werden. Zentral für das Verständnis der Ziele dieses Programms ist die Berücksichtigung des durch Herbert Simon populär gemachten Konzepts der begrenzten Rationalität[1] und des von der Gruppe selbst entwickelten Begriffs der ökologischen Rationalität.

Beide genannten Forschungsprogramme beschäftigen sich mit der Auswirkung der Verwendung von Heuristiken auf jenes elusive Phänomen, das als „menschliche Rationalität“ bezeichnet wird. Das ältere Programm kommt in der Analyse zu einem pessimistischeren Ergebnis und fokussiert auf Fehler und Verzerrungen, die durch den *fehlgeleiteten* Gebrauch einfacher Heuristiken entstehen. Menschliche Rationalität wird stets begleitet gesehen von menschlicher Irrationalität. Das neuere Programm fokussiert auf den Aspekt der Ermöglichung und erfolgreichen Problemlösung durch den *adaptiven* Gebrauch einfacher Heuristiken. Die beiden Programme benutzen allerdings sehr unterschiedliche Begriffe, so dass kein Konsens darüber besteht, was als Heuristik und was als Rationalität bezeichnet werden soll und welche Normen sich als geeignete Normen für die Einschätzung menschlicher Rationalität qualifizieren können. Dieser Disput soll ebenfalls näher betrachtet werden.

Weiterhin soll eine Gruppe von Klassifikationsheuristiken des ABC-Programms, die so genannten *Fast and Frugal Trees*, der Software hinzugefügt und in einer extensiven Simulation an einer Bandbreite von Datensätzen mit klassischen Algorithmen verglichen werden.

[1] Eine frühe Verwendung von „bounded“ in einem verwandten Sinne ist bei John Locke zu finden:

> [...] men, extending their enquiries beyond their capacities, and letting their thoughts wander into those depths where they can find no sure footing, it is no wonder that they raise questions and multiply disputes which, never coming to any clear resolution, are proper only to continue and increase their doubts, and to confirm them at least in perfect scepticism. Whereas, were the capacities of our understanding well considered, the extent of our knowledge once discovered, and the horizon found which sets the ***bounds*** between the enlightened and dark parts of things, between what is and what is not comprehensible by us, men would perhaps with less scruple acquiesce in the avowed ignorance of the one, and employ their thoughts and discourse with more satisfaction in the other. (Locke, 1690/1998, S. 16-17, Hervorhebung hinzugefügt)

1.3 Paradoxien und Heuristiken

Mit Kannetzky (2000) können Paradoxien in folgender Weise über ihren Effekt auf den Beobachter charakterisiert werden:

> Paradoxien sind in einer ersten Näherung ganz allgemein solche Aussagen (bzw. die entsprechenden Sachverhalte), die einer allgemein akzeptierten Meinung entgegenstehen bzw. widersprechen und unerwartet, überraschend, unglaubwürdig oder seltsam sind, weil sie, wenigstens fürs erste und bezogen auf diese Meinung, widersinnig, absurd oder sinnlos erscheinen. (Kannetzky, 2000, S. 13)

Zu den Paradoxien gehört eine sehr heterogene Gruppe von Aussagen und Problemen, die unterschiedlich schwer aufzulösen oder zu beantworten sind, darunter Olbers Paradox, Zenons Paradox, Haufenparadoxien, aber auch Russells Antinomie zu Freges Grundlagen der Arithmetik, die Lügnerantinomie und das Gefangenendilemma. Dabei sollten Paradoxien nicht als Belästigungen, Belanglosigkeiten oder künstlich erzeugte Pseudoprobleme abgetan werden:

> Paradoxien sind deshalb Glücksfälle, weil sie uns darauf aufmerksam machen, daß wir Meinungen anhängen, die falsch sind, auch wenn sie uns selbstverständlich scheinen. Aber zugleich sind sie Ärgernisse, denn sie verraten uns nicht, auf welchem Holzweg wir uns gerade befinden. (Kannetzky, 2000, S. 13)

Paradoxien und ihre Erklärungen verhalten sich dabei wie Feuer und Wasser, eine Erklärung höhlt die Paradoxie aus, bis nichts mehr von ihr übrig ist, solange die Erklärung überzeugend ist. Allerdings sperren sich Paradoxien gegen eine derartige Erklärung, und in vielen Fällen begegnen Erklärungen den Paradoxien aus einer schiefen Richtung, was einen Moment der Unbeholfenheit und Unschicklichkeit offenbaren kann:

> Im gleichen Maße, wie die Voraussetzungen eines Paradoxes erläutert und geklärt werden, verschwindet das Paradox. Ein hinreichend klar dargestelltes Paradox ist kein Paradox. Es ist so, wie wenn man einen Witz erklären muß: Sobald man dies tut, bleibt nichts vom Witz, weil dieser auf bestimmten, wenigstens scheinbar selbstverständlichen, aber irreführenden, oder auch auf unbewußten Voraussetzungen beruht. (Kannetzky, 2000, S. 21)

Paradoxien kennzeichnet vielleicht die Gemeinsamkeit einer „Überdehnung von Abstraktionen, Idealisierungen, Analogien und Modellvorstellungen über den Bereich ihrer sinnvollen Verwendung hinaus“ (Kannetzky, 2000). Kannetzky sieht eine heuristische Funktion in der Beschäftigung mit Paradoxien, da (ungültige) Selbstverständlichkeiten aufgedeckt werden können. Der Titel der Arbeit lautet aber nicht „Die heuristische Rationalität einfacher Paradoxien“, sondern „Die paradoxe Rationalität einfacher Heuristiken“. Angedeutet ist hier zweierlei: Im Sinne des ABC-Programms wird zum einen einfachen Heuristiken die Eigenschaft der Rationalität eingeräumt[2], gleichzeitig wird aber auf den Umstand hingewiesen, dass diese Eigenschaft der üblichen Auffassung der Zuschreibungsbedingungen für Rationalität in entscheidender Hinsicht widerspricht. In dieser Arbeit wird die Rationalität des Heuristikgebrauchs explizit zugestanden; diese Einräumung soll aber nicht voraussetzungsfrei erfolgen. Die Betrachtung wird sich als in weiten Zügen kompatibel mit der Betrachtung des ABC-Programms und in gleicher Weise inkompatibel mit der Betrachtung des *Heuristics and Biases*-Programms erweisen, sich aber stellenweise kritisch von beiden Positionen absetzen.

1.4 Methodische Innovationen

Ein Schwerpunkt dieser Arbeit ist sicherlich die Überprüfung neu entwickelter Methoden auf ihre Tauglichkeit für die psychologische Forschung. Dabei soll die Problematik der Bereitstellung neuer Werkzeuge nicht verkannt werden (vgl. Gigerenzer & Goldstein, 1996a; Gigerenzer, 1991). Insbesondere droht bei der Beschäftigung mit neuen Methoden jenes Phänomen, das Quine (1981) als Methodolatrie - als die Verehrung von Methoden - bezeichnet:

[2] Dabei soll bereits angemerkt werden, dass dieser Sprachgebrauch vom Verfasser als elliptisch verstanden wird: Rationalität kommt dementsprechend entweder dem Anwender oder den Handlungen oder Überzeugungen zu, die aus der Anwendung von Heuristiken resultieren.

> There is methodolatry, or the love of gadgetry: the tendency to take more satisfaction in methods than in the results. Also there is the repose, the respite from hard thought and hairy decisions, that a smooth algorithm can bring. In these ways one may be lured into problems that lend themselves to favorable techniques, though they not be the problems most central to one's concerns. The rise of the computer aggravates this danger. (Quine, 1981, S. 153 f.)

In der Konsequenz droht hier die Vereinnahmung durch das Werkzeug, das seine eigenen Ziele findet:

> [...] as methods and techniques get more complicated, the role of theory in research is being dangerously ignored in favor of purely empirical work that proceeds without so much as a hypothesis. Like Pirandello's characters in search of an author, many of today's researchers seem to have an assortment of techniques in search of a substantive problem. (Einhorn, 1972, S. 367)

Hier stellt die gewählte Unterstützung durch neue Technologien keinen nennenswerten Vorteil dar, im Gegenteil, die Beschleunigung der Bearbeitung mag den Blick auf die inhärenten Schwierigkeiten verstellen:

> Just as the ancient alchemists were not successful in turning base metal into gold, the modern researcher cannot rely on the „computer“ to turn his data into meaningful and valuable scientific information. (Einhorn, 1972, S. 378)

Eine unkritisch übernommene Methode verlässt aber den Bereich einer Entscheidungsfindung, die noch als rational bezeichnet werden könnte:

> Institutionalization of new varieties of reason of course often leads to fake rationality. People use the newest statistical procedures when they have not yet learned to manage them. (Stinchcombe, 1990, S. 310)

Schimank (2005) bezeichnet Rationalitätsfassaden als die Darstellung des eigenen Handelns in einem Sinne, der den Rationalitätserwartungen der Bezugsakteure entspricht. Schimank selbst betrachtet viele Personalentscheidungen als ein Beispiel, weil es für ihn in vielen Fällen „ehrlicher“ wäre, zwischen verbleibenden Bewerbern zu losen, als ein aufwändiges, aber unkontrolliertes, vorgeblich „rationales“ Verfahren der Auswahl zu bemühen. Diese Überlegungen haben den Zweifel genährt, dass es unvollständig erscheint, Verfahren zu produzieren, ohne deren Einbettung in ein greifbarer gewordenes Konzept von Rationalität zu leisten und den Anwendungskontext einer Methode zu prüfen.

1.5 Die Struktur der Arbeit

An dieser Stelle soll die Gliederung der folgenden Kapitel vorgestellt und damit gleichzeitig eine Skizze des zu erwartenden Inhalts gegeben werden. Das zweite Kapitel wendet sich zunächst einer Klärung des Begriffes *Data Mining* zu **(Abschnitt 2.1)**, da viele der vorgestellten Verfahren diesem Bereich zugeordnet werden können, der in einem dynamischen Spannungsverhältnis zur klassischen Statistik steht. Dabei wird auch die Beziehung zwischen Heuristiken und Algorithmen erläutert. Die Relevanz des Ausgangsproblems für die psychologische Diagnostik und insbesondere das Internet Recruiting im Rahmen der DIN 33430 wird erläutert **(Abschnitt 2.2)** - mit einigen Beispielen für die Implementierung von Entscheidungsregeln in computerbasierten Testverfahren - bevor das Regelinduktionsproblem schrittweise statistisch präzise formuliert wird. Dies schließt die Wahl eines Kriteriums zur Bestimmung der Klassifikationsgüte ein sowie eine Diskussion des *Overfitting*-Problems **(Abschnitt 2.3)**. Im Folgenden wird eine Möglichkeit der Visualisierung von Klassifikationsergebnissen entwickelt **(Abschnitt 2.4)** - das ROCS-Diagramm -, auf die im Verlauf der Arbeit zurückgegriffen werden soll. Sie wird abschließend ebenfalls genutzt, um die Ergebnisse einer Untersuchung von Montel (2005) vorzustellen, die sich teilweise auf eine ähnliche Grundfragestellung beziehen und an Datensätzen erbracht worden sind, die auch in dieser Arbeit verwendet werden **(Abschnitt 2.5)**.

Das dritte Kapitel ist der Darstellung von Heuristiken des künstlichen Lebens gewidmet; es werden genetische Algorithmen (GA) und die Partikelschwarmoptimierung vorgestellt. Zunächst wird die Grundform

des GA erläutert **(Abschnitt 3.1)**, im Anschluss werden der Reihe nach die Komponenten von genetischen Algorithmen diskutiert, was Repräsentationsformen, die Erstellung von Anfangspopulationen, die Wahl von Fitnessfunktion und Selektionsprinzip sowie Methoden der Fitnessskalierung einschließt **(Abschnitt 3.2)**. Ein gesonderter Abschnitt ist den genetischen Mutations- und Rekombinationsoperatoren gewidmet **(Abschnitt 3.3)**. Die Algorithmenklasse der genetischen Algorithmen wird an die Ausgangsfragestellung der Arbeit adaptiert, so dass die Struktur von Algorithmen beschrieben werden kann, die zur Lösung des Problems eingesetzt werden können **(Abschnitt 3.4)**. Zudem wird dem Problem von eingeschränkten Lösungsräumen, Belohnungs- und Reparaturmechanismen ein eigener Abschnitt gewidmet **(Abschnitt 3.5)**. Weiterhin wird die Klasse der Schwarmalgorithmen vorgestellt am Beispiel der Partikelschwarmoptimierung (PSO). Dabei werden der Algorithmus mit seinen Komponenten präsentiert, die Struktur der Partikel und der Kommunikationstopologie von Schwärmen sowie die Regeln für ihre Bewegung und Beschleunigung beschrieben, zudem die Schwarmevaluation und das Abbruchkriterium erläutert und eine Pseudocode-Formulierung gegeben **(Abschnitt 3.6)**. Das Kapitel wird abgeschlossen durch einen Überblick über bisherige Einsätze genetischer Algorithmen in einer großen Bandbreite an Anwendungsbereichen und einen Anwendungsvergleich, in dem sowohl die Anpassung von Partikelschwärmen an verschiedene Aufgabenstrukturen erläutert als auch das Problem der Parameterwahlen exemplifiziert werden kann **(Abschnitt 3.7)**.

Das vierte Kapitel widmet sich den Grenzen menschlicher Rationalität und dem Forschungsprogramm der adaptiven Werkzeugkiste. Dazu werden zunächst einige basale begriffliche Unterscheidungen dargestellt, die für verschiedene Auffassungen von Rationalität charakteristisch sind, und es wird die Trennung zwischen Rationalität und Irrationalität diskutiert. Es folgt eine Darstellung der Auseinandersetzung zwischen den beiden psychologischen Forschungsprogrammen zu Heuristiken, und die normative Rolle von Logik und Wahrscheinlichkeitstheorie werden hinterfragt **(Abschnitt 4.1)**. Das von Herbert A. Simon entwickelte Konzept der *Bounded Rationality* wird präsentiert und in den Verbindungen zu Heuristiken beleuchtet, zudem wird die Rolle der äußeren Umgebung rekonstruiert **(Abschnitt 4.2)**. Es folgt eine Darstellung des Konzeptes der ökologischen Rationalität, das in Bezugnahme auf Simon im Forschungsprogramm der ABC-Gruppe entwickelt wurde. Diese Darstellung beinhaltet auch einen Überblick über die Metapher der Adaptiven Werkzeugkiste **(Abschnitt 4.3)**. Das Programm der ABC-Gruppe wird an einigen Beispielen vorgestellt. Zunächst wird die Rekognitionsheuristik präsentiert, zusammen mit einigen analytischen und empirischen Ergebnissen **(Abschnitt 4.4)**. Im Anschluss wird die *Take-The-Best*-Heuristik diskutiert, dazu wird der Algorithmus präsentiert mit Beispielen von Simulationsstudien und empirischen Ergebnissen **(Abschnitt 4.5)**.

Im fünften Kapitel wird eine Klasse einfacher Klassifikationsheuristiken dargestellt und weiterentwickelt, die ebenfalls im ABC-Programm eingeführt wurden, die *Fast and Frugal Trees*. Diese Bäume werden zunächst motiviert und in ihrer Struktur erläutert **(Abschnitt 5.1)**. Die Eigenschaften des Klassifikationsmodell werden betrachtet und im Hinblick auf Anwendungsmöglichkeiten des Modells bewertet **(Abschnitt 5.2)**. Es wird ein Überblick über eine Auswahl empirischer Studien gegeben, die Modelle zum Einsatz bringen, die sich als *Fast and Frugal Trees* beschreiben lassen **(Abschnitt 5.3)**. Die Frage nach geeigneten Konstruktionsregeln wird thematisiert und erste Lösungen werden weiterentwickelt **(Abschnitt 5.4)**. Zuletzt werden einige analytische Ergebnisse zu diesem Klassifikationsverfahren berichtet **(Abschnitt 5.5)**.

Das sechste Kapitel berichtet über eine Reihe von Simulationsstudien und stellt eine Sammlung von Ergebnissen vor. Im ersten Abschnitt werden die Ergebnisse des $\textsc{Gier}_{[1.0]}$.downhill-Algorithmus für die Datensätze *Hochschule* und *Beratungsunternehmen* präsentiert, zusammen mit einer Analyse der Ergebnisqualität **(Abschnitt 6.1)**. In einer umfangreichen Simulationsstudie an insgesamt dreiunddreißig Datensätzen wird weiterhin die Algorithmenklasse der *Fast and Frugal Trees* analysiert **(Abschnitt 6.2)**. Im dritten Abschnitt werden die gleichen Datensätze hinsichtlich einer Überprüfung der Leistung der Datensätze $\textsc{Gier}_{[1.0]}$.ga und $\textsc{Gier}_{[1.0]}$.swarm genutzt **(Abschnitt 6.3)**. Zuletzt werden die Ergebnisse diskutiert, und weitere Entwicklungsmöglichkeiten werden skizziert **(Abschnitt 6.4)**.

Das siebte Kapitel wird durch eine auf den ABC-Ansatz zugeschnittene Diskussion des *Frame*-Problems begonnen. Forsters Hypothese des Komplexitätserhaltungssatzes wird von der Diskussion der vom *Frame*-Problem affligierten Werkzeugauswahl gefolgt **(Abschnitt 7.1)**. Um die Grenzen der in dieser Arbeit unternommenen Problemlösung fassen zu können, wird eine Modellierung der Aufgabenstellung als q-Morphismus vorgeschlagen, die als Grundlage genommen wird, um das bearbeitete Problem von verwandten Fragen und unbearbeiteten Aspekten abzugrenzen **(Abschnitt 7.2)**. Im letzten Abschnitt wird

Tabelle 1.1: Module und Codezeilen der Begleitsoftware

Modulname	integriert in	Codezeilen
GIER$_{[1.0]}$		ca. 62 000
GIER$_{[1.0]}$.life	GIER$_{[1.0]}$	
GIER$_{[1.0]}$.TSP	GIER$_{[1.0]}$	
GIER$_{[1.0]}$.ACO	GIER$_{[1.0]}$	
GIER$_{[1.0]}$.simple_heuristics		ca. 66 250
GIER$_{[1.0]}$.boid		1558
GIER$_{[1.0]}$.norms		939
GIER$_{[1.0]}$.ROCS		268
GIER$_{[1.0]}$.swarmtest		1161
Gesamtcode		ca. 132 000

die epistemische Bedeutsamkeit von Heuristiken für die Entscheidungsfindung diskutiert. Das Verhältnis von Einfachheit und Komplexität und damit die paradoxe Natur der Rationalität einfacher Heuristiken wird zum Thema gemacht, und es wird über die Rolle von Unsicherheit reflektiert **(Abschnitt 7.3)**.

1.6 Programmierte Softwaremodule

Die Begleitsoftware zur Arbeit besteht aus mehreren Paketen. Im Anhang ausführlich dargestellt wird die Entwicklung und Struktur der Software GIER$_{[1.0]}$. Mit dieser Software ist die Erstellung hierarchischer Klassifikationsregelmengen möglich unter Verwendung verschiedener Algorithmen des Künstlichen Lebens und gieriger Algorithmen. Gleichzeitig sind Datenvorbereitungs- und Visualisierungsroutinen enthalten. Das zweite Paket - GIER$_{[1.0]}$.simple_heuristics - ist ein exploratives Paket in dem Sinne, dass eine Menge von Werkzeugen zusammengetragen worden ist, um den postulierten Inhalt der Adaptiven Werkzeugkiste - also einige der beschriebenen Heuristiken - in künstlichen und natürlichen Umgebungen zu testen. Die Software enthält Werkzeuge zum Vergleich einer Reihe von Heuristiken für paarweise Vergleichsentscheidungen, Klassifikation und Schätzung[3]. In dieser Arbeit werden Ergebnisse der in diesem Paket enthaltenen *fast and frugal heuristics* präsentiert: Die *Estimation Trees* und *QuickEst*-Algorithmen (Hertwig, Hoffrage & Martignon, 1999) werden in dieser Arbeit nicht weiter ausgeführt, stehen aber für die Exploration zur Verfügung. Gleiches gilt für eine Reihe von Werkzeugen, mit denen Prinzipien und Techniken des Künstlichen Lebens exploriert werden können. Die Software enthält weitere speziellere Analysemodule für die Rekognitionsheuristik und die Take-The-Best-Heuristik, auf die ebenfalls in Berücksichtigung des bereits beträchtlichen Umfangs der Arbeit nicht näher eingegangen werden kann: Es können Entscheidungen der TTB-Heuristik analysiert und visualisiert werden, und eine ausführliche Untersuchung der Robustheit der Heuristik bei Datensatzaufteilungen ist möglich. Eine vereinfachte Version der GIER$_{[1.0]}$.simple_heuristics-Software kam während der *Summer School* 2006 des Max-Planck-Instituts für Bildungsforschung in Berlin zur eigenständigen Exploration von Heuristiken zum Einsatz und wurde bereits für weitere Lehrveranstaltungen angefragt. Tabelle 1.1 zeigt die Reihe der entwickelten Module und die Länge des jeweiligen Programmcodes. Eine Aufnahme des Codes in den Anhang hätte den Umfang der Arbeit um mehrere Tausend Seiten vergrößert. Aus diesem Grund soll in Anhang C nur sehr exemplarisch auf bestimmte Programmierprinzipien hingewiesen werden. Damit ist die Art der Behandlung des Themas dieser Arbeit vorgezeichnet, die eine Bandbreite von inhaltlichen Bereichen und methodischen Ansätzen zu integrieren versucht. Gewünscht ist eine Annäherung an die Schwierigkeiten, aber auch an den Reiz der Frage nach der möglichen Rationalität heuristischer Ansätze des Denkens und Entscheidens.

[3] Für die Umsetzung der linearen Regression in *Object Pascal* wurden zwei Gruppen von Routinen adaptiert, die dem Verfasser von Magnus Persson einerseits und über das Internet durh Fred J. Edberg (*Regcomp)* zur Verfügung gestellt wurden.

Induction, which has been called the „scandal of philosophy“, has become the scandal of psychology and artificial intelligence as well.

(Holland, Holyoak, Nisbett & Thagard, 1986)

2 Ein heuristischer Ansatz zur Regelinduktion

In diesem Kapitel soll zunächst der Bereich des *Data Mining* vorgestellt werden, innerhalb dessen ein Platz für heuristische Methoden gefunden werden kann. Das Problem der Regelinduktion wird mit Blickrichtung auf Anwendungsfragen der psychologischen Diagnostik und den Kontext ihrer Anwendung reflektiert, um dann eine formale Definition des bearbeiteten Problems geben zu können. Eine Methode der Visualisierung von Klassifikationsergebnissen wird vorgestellt und an einem Beispiel zum Einsatz gebracht. Eine Modellierung der Fragestellung ermöglicht abschließend eine klare Abgrenzung des Problems von nicht-bearbeiteten Problemen.

2.1 Data Mining: Der Datensatz als Goldmine

Die in der vorliegenden Arbeit thematisierten einfachen Heuristiken, die im folgenden Abschnitt näher charakterisiert werden sollen, haben eine deutlich stärkere Nähe zur Praxis des *Data Mining* als zur herkömmlichen Statistik. Deshalb erscheint es sinnvoll, die Unterschiede hinsichtlich des üblichen Vorgehens in beiden Disziplinen auszuführen. Anschließend soll die Frage nach dem Unterschied zwischen Heuristiken und Algorithmen beantwortet werden und zuletzt soll das Problem des *Bias* im Kontext des *Data Mining* diskutiert werden.

2.1.1 Data Mining und traditionelle Statistik

„Data Mining“ klingt in den Ohren des Hörers als Beschreibung einer Tätigkeit vielleicht aufregender, abenteuerlicher und auch anstrengender als der Begriff „statistische Analyse“. Insbesondere die Statistik scheint trotz hoher Akzeptanz ihrer Ergebnisse ein Image-Problem aufzuweisen:

> „Statistics 382“, „Introduction to Statistics“, and Religion 334, „Death and Dying“, are both being offered as courses on the World Wide Web, some people think they are equivalent“. [Anonymous on the web] (Unwin, 2000, S. 303)

Data Mining scheint hier eine lebendigere Alternative zu bieten. Die Metapher des Bergwerks kann allerdings in zweierlei Hinsicht verstanden werden, als Bezugnahme auf einen schweißtreibenden Förderprozess oder als Bezugnahme auf den Ertrag der Minenarbeit, also auf die Gewinnung zuvor verborgener Bodenschätze. Insbesondere Statistiker, die dem sich ausbreitenden Feld kritischer gegenüber stehen, neigen zur ersten Perspektive mit negativer Konnotation, während Computerwissenschaftler eher der zweiten Sicht verschrieben scheinen:

> Statisticians have used the term data mining to denote unsavoury 'data dredging' or 'fishing expeditions' in search of publishable P-values. In contrast, the computing community in the past few years has adopted the term data mining (DM) to denote 'the extraction of previously unknown information from databases that may be large, noisy and have missing data' [...]. (Mackinnon & Glick, 1999, S. 255)

Ein Grund für den Fortbestand der Differenz zwischen den beiden Gruppen könnte in der Geschichte der beiden Disziplinen begründet liegen, die das Feld *Data Mining* als - etwas respektlosen -Eindringling in ein etabliertes wissenschaftliches Programm ausweist:

> Many statisticians look askance when they see conference posters advertising meetings on data mining or intelligent data analysis. The descriptions of the aims of these conferences sound suspiciously like the descriptions of the aims of statistics conferences. It seems as if others, who do not call themselves statisticians, are setting up in competition and presenting themselves as experts in data analysis. Worse still, these newcomers (who have often coined exciting new names for their activities, such as „expert systems," „data mining," „neural networks," or „intelligent data analysis" [What does that make statistical data analysis?]) often seem to attract research funding that makes funding awarded to statisticians appear rather thin. As if all this were not enough, the newcomers sometimes appear to have relatively little formal statistical expertise, so that they sometimes make errors that statisticians would avoid as obvious. (Hand, 2000, S. 442)

Manche Kritiker sehen im Data Mining deshalb den Versuch, eine etablierte Disziplin aufzubohren und in trivialisierten Einzelteilen leichtfertig an den Meistbietenden zu verkaufen mit verheerenden Folgen für die ursprüngliche Disziplin: In der Tat ist die Gruppe der „Data Miner" alles andere als disziplinär geschlossen. Zhou (2003) sichtet die Reihen der praktizierenden Mitglieder in folgender Weise:

> Due to its interdisciplinary nature, data mining has received contributions from a lot of disciplines such as databases, machine learning, statistics, information retrieval, data visualization, parallel and distributed computing, etc. The first three in the list, i.e., database, machine learning, and statistics, are undoubtedly the primary contributors. (S. 140)

Eine ähnliche Spannung wie die zwischen Data Mining und Statistik besteht allerdings auch zwischen dem sogenannten *Machine Learning*[1] und der Statistik:

> Was ist der Unterschied zwischen maschinellem Lernen und Statistik? Für Zyniker, die ironisch auf das explosionsartig anwachsende kommerzielle Interesse (und Hochpushen) in diesem Bereich blicken, ist Data Mining gleichbedeutend mit Statistik plus Marketing. (Witten & Frank, 2001, S. 28)

Aus der Aufzählung wird allerdings ebenfalls deutlich, dass Statistik eine, wenn auch nicht die einzige Grundlage für *Data Mining* darstellt. Zhou sieht die drei zentralen Bereiche als komplementär und nicht als auf sich allein gestellt lebensfähig:

> It is obvious that without the powerful data management techniques donated by the database community and the practical data analysis techniques donated by the machine learning community, data mining would be seeking a needle in the haystack. [...] However, it is still clear that without the solid theoretical foundation donated by the statistics community, data mining will be building a castle in the air. (Zhou, 2003, S. 140)

Zu den Anwendungsgebieten gehören Clusterverfahren, Klassifikationsverfahren, Prädiktionsmodelle, Entscheidungshilfen, Datenreduktionsverfahren etc.; die Vielfalt der Aufgaben wird nur von der Vielfalt der Lösungansätze übertroffen. Vor diesem Hintergrund kann die Data Mining-Defintion von Witten und Frank (2001) sinnvoll eingeführt werden:

> Data Mining definiert sich als der Prozess, Muster in Daten zu erkennen. Der Prozess hat automatisch oder (was häufiger der Fall ist) halbautomatisch stattzufinden. Dabei müssen sinnvolle Muster erkannt werden, die zu einem Vorteil in der Regel wirtschaftlicher Art führen. Die Daten stehen immer in ausreichender Menge zur Verfügung. (Witten & Frank, 2001, S. 3)

[1] Machine Learning hat nicht den Anspruch zu erklären, was Lernen ist oder sein kann und will auch nicht explizit biologisches Lernen simulieren, Ziel ist eher das Erlernen von Wissen und die Verarbeitung von Informationen (vgl. Witten & Frank, 2001).

Der letzte Satz verweist auf eine fundamentale Einschränkung der Reichweite von *Data Mining*: Das Erheben von Daten und damit verbunden die Planung von Datenerhebungen, die Durchführung von Beobachtungen oder Experimenten fallen nicht unter das Konzept. Daten sind der Ausgangspunkt für *Data Mining*, ein praktischer Nutzen ist das Ergebnis der Anwendung. Dieser Nutzen beschränkt sich nicht nur auf hilfreiche Vorhersagemöglichkeiten, sondern kann auch die Sammlung und Aufbereitung von Informationen umfassen. Zu einer vollständigen Datenauswertung im wirtschaftlichen Bereich gehören nach Gür Ali und Wallace (1997) die folgenden Schritte:

1. das wirtschaftliche Problem muss verstanden werden,
2. der Datensatz muss ausgewählt und strukturiert werden,
3. angemessene Werkzeuge müssen ausgewählt und appliziert werden,
4. die Ausgabe muss analysiert werden,
5. die Ergebnisse müssen dem Management präsentiert werden.

Dieses Schema kann als ein Spezialfall des Konzepts der *knowledge discovery in databases (KDD)* aufgefasst werden, also eines Prozesses, den Mackinnon und Glick (1999) in folgender Weise strukturiert sehen:

- data selection and sampling;
- pre-processing and cleaning;
- transformation and dimension reduction;
- data mining or analysis;
- visualization and evaluation. (S. 256)

Dabei wird *Data Mining* hauptsächlich im vierten Schritt platziert, auch wenn die Phasen aufeinander abgestimmt sein müssen. Nicht jedes Verfahren liefert in gleicher Weise präsentable Ergebnisse, die Anwendung spezieller Verfahren ist wiederum abhängig von der Qualität des Datensatzes. Es ist für das *Data Mining* aber entscheidend, diesen Prozess nicht als einen reflexähnlichen Automatismus bei Bereitstellung von Daten aufzufassen. Die Lösung eines Problems bzw. die Befriedigung eines differenzierten Erkenntnisinteresses stehen im Vordergrund:

> We are constantly being reminded (and rightly so) that data mining must be more than just discovering knowledge for its own sake. The knowledge must provide insights into a business problem or opportunity and guidelines for action, i.e., support for decision making. (Gür Ali & Wallace, 1997, S. 3)

Damit werden eine Reihe von inhaltlichen Unterschieden zur traditionellen Statistik deutlich. Witten und Frank (2001) weisen darauf hin, dass sich letztere - vereinfacht - stärker mit dem Überprüfen von einzelnen Hypothesen beschäftigt, während im maschinellen Lernen stärker die Suche durch eine Hypothesenmenge im Vordergrund steht. Weiterhin ist der Kanon zulässiger Methoden im *Data Mining* größer, allerdings nicht unbedingt immer zugunsten der Qualität der Anwendungen:

> [...] whereas modern statistics is heavily model-driven, data mining places much greater emphasis on the algorithms. Indeed, I would go so far as to say that there is an overemphasis on algorithms. Recent data mining conference proceedings show a surfeit of new algorithms, with very little critical assessment and comparison. Because the development of these algorithms is in the absence of a unifying theoretical underpinning analogous to that underlying statistical modeling, the discipline is in danger of losing direction. (Hand, 2000, S. 448)

Die Vielfalt der Algorithmen wird potenziert durch den explorativen Fokus der Auswertung. Während die konventionelle hypothesenprüfende Statistik die Anzahl der Modelle zu minimieren sucht, ist die diesbezügliche Praxis im Data Mining sogar gegenläufig:

> [...] anyone with an important data mining problem to be solved is well advised to consider the possibility of solving it with different data mining paradigms, trying to determine the best paradigm and the best particular algorithm for the specific target problem. (Freitas, 2002, S. 6)

Vor- und Nachteile der beiden Strategien sind offenkundig: Die schmale Analyse führt zu einer Aussage, die durch eine gut untersuchte und begründbare Methode gewonnen wurde und für deren Interpretation es Erfahrungswerte gibt. Die breitere Strategie hingegen ermöglicht das Auffinden ungewöhnlicherer Zusammenhänge, allerdings nicht unbedingt immer inhaltlich substanzieller Zusammenhänge. Die Inflation der Verfahren und Anwendungen lässt anraten, das Vertrauen in die gewonnenen Ergebnisse zu senken, ermöglicht aber das Gewinnen von Ergebnissen, die durch die Maschen des statistischen Standardnetzes gleiten würden. Diese Ergebnisse ermöglichen Handlungsgrundlagen, wenn auch nicht notwendigerweise die bestmöglichen:

> [...] data mining lies at the interface between statistics and computer science. Implicit in the breadth of its parentage is a willingness, on the part of its practitioners, to try approaches that deviate from the [...] statistical norm. Moreover, they will be willing to confront problems that statisticians, bound by the constraints of their relatively formal perspectives on what can and cannot be done, would not try to tackle. The solutions reached by the less formal approaches may or may not be very good solutions. But often solutions of some kind are reached, and some solution, albeit imperfect, is often better than none: In a commercial environment, for example, action must be taken (inaction is not an absence of action but merely a particular choice of action) so that one must do the best one can, even if it may not be ideal. (Hand, 2000, S. 443)

Für Disziplinen, die Ergebnisse statistischer Forschung lediglich nutzen bzw. zu nutzen glauben, muss zudem angeführt werden, dass die Anwendungsindikationen für bestimmte Methoden oft unzureichend geprüft werden bzw. die Passung von Methode und Problem nicht zufriedenstellend begründet wird. Freedman (1991) kritisiert etwa die ubiquitäre Anwendung von Regressionsmodellen in den Sozialwissenschaften:

> Regression modeling is a dominant paradigm, and many investigators seem to consider that any piece of empirical research has to be equivalent to a regression model. Questioning the value of regression is then tantamount to denying the value of data. (S. 292)

Die imaginierte Verfügbarkeit einer Methode zur Beantwortung einer großen Bandbreite an Fragen kann nach Freedman zu einer Selbstüberschätzung der Disziplin führen, alle gesellschaftlichen Fragen prinzipiell erfolgreich in Angriff nehmen zu können; ganz abgesehen davon, dass der Schritt der Variablenauswahl in der Modellierung oft nicht den notwendigen Respekt erhält. Die Tests setzen neben einer adäquaten Stichprobengewinnung die korrekte Variablenauswahl voraus, um als valide interpretiert werden zu können. Die Problematisierung des psychologischen Gebrauchs bzw. Missbrauchs statistischer Signifikanztestmethoden in einer hybriden Mischform der statistischen Ansätze von Fischer sowie von Neyman und Pearson (vgl. Gigerenzer et al., 1989) ist ein weiteres Beispiel für das beschriebene Phänomen (s.a. Falk & Greenbaum, 1995; Sedlmeier, 1996; Hubbard, Parsa & Luthy, 1997; Sohn, 1998; Sedlmeier, 1998; Haller & Krauss, 2002). Mit einer solchen Abweichung gehen allerdings die genannten Vorteile statistischer Methoden gegenüber *Data Mining*-Ansätzen vollständig verloren. Aufgrund der stärkeren Konzentration des *Data Mining* auf Exploration und Deskription zugunsten von Modellierung und Inferenz (Hand, 2000) kann es als gute Alternative auch für wissenschaftliche Studien in der Phase der Exploration eines Forschungsbereichs Stärken zeigen. Diese liegen vor allem in der Anschaulichkeit von Ergebnissen:

> Data miners are often more interested in understandability than accuracy or predictability per se. Thus, there is a focus on relatively simple interpretable models involving rules, trees, graphs, and so forth. (Glymour, Madigan, Pregibon & Smyth, 1997, S. 11 f.)

Genau dies ist der Punkt, warum Heuristiken einen festen Platz im *Data Mining* beanspruchen können, wie der folgende Abschnitt zeigen wird. Der Prozess des *Data Mining* kann zudem als iterativ charakterisiert werden, der wiederholte Rückgriff auf den Datensatz nach Begutachtung gefundener Muster durch Experten mit Bereichswissen kann nicht durch eine Standardlösung mit garantiertem Erfolg ersetzt werden, Versprechungen dieser Art müssen als unseriös (dis-)qualifiziert werden (Hand, 2000). Insbesondere eine vollständige Automatisierung ist hier in vielen Fällen nicht denkbar:

> [...] human judgment is essential for many non-trivial inference problems. Thus, automation can at best only partially guide the data analysis process. Properly defining the goals of an analysis remains a human-centred, and often difficult, process. (Glymour et al., 1997, S. 17 f.)

Eine unsachgemäße Anwendung von *Data Mining*-Verfahren ist als ebenso schädlich einzustufen wie der Missbrauch traditioneller Statistik. Die Vielzahl der entwickelten Modelle und Methoden ist für Qualitätskontrolle und Selbstreflexion der Disziplin keine günstige Voraussetzung. Die Betonung der Auswertungsmethode kann zudem zu einem Rückgang in der Sorgfalt der Sicherstellung der am Ausgangspunkt stehenden Datenqualität führen, was den Erfolg der Unternehmung ernsthaft in Frage stellen würde:

> I would not be at all surprised to see a backlash against data mining as it fails to provide the (media-hyped) results. Many of the results will, on closer examination, turn out to be due to data anomalies and peculiarities and to be of no intrinsic interest or value. (Hand, 2000, S. 448 f.)

2.1.2 Heuristiken und Algorithmen

Die Abgrenzung von Heuristiken und Algorithmen[2] ist nicht unbedingt eindeutig, in vielen Fällen wird der Begriff „Algorithmus“ auch für streng-genommen heuristische Verfahren gebraucht, wenn auch die umgekehrte Domänenerweiterung selten ist.

Ein Algorithmus kann definiert werden als „allgemeines Verfahren, das die Ausführung eines Prozesses eindeutig regelt und das in einem endlichen Text angegeben werden kann“ (Tack, 1991, S. 264). Der Klasse von heuristischen Lösungsansätzen oder *Heuristiken* fehlt eine derartig eindeutige Definition. Das Wort selbst lässt sich auf einen altgriechischen Stamm zurückverfolgen:

> *Heuristic* comes from the Greek *heuriskein*, meaning „to find“, hence *heureka*, meaning „I found it (out)“. Since its introduction to English in the early 1800s, the term has acquired a range of meanings. (Hoffrage & Reimer, 2004, S. 439)

Eine der Bedeutungen, die Heuristik annimmt, ist die Verwendung durch Einstein als fehlerbehaftete Methode:

> As used by Einstein [...] a heuristic is an approach to a problem that is necessarily incomplete given the knowledge available, and hence unavoidably false, but which is useful nonetheless for guiding thinking in appropriate directions [...]. (Gigerenzer & Todd, 1999, S. 26)

Die negative Konnotation des Unzureichenden wird im Verlauf der kognitiven Wende zurückgenommen:

> With the advent of information theory in cognitive psychology, the term heuristic finally came to mean a useful shortcut, an approximation, or a rule of thumb for searching through a space of possible solutions. (Hoffrage & Reimer, 2004, S. 439)

Die Verbform *heurisko* [3] lässt sich übersetzen als „ich finde das, was ich suche“ und als „ich erforsche“, aber sie hat auch die Bedeutung von „ich finde zufällig“ bzw. „ich treffe an“. Dieses Element der Zufälligkeit ist für eine nähere Bestimmung der Wortbedeutung entscheidend:

[2] Nach Groner und Groner (1991) entspringt der Name „Algorithmus“ einem historischen Missverständnis: Mohammed Ben Musa al Khovarezmi, der um 830 am Hofe des Kalifen Mammum in Baghdad lebte hat demnach ein Buch über die von Indern entwickelten Regeln der Arithmetik verfaßt. Dieses Buch wurde im Mittelalter in der lateinischen Übersetzung mit *algoritmi de numero indorum* betitelt, wobei „algoritmi“ den Namen des Autoren verschriftlichen sollte, später aber als Terminus technicus missverstanden wurde. Eine orthographische Normalisierung verwandelte es dann letztendlich zu dem pseudogriechischem wort „algorithmos“ ohne altsprachlichen Stamm.

[3] $\left(\acute{\epsilon}\upsilon\rho\acute{\iota}\sigma\kappa\omega\right)$

> Ihr hervorstechendes negatives Definitionsmerkmal ist das Fehlen der Lösungsgarantie: Wir wissen bei einem bestimmten Problem nicht im voraus, ob eine bestimmte Heuristik zum Erfolg führen wird oder nicht. Aus diesem Grunde ist auch die Klasse von Problemen, auf die eine Heuristik angewendet werden kann, prinzipiell offen. (Groner & Groner, 1991, S. 318)

Simon (1967) macht das Merkmal der Lösungsgewissheit bzw. Ungewissheit zum zentralen Unterscheidungskriterium. Zunächst definiert er das Konzept der Problemlösungswahrscheinlichkeit:

> Consider, next, some population of problems and a problem-solving program S. We define $p(S,t)$ as the probablity that a problem, drawn at random from the domain, will be solved by the program, S, in t minutes or less. The function $p(S,t)$ wil then be a non-decreasing function of t.

Mit diesem Konzept führt er nun die Unterscheidung zwischen algorithmischen und nicht-algorithmischen Programmen an:

> ... we might call a program algorithmic if, for some t_1, $p(S,t_1) = 1$. Or we might call a program asymptotically algorithmic if we can make $p(S,t)$ as close to 1 as we please by making t sufficiently large. The important point is that whether a particular program does or does not possess such algorithmic properties is quite independent of its heuristic power for moderate t. For example, the only algorithm for chess (which consists of considering all possible games and minimaxing backward from the end) would almost never find a move during the lifetime of the universe, much less in a reasonable computing time. (Simon, 1967)

Damit werden auch bereits Gründe genannt, warum eine Methode ohne Gewährleistung der Problemlösung überhaupt als vernünftige Methode konzeptualisiert werden kann, selbst wenn es eine Lösung gibt, die durch einen Algorithmus gefunden werden könnte. Bei einer Beschränkung der Lösungszeit kann ein heuristischer Ansatz der Ansatz mit einer höheren Problemlösungswahrscheinlichkeit sein, da der Punkt, an dem die Lösungswahrscheinlichkeit des Algorithmus die der Heuristik übertrifft zeitlich hinter dem für die Lösung einberaumten Zeitfenster liegen kann. Dies wird im Allgemeinen durch eine selektive Suche erreicht:

> Ein wesentliches Merkmal der meisten Heuristiken besteht auch darin, dass der Suchraum, innerhalb dessen die gesuchte Lösung liegen sollte, drastisch eingeschränkt wird, allerdings auch mit dem Preis, die Lösung selbst zu verlieren. Algorithmen sind wohldefiniert, formal korrekt und erzwingen eine Lösung, falls es die Zeit und Problemgröße erlauben, sozusagen mechanisch. Heuristiken dagegen lassen sich viel weniger gut spezifizieren, sie beinhalten den Mut zum Risiko bei der Abkürzung auf langen und beschwerlichen Lösungswegen, das Eingehen auf die Besonderheiten einer bestimmten Problemsituation, die Nutzung von eigenen Erfahrungen und von Empfehlungen anderer. (Groner & Groner, 1991, S. 318) (Groner & Groner, 1991, S. 318)

Dabei sollte darauf hingewiesen werden, dass die Entwicklung von Algorithmen selbst in vielen Fällen einem nicht-algorithmischen Prozess entspringt, die Anwendung von Algorithmen setzt eine bereits abgeschlossene Entwicklung voraus (Groner & Groner, 1991).

Wird eine Heuristik auf einem Computer umgesetzt, so ist für ihre Anwendung ein Algorithmus notwendig, dem der Computer folgen kann: Auch wenn das Ziel der Suche nicht deterministisch erreicht wird, so wird zumindest das Ende der Suche deterministisch erreicht: Auch aus diesem Grund wird in der Sprache von Heuristiken oft der Begriff des Algorithmus bemüht.

2.1.3 Data Mining und Bias

Dem Problem der Überanpassung von Modellen an die vorliegenden Daten (vgl. Abschnitt 2.3.5) kann unter Umständen durch große Stichproben, die in vielen *Data Mining*-Anwendungen zur Verfügung stehen, teilweise vorgebeugt werden, obwohl die Verwendung einer großen Zahl an Algorithmen mit unter Umständen zahlreichen Parameterkombinationen die Gefahr der Aufdeckung unzuverlässiger Effekte erhöht.

Das *Data Mining* ist allerdings anfällig für eine Reihe weiterer Verzerrungen. Witten und Frank (2001) unterscheiden das Phänomen des Sprach-Bias von dem des Such-Bias: Ein *Sprach-Bias* wird durch jede Konzeptbeschreibungssprache eingeführt, die zu einer differenziellen Beschreibbarkeit plausibler Konzepte führt bzw. die Darstellung empirisch unmöglicher Verhältnisse ermöglicht. Jede formale Repräsentation führt zu einer Betonung bestimmter Eigenschaften des Problems auf Kosten anderer, wenn das Problem nicht auf eine bestimmte Sprache fixiert sein sollte. Zu diesen Verzerrungen kann auch die Wahl einer bestimmten Auflösung der Suche gezählt werden, durch die minimale Suchinkremente definiert werden, aber auch bestimmte Werte von Variablen sprachlich ausgeschlossen werden. Bereits hier findet eine Reduktion des Suchraums statt, die je nach Problembeschaffenheit harmlos oder gravierend sein kann. Häufig wird der Problematik der Sprache zu wenig Beachtung geschenkt.

Einer zweiten Gruppe von Verzerrungen geben Witten und Frank (2001) den Namen *Such-Bias*. Diese Vezerrung tritt durch die Art und Weise hervor, mit der der Suchraum nach der gewünschten Lösung durchsucht wird. Jedes heuristische Verfahren verkürzt den Suchraum durch charakteristische Vereinfachungen. Als Beispiel soll hier das sogenannte *Hill-Climbing*-Verfahren zur Funktionsoptimierung[4] herangezogen werden: Bei der Suche nach einer Parameterkombination mit maximalem assoziierten Funktionswert wird ein zufälliger Ausgangspunkt gewählt, und die Heuristik folgt von diesem Punkt aus immer dem größten Steigungsgradienten, bis ein Gipfel der Funktion erreicht wird, von dem aus keine Verbesserung mehr möglich ist [5]. Ein *Hill-Climbing*-Algorithmus wird deshalb auch als gieriger Algorithmus (*greedy algorithm*) bezeichnet, da er in jedem Schritt versucht, die maximal mögliche Verbesserung der Lösung zu erreichen. Der Suchraum aller möglichen Parameterkombinationen wird durch diese Methode drastisch eingeschränkt, da nur die Punkte auf der Spur zum Gipfel sowie die Steigungsgradienten in diesen Punkten zu überprüfen sind. Allerdings stellt nichts in der Prozedur sicher, dass der erreichte Gipfel der global höchste Gipfel ist, obwohl das Ergebnis der Prozedur mit Sicherheit einen lokalen Gipfel findet. Sicherlich kann die Prozedur mehrfach gestartet werden mit unterschiedlichen Anfangspunkten, aber auch eine Konvergenz der Ergebnisse kann nicht die Optimalität des Ergebnisses sicherstellen.

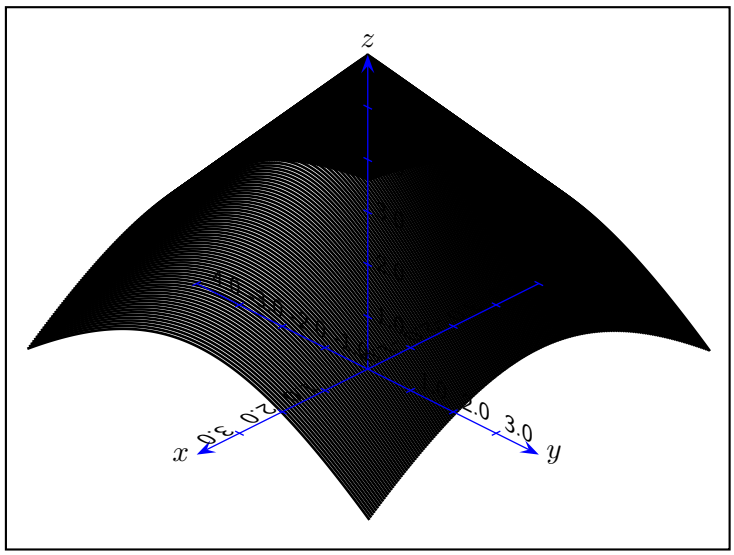


Abbildung 2.1: Günstige Funktion für die Anwendungen der *Hill-Climbing*-Heuristik: Das einzige lokale Maximum ist auch das globale Maximum. Die Abbildung zeigt die Funktion $z = 6 - \sqrt{x^2 + y^2}$

Zur Veranschaulichung seien zwei Funktionen gezeigt, für deren Analyse ein *Hill-Climbing*-Algorithmus

[4] Auf eine Ambiguität von „Funktionsoptimierung“ sei an dieser Stelle hingewiesen:

> [...] two phenomena are often confused under the term „optimise“. One concerns the outcome of performing optimally. [...] this sense of optimise can only be defined with respect to both a currency and a specified environment. The other meaning of optimise is the process of finding an optimum. Often the term in this sense is applied to a method that tries to find the optimum but which might not get there [...]. (Hutchinson & Gigerenzer, 2005a, S. 161)

Im Rahmen des *Data Mining* bezeichnet die Aufgabe der Funktionsoptimierung auch den *Versuch* der Funktionsoptimierung, selbst wenn nicht garantiert oder auch nur bestimmt werden kann, ob ein Optimum erreicht wird.

[5] Auf Varianten der Heuristik, bei denen feste Schrittgrößen zum Einsatz kommen oder eine Überschussbewegung eingeführt wird, sei nur hingewiesen.

unterschiedlich gut geeignet ist. Die Funktion in Abbildung 2.1 stellt den günstigsten Fall für die Anwendung von *Hill-Climbing* dar: das einzige lokale Maximum der Funktion ist auch das globale Maximum. Ganz anders hingegen präsentiert sich die zweite Funktion in Abbildung 2.2. Kann die erste Funktion visuell noch als Berggipfel beschrieben werden, gleicht die zweite Funktion eher einem Waldhügel, bei dem jeder Baum ein lokales Maximum bildet. Der Algorithmus hingegen ist blind für diesen Gestaltunterschied: Bei der ersten Funktion könnte der Gipfel von jedem Startpunkt aus erreicht werden, bei der zweiten Funktion würde jeweils nur der nächstliegende „Baum“ erklommen. Man sieht, dass die durch *Hill-Climbing* erreichte Einfachheit des Suchprozesses ihren Preis haben kann, der sich nicht notwendigerweise im Ergebnis der Anwendung zeigt: die Optimierung von visualisierbaren Funktionen ist ein so genanntes *Toy Problem*, bei dem die optimale Lösung bekannt ist und auch das Ausmaß der Abweichung eingeschätzt werden kann, um die Funktionsweise einer Prozedur zu analysieren. Im üblichen Anwendungsfall schweigt sich die Lösung über ihre eigene Angemessenheit aus. Man kann sagen, dass bei Anwendung von *Hill-Climbing*-Algorithmen in die Hypothese investiert wird, dass die Funktion eher der Funktion im ersten Fall gleicht. Je stärker die Abweichung von diesem Idealzustand ausfällt, desto stärker ist die Ausprägung der negativen Konsequenzen des Such-Bias, der durch *Hill-Climbing* eingeführt wird.

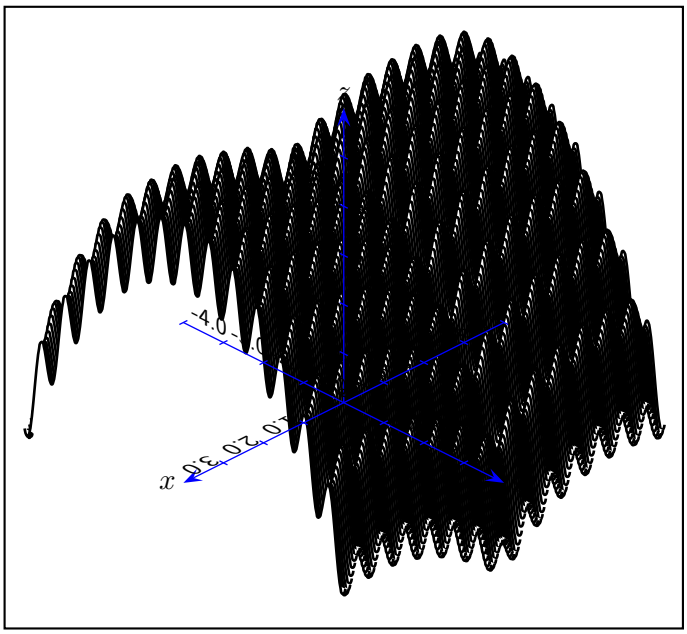


Abbildung 2.2: Ungünstige Funktion für Anwendungen der *Hill-Climbing*-Heuristik: Es gibt zahlreiche lokale Maxima mit kleinerem Funktionswert als das globale Maximum. Die Abbildung zeigt die Funktion $z = 5 + \frac{1}{3}(\cos 600x + 2 \cdot \cos 600y) - \frac{3}{10}(\frac{1}{2}(|x| - 3)^2 + y^2)$

Witten und Frank (2001) sehen auch Bemühungen zur Vermeidung von Überanpassung an Datensätze als einen Bias, da jede hierzu eingesetzte Methode erneut in bestimmte Hypothesen investiert. Bei Klassifikationsbäumen etwa kann einer solchen Überanpassung durch *Pruning* vorgebeugt werden, also durch das Beschneiden von Bäumen. Dies kann entweder vor dem Aufbau von Ästen eingesetzt werden als *Forward Pruning* oder nach Aufbau des vollständigen Baums als *Backward Pruning*. Die erste Methode investiert in die Hypothese, dass die Problemstruktur keinen komplexeren Baum erfordert, die zweite Methode investiert in die Hypothese, dass die konditionalen Verzweigungen in nicht-beschnittenen Ästen des Baums robust sind.

Eine weitere Verzerrung kann in der Stichprobenziehung auftreten, so dass Daten vorliegen, die in bestimmter Hinsicht ausgewählt worden sind. Dies kann zu einem großen Problem werden, wenn die Daten trotz einer Vorauswahl als Zufallsstichprobe analysiert werden. Glymour et al. (1997) betonen deshalb die folgenden beiden Konsequenzen:

- ... *any* technique used to analyze truncated data as if it was a random sample, can be fooled, regardless of how the truncation was induced;
- the data themselves are seldom capable to alert the analyst that a selection mechanism is operating - information external to the data at hand is critical in understanding the nature and extent of potential biases. (S. 22)

Die Autoren nennen eine derartige Verzerrung einen *Selection Bias*. Als drastisches Beispiel für die Konsequenzen mangelnder Vorsicht verweisen die Autoren auf die Vorgeschichte des *Challenger-Space-Shuttle*-Unglücks:

> The Rogers Commission concluded that an O-ring failure in the solid rocket booster led to the structural breakup and loss of the Challenger. In reconstructing the events leading up to the decision to launch, the Commission noted a mistake in the analysis of thermal-distress data whereby flights with no (i.e., zero) incidents of O-ring damage were excluded from critical plots of O-ring damage and ambient launch temperature since it was felt that they did not contribute any information about the temperature effect. This truncation of the data led to the conclusion that *no* relationship between O-ring damage and temperature existed, and ultimately, the decision to launch. (Glymour et al., 1997, S. 22)

Zuletzt sei noch auf einen Bias hingewiesen, der dem menschlichen Anwender von *Data Mining*-Verfahren zu verdanken ist, dessen Eliminierung aber selten wünschenswert erscheint. Die Ermittlung von statistischen Zusammenhängen steht in keiner direkten kausalen Beziehung zur Etablierung normativer Praktiken oder Entscheidungen. Witten und Frank (2001) nennen als Beispiel ein Analyseergebnis, demnach Menschen mit roten Autos ihre Autokredite mit der geringsten Wahrscheinlichkeit zurückzahlen. Eine Regel, nach der Menschen aufgrund einer Präferenz für eine bestimmte Wagenfarbe der Kredit verwehrt bleiben soll, erscheint aber mehr als fragwürdig. Subtiler noch sind hier allerdings vermittelte Zusammenhänge mit Geschlecht, Nationalität, Religionszugehörigkeit etc. Auch wenn mehr und mehr fragwürdig erworbene Informationen wie Summenstatistiken über Wohngebiete, Einkaufslisten auf „Kundenkarten“ oder ähnliches die Datenbanken von Data Minern ergänzen, ist die Reflexion über ethische Fragen der Anwendung und selbst bei Verzicht auf diese Perspektive die Berücksichtigung legaler und monetärer Konsequenzen von Diskriminierungen und Veränderungen im Image aufgrund unvorsichtig gewählter Regeln unverzichtbar. Da mit dieser Vorgehensweise aber Lösungen zurückgewiesen werden, deren Unzulässigkeit nicht statistisch nachgewiesen werden kann, ist somit ein ethischer oder pragmatischer Bias in das *Data Mining* eingeführt, der im Gegensatz zu anderen Fehlerquellen stärker der bewussten Kontrolle unterliegt.

2.2 Die Relevanz von Regelinduktion für die psychologische Diagnostik

Die Anwendung von Entscheidungsregeln ist für die praktische Diagnostik unerlässlich. Allerdings liegen diese Regeln in vielen Fällen lediglich implizit vor und sind in der Praxis eines Entscheidungsprozesses verborgen, der persönlicher Erfahrung und Expertenwissen, mitunter sogar Übung oder Intuition primäre Bedeutung zuteilt (vgl. Miesen, Schuhfried & Wottawa, 1999). Dieser Umstand ist nicht in jedem Fall unproblematisch:

> Sicher spielt die praktische Erfahrung der Diagnostiker eine Rolle. Doch oft fehlen ihnen die notwendigen Rückmeldungen. Sie erfahren zu selten, welche Entscheidung richtig war und welche nicht. Und selbst wenn ein Diagnostiker erfährt, was für Treffer er gelandet hat, weiß er ohne explizites Regelsystem nach ein oder zwei Jahren nicht mehr, welche konkreten Regeln er damals verwendete und heute wieder einsetzen und ausbauen könnte. (Miesen et al., 1999, S. 18)

Data-Mining-Verfahren für die Induktion expliziter Entscheidungsregeln sollen hier hinsichtlich ihrer Relevanz für die psychologische Diagnostik untersucht werden. Zunächst soll allerdings an einem Beispiel von Dawes gezeigt werden, dass die Suche nach Entscheidungsalgorithmen und die Akzeptanz dieser Prozeduren zwei sehr verschiedene Prozesse sein können. Gleichzeitig kann das Beispiel auch einen ersten Beleg für kontraintuitive Qualitäten simpler Entscheidungsstrategien liefern.

2.2.1 Experten und einfache lineare Modelle

Dawes (1988) zeichnet den Konflikt nach, den der Entscheider erlebt, der seine eigene Entscheidungskompetenz an ein algorithmisches Entscheidungsverfahren überträgt:

> The idea that a self-imposed external constraint on action can actually enhance our freedom by releasing us from predictable und undesirable *internal* constraints is not a popular one. It is hard to play Ulysses. The idea that such internal constraints can be cognitive, as well as emotional, is even less palatable. Thus, to allow our judgment to be constrained - or even influenced - by the „mere numbers“ or pictures or external aids offered by computer printouts is anathema to many people. In fact, there is even evidence that when such aids are offered, many experts attempt to „improve upon“ these aids' predictions - and they do worse than they would have had they „mindlessly“ adhered to them. (Dawes, 1988, S. 143)

Eines der frühesten formalen Entscheidungsverfahren wird 1772 von Franklin in einem Brief an Joseph Priestley berichtet, eine moralische Algebra, die auch als *Franklin's Rule* bezeichnet worden ist. Franklin benutzte dazu ein Blatt Papier mit zwei Spalten, auf dem er im Laufe einer halben Woche Pro- und Kontra-Argumente für ein Enscheidungsproblem sammelte, um dann in einer Art Gesamtüberschau die relativen Gewichte der Argumente zu bestimmen:

> Franklin's „prudential algebra“ involves nothing more than a *weighted average* of reasons for and against a particular course of action. Each positive reason is assigned a score of +1, each negative one a score of -1. The scores are then given intuitive *importance weights*. The course of action with the higher resulting sum is judged to be the more desirable one. (Dawes, 1988, S. 202)

Nach Dawes (1988) sind Gewichtungsschemata beschriebener Art im Allgemeinen bessere Entscheidungsverfahren als die reine, ununterstützte Intution:

> When judgment can be compared with [...] external standards of accuracy, weighting schemes *consistently* are found to be superior to intuitive global judgment, and judges consistently are found to be overconfident in their global judgments. (Dawes, 1988, S. 203)

Dawes (1979) stellt eine Untersuchung mit drei Entscheidungsaufgaben an, bei denen die Ergebnisse linearer Entscheidungsmodelle mit den Urteilen von Experten verglichen werden. Die Aufgaben sind im einzelnen:

1. Die Entscheidung einer Klassifikation von Neurose vs. Psychose bei 860 Psychiatriepatienten auf der Grundlage von MMPI-Ergebnissen;
2. die Vorhersage durchschnittlicher Abschlussnoten (*GPA*) von Studenten im ersten Jahr des Graduiertenstudiengangs auf der Basis von akademischen Vorleistungen und Persönlichkeitseigenschaften;
3. die Vorhersage der Einschätzung von Studenten in Graduiertenstudiengängen durch Mitglieder der Fakultät nach zwei bis fünf Jahren auf der Basis ihrer Leistungen im Vorgraduiertenbereich, ihrer GRE-Ergebnisse und eines Maßes für die Selektivität der Universität ihres ersten Abschlusses.

Dawes (1979) stellt fest, dass die linearen Modelle im Vergleich zu den Experten deutlich mehr Varianz aufklären. Dabei entscheidet sich Dawes bewusst für sehr einfache lineare Modelle, die auch als „unangemessene“ lineare Modelle (*improper linear models*) bezeichnet werden können:

> An *improper linear model* is one in which weights are chosen by some nonoptimal method. They may be chosen to be equal, they may be chosen on the basis of the intuition making the prediciton, or they may be chosen at random. Nervertheless, improper models may have great utility. (Dawes, 1982, S. 392)

Tatsächlich klärten lineare Modelle mit zufälliger Gewichtung der Prädiktoren 150% mehr Kriteriumsvarianz auf als die Experten, ein lineares Modell mit einheitlicher Gewichtung der Prädiktoren sogar 261% mehr Varianz (Dawes, 1979). Eine Erklärung für diese zunächst im doppelten Sinne kontraintuitiven Befunde stellt nach Dawes (1982) die Überlegung dar, dass Menschen zwar besser als Algorithmen dazu in der Lage seien, Informationen, die eine Unterscheidung unterstützen können, auszuwählen und zu kodieren, dass Modelle aber besser in der Lage seien, diese Informationen auszuwerten und zu integrieren, um zu einer Entscheidung zu gelangen. Als möglichen Grund für das gute Abschneiden der

unangemessenen Modelle sieht Dawes (1982) die Notwendigkeit einer genügend großen Fallzahl für die Schätzung von Regressionsgewichten, die er bei einem Minimum von zwanzig Fällen pro Gewicht ansetzt und selten in praktischen Entscheidungssituationen erreicht findet. Unterhalb dieses Minimums schneiden seiner Meinung nach die einheitlich gewichteten Modelle besser ab als die geschätzten[6].

Ist keine messbare Kriteriumsvariable vorhanden, zeigen Studien zum *Bootstrapping* eine zweite Möglichkeit, wie Modelle Experten übertreffen können, welche auf den ersten Blick unglaublicher erscheint: Wird das Entscheidungsverhalten des Experten genutzt, um ein lineares Modell zu spezifizieren, ist das entstandene lineare Modell oft besser in der Lage, zukünftige Entscheidungen korrekt zu treffen als der Experte selbst (Dawes, 1982). Dies scheint unmöglich, da jener schließlich als Original diente und die Kopie mit zahlreichen Einschränkungen versehen ist. Allerdings erweisen sich diese Einschränkungen gerade in manchen Situationen als vorteilhaft: Zum einen gleichen lineare Modelle zufällige Abweichungen von einer zugrundeliegenden Entscheidungsregel bzw. *decision policy* aus, wenn es eine solche gibt. Zudem könnte das Modell dazu in der Lage sein, selektiv die Essenz der Expertise zu erfassen, während der Experte nicht in der Lage zu sein scheint, validen Prinzipien zu folgen, was die Ähnlichkeit seines Abschneidens mit dem von zufallsgewichteten Modellen illustriert. Dawes (1982) beweist zudem folgenden Satz:

Satz 2.1 (Korrelation zwischen Prädiktorvariablen und Kriterium) *Die durchschnittliche Korrelation zwischen Prädiktorvariablen und Kriterium ist kleiner als die Korrelation zwischen dem Durchschnittswert der Prädiktorvariablen und dem Kriterium.*

Es seien $X_1, X_2, \ldots, X_m$ standardisierte Variablen, die alle positiv korreliert seien mit der standardisierten Variablen Y, also $r_{x_k,y} > 0$. Die Korrelation des durchschnittlichen Variablenwerts über alle Variablen X_k entspricht damit der Korrelation der Summe der Variablen mit Y. Die Kovarianz zwischen Y und der Summe der Prädiktorwerte lässt sich aufgrund der Standardisierung der Variablen schreiben als:

$$\begin{aligned} s_{x_{i1}+\ldots+x_{im},y_i} &= \frac{1}{n}\sum_i (y_i(x_{i1}+\ldots+x_{im})) \\ &= \frac{1}{n}\sum_i (y_i x_{i1}) + \ldots + \frac{1}{n}\sum_i (y_i x_{im}) = r_{x_1,y} + \ldots + r_{x_m,y} \end{aligned} \tag{2.1}$$

Die Varianz des (standardisierten) Kriteriums ist $s_y^2 = 1$, die Varianz der Summe der Prädiktoren ist:

$$s^2_{x_{i1}+\ldots+x_{im}} = m + m(m-1)\bar{r}, \tag{2.2}$$

wobei $\bar{r}$ der durchschnittlichen Interkorrelation zwischen den Pädiktorvariablen entspricht. Daher gilt mit $\bar{x}_i = \frac{1}{m}\sum_{k=1}^m x_{ik}$

$$r_{\bar{x}_i,y} = \frac{\sum_i r_{x_i,y}}{\sqrt{m+m(m-1)\bar{r}}}. \tag{2.3}$$

Zudem ist damit für $\bar{r} < 1$:

$$r_{\bar{x}_i,y} = \frac{\sum_i r_{x_i,y}}{\sqrt{m+m(m-1)\bar{r}}} > \frac{\sum_i r_{x_i,y}}{\sqrt{m+m(m-1)}} = \frac{\sum_i r_{x_i,y}}{m} \tag{2.4}$$

■.

Dieser Satz hilft zur Erklärung, warum das zufällige Modell schlechter abschneidet als das gleichgewichtete Modell: Weil jedes zufällig ermittelte Modell positiv mit dem Kriterium korreliert ist und der Durchschnitt der Zufallsmodelle probabilistisch dem Gleichgewichtungsmodell entsprechen muss, wenn die relativen Gewichte unabhängig voneinander mit gleichem Erwartungswert bestimmt werden, ist die

[6] Die vereinfachten Modelle stellen zwar einen Beleg für mögliche Vorteile von Einfachheit dar, Dawes versteht sie aber nicht als Heuristiken, und er lässt sich entschieden auch nicht in diesem Sinne vereinnahmen, da er einen pejorativen Heuristik-Begriff verwendet, den er von den vorgestellten Modellen abgrenzt:

> The basic point of this book is that we often think in automatic ways about choice situations, that these automatic thinking processes can be described by certain psychological rules („heuristics“), and that they systematically lead us to make poorer choices than we would by thinking in a more controlled manner about our decisions. (Dawes, 1988, S. 7)

durchschnittliche Korrelation zwischen Zufallsmodellen und Kriterium dementsprechend kleiner als die Korrelation zwischen gleich gewichtetem Modell und Kriterium. Die beschriebene Form der Gleichgewichtung wird mitunter auch als *Dawes's Rule* bezeichnet. Weiterhin beobachtet Dawes (1982), dass die Lösung für die Bestimmung der optimalen Gewichte in Regressionsmodellen ein flaches Maximum (*flat maximum*) hat, d.h. lineare Modelle erweisen sich im Allgemeinen als robust gegenüber Abweichungen von den optimalen Gewichten:

> [...] coefficients are not as important in linear models as their signs. Thus arbitrary (random) linear models can approximate monotone interaction effects. And of course, they approximate main effects. (Dawes, 1988, S. 214)

Ist die Beziehung zwischen Prädiktoren und Kriterium konditional monoton, empfehlen sich einfache lineare Modelle in dieser Betrachtung für die Erstellung von Entscheidungsregeln, wenn von menschlicher Seite aus Variablen identifiziert werden können, die diese Eigenschaften mitbringen. Dawes (1988) weist darauf hin, dass disordinale Interaktionen leichter hypothetisch zu ersinnen als empirisch zu finden seien. Dementsprechend ist eine Arbeitsteilung zwischen Experten und Modellen die Empfehlung, die Dawes hier vorschlägt (vgl. Whitecotton, Sanders & Norris, 1998):

> [...] because most interactions that do exist in nature are monotone and because people have difficulty integrating information from noncomparable dimensions, linear models will outperform clinical judgment. The only way to avoid this broad conclusion is to claim that training makes experts superior to other people at integrating information (as opposed, for example, to knowing what information to look for), and there is no evidence that experts *think differently* from others. (Dawes, 1988, S. 215)

Dawes (1982) nennt eine Reihe von Einwänden, die gegen den Einsatz dieser Modelle in sozial bedeutsamen Entscheidungssituationen vorgebracht werden, auch wenn die Auswahl der Variablen menschlichen Entscheidern überlassen wird. Neben technischen Einwänden, dass nicht Korrelationen, sondern Nutzenwerte interessieren mögen[7], sind dies vor allem auch psychologische Schwierigkeiten:

- Einzelfälle werden zur Widerlegung des Modells angeführt: Sorgfältige Aufzeichnungen könnten die Bedeutsamkeit dieser Fälle eventuell reduzieren.
- Es kommt zu selbsterfüllenden Prophezeiungen: Entscheidungen des Modells werden so lange bekämpft, bis sie zu schlechten Entscheidungen werden müssen. Eigene Entscheidungen können nachträglich in ähnlicher Weise zu guten Entscheidungen gemacht oder befunden werden[8].
- Es gibt eine Verwechslung zwischen der Möglichkeit der Überlegenheit eines zweiten Modells und der faktischen Überlegenheit.
- Es wird fälschlicherweise angenommen, dass ein Großteil der Varianz erklärbar sein muss, so dass die Leistung des Modells als enttäuschend ausgewiesen wird.

Der letzte Punkt ist insbesondere interessant, da hier ein Wunsch nach Vorhersagbarkeit die Wahrnehmung realistischer Prognosemöglichkeiten verzerrt:

> *I want it to be predictable, especially when the aspect of it that I want to predict is important to me.* This desire, I suggest, translates itself into an implicit assumption that the future is in fact highly predictable, and it would then logically follow that if something is not a very good predictor, something else might do better (although it is never correct to argue that it necessarily will). (Dawes, 1982, S.406)

[7] Bei komparativen Wahlen wären oft aber eher Rangkorrelationen entscheidend (Dawes, 1982).

[8] Dazu notiert Dawes (1979):

> [...] admission officers who judge that a candidate is particularly qualified for a graduate program may feel that their judgment is exonerated when the candidate does well, even though the candidate's success is in large part due to the positive effects of the program. (S. 580)

Aus dieser Haltung heraus wird jedes schlechte Ergebnis einer statistischen Vorhersage auf ein Versagen der statistischen Methode attribuiert, und nicht etwa auf die prinzipielle Unmöglichkeit der Prognose:

> Statistical prediction, because it includes the specification (usually a low correlation coefficient) of exactly how poorly we can predict, bluntly strikes us with the fact that life is not all that predictable. Unsystematic clinical prediction (or „postdiction“), in contrast, allows us the comforting illusion that life is in fact predictable and that we can predict it. (Dawes, 1982, S.406)

Als Beispiel für die notwendige Bescheidenheit in der Prognose von Studienerfolg zählt Dawes Einflussfaktoren auf dieses Kriterium auf:

> Academic success [...] is influenced by whom one shares an office with as a graduate student, by which professors happen to have positions available for research assistants, by the person or people with whom one has libidinal involvement (often met on a „chance“ basis), by the relative strengths of those with whom one competes for the first job (as judged by the professors who happen to be appointed to the „search committee“), and so on. (Dawes, 1988, S. 218)

Nach diesen ersten Entscheidungen führen Verstärkungen von Entwicklungen und Pfadabhängigkeiten zu sehr differenziellen Werdegängen, obwohl die Kandidaten als auch ihre Leistungen zu einem bestimmten Zeitpunkt sehr ähnlich gewesen sein könnten.

Als weitere Gruppe von Einwänden nennt Dawes ethische Erwägungen, nach denen der Verzicht auf ein Interview etwa als dehumanisierende Behandlung kritisiert wird. Hier ist das Gegenargument von Dawes (1982) sicherlich kontroverser zu diskutieren: Seiner Meinung nach drückt eine solche Haltung - vom Interviewenden geäußert - eine kognitive Eitelkeit aus, besser urteilen zu können nach einem kurzen Interview als akkumulierte Bewertungen aus fast vier Jahren Unterricht (selbst das Vorhandensein eines gefundenen speziellen Talentes könnte bei Bewerbern mit höherem Testergebnis auch gegeben sein[9]). Dementsprechend hält es Dawes trotz einer möglichen „ethischen Ungemütlichkeit“ für eine Tatsache, dass Menschen in der bestmöglichen Weise behandelt werden sollen, auch wenn dies bedeutet - wie seine Ergebnisse demonstrieren - dass Selektion, Diagnose und Prognose durch nichts mehr als durch die Addition einiger Werte auf wichtigen Attributen bewerkstelligt werden sollen. Ein abweichendes Verhalten wäre für Dawes dementsprechend ein Betrug am Klienten.

Die Fragen, die Dawes bei dieser lapidaren Einschätzung allerdings nicht berücksichtigt, sind Fragen bezüglich eines möglichen Dissens bezüglich des Kriteriums. Eine Automatisierung des Verfahrens würde ein Restmoment von Wertepluralität vernichten können, dass durch einen persönlichen Handlungsspielraum von menschlichen Entscheidern vorhanden sein könnte - im Guten wie im Schlechten. Sicherlich sollte nicht ausgeschlossen werden, dass es verantwortungsbewusste Entscheider gibt, deren Formalisierung einen Verlust bedeutet, doch leider gibt es unter Umständen mehr selbstüberzeugte als überzeugende Entscheider. Dawes nennt das Beispiel einer Untersuchung, in der die Suche nach einer Begründung für schlechte Rekrutierungsergebnisse von Bewerberinnen im Zentrum stand. Dabei konnte Dawes einen der Faktoren identifizieren:

> One of the older professors had cut back on his practice to devote time to interviewing applicants to the school. He assessed such characteristics as „emotional maturity,“ „seriousness of interest in medicine,“ and „neuroticism.“ Whenever he interviewed an unmarried female applicant, he concluded she was „not sufficiently interested in medicine,“ and when he interviewed a divorced one, he concluded she was „neurotic.“ Not many women were positively evaluated on these dimensions, which of course had nothing to do with gender. (Dawes, 1988, S. 219)

In diesem Sinne kann die Betonung von menschlicher Urteilskraft bei einem Fehlen geeigneter und erfahrener Entscheider zu einem tatsächlichen Schaden führen:

[9] Trotzdem bietet das Interview potenziell die Möglichkeit der Messung von Charakteristika eines Bewerbers, die durch kein anderes Verfahren geprüft werden können. In diesem Fall spielt nach Guion (1991) die Messgenauigkeit eine nachgeordnete Rolle, zudem Interviews eine *Public Relations*-Funktion für Unternehmen besetzen können, im Sinne einer Unternehmenspräsentation, aber auch einer Verhinderung von Entfremdungsempfindungen bei den Abgelehnten.

> It opens the door to shoddy practices, trivialities, snap judgments, and a willingness to defend what is easy - all on the grounds of professional judgment; it may permit or even encourage people who are neither well informed nor have much depth of understanding to say, „My judgment is just as good as „anyone's.“ (Guion, 1991, S. 388)

Allerdings beugt die von Dawes vorgeschlagene Methode - wie die Behandlung durch menschliche Interviewer - nicht der Möglichkeit einer systematischen Diskriminierung von Teilgruppen vor: Eine Reflexion des Verfahrens und eine Beobachtung seiner Entscheidungen bleibt also unabdingbar.

2.2.2 Computergestütztes Recruiting und Regelinduktion

Anders als in den von Dawes beschriebenen Settings ist es in computernahen Entscheidungssituationen einfacher, Algorithmen zu implementieren. Aus einer Vertrautheit mit algorithmischen Prozeduren kann - mit allen Gefahren einer unreflektierten Technikakquieszenz - auch eine höhere Akzeptanz für sachgerecht implementierte Regeln entstehen.

Zunehmend nutzt die psychologische Diagnostik computerbasierte Verfahren und Prozeduren, und neben PC-Verfahren, die sich schon seit Jahrzehnten im Einsatz befinden, rücken diagnostische Verfahren im Internet z.B. im Bereich des *Internet-Recruiting* nach vorn. Die Bedeutung des *Internet-Recruiting* für die Eignungsdiagnostik vor allem im wirtschaftlichen Sektor ist in den letzten Jahren massiv angewachsen. Das Internet dient dabei sowohl als Informationsquelle über vakante Positionen und Karrieremöglichkeiten als auch als Kontaktaufnahmemöglichkeit für Bewerber und Unternehmen. Ein Großteil der bedeutenderen Unternehmen betreibt ein aktives *Online-Recruiting* (vgl. Wottawa & Woike, 2002). Die Formen des *Internet-Recruiting* reichen dabei von einer einfachen Digitalisierung des Postwegs über die interaktive Bearbeitung von Formularen bis hin zu komplexeren Einstellungsprozeduren wie Informationsspielen, Einstellungstests und Kommunikation über das Internet.

Die Akzeptanz von netzbasierten Verfahren ist nicht immer gegeben, unterschiedliche Geräteausstattung, Probleme des Identitätsnachweises und die fehlende Kontrolle über bereitgestellte Inhalte werden häufig als Gegenargumente angeführt. Diesen Argumenten können einige Vorteile internetbasierter Verfahren gegenübergestellt werden: Das Internet ermöglicht die Kombination aus zentraler Verwaltung von Daten und Inhalten einerseits und distribuierter Nutzung andererseits, so dass neben den durch Virtualisierung gesparten Materialkosten auch Reisekosten gespart werden können und von vielen Orten aus Zugriffsmöglichkeiten bestehen. Gleichzeitig kann von zentraler Stelle aus ein Update vorgenommen werden, ohne dass logistische Zusatzkosten entstehen. Zudem ermöglicht die virtuelle Bewerbung eine Beschleunigung der Bearbeitung, was bei starker Konkurrenz um geeignete Bewerber zu einem Vorteil werden kann (vgl. Wottawa, Kirbach, Montel & Oenning, 2005). Die Probleme hinsichtlich der Ungleichheit von internetbezogenen Fähigkeiten und Zugangsschwierigkeiten werden zudem seit Beginn der zweiten Generation von Webbenutzern, bei denen ein Internetanschluss zuhause und am Arbeitsplatz nicht mehr erst eingerichtet werden musste (in den USA seit ca. 1997, in Europa anderthalb Jahre später), zunehmend nivelliert (Bartram, 2000). Dabei kann Internet-Recruiting eine ganze Reihe von Zielen verfolgen. Dazu zählen:

- Personalmarketing
- Bereitstellung von Wettbewerben und Spielen
- Aufbau von Bewerberdatenbanken
- Bewerbertestung
- Screening von Bewerbern
- Anonymes Self-Assessment
- Trainingsvorbereitung und -durchführung

Das Screening wird bei großen Bewerberzahlen zu einer entscheidenden Aufgabe, eine reine Analyse der Lebenslaufdaten und der Bewerbungsmappen ist aber oft zu grob für eine sachgerechte Entscheidung. Hier können Online-Verfahren Abhilfe schaffen (Wottawa et al., 2005). Feldman und Klaas (2002) beobachten

eine bevorzugte Nutzung des Internets bei der Stellensuche bei größeren gewünschten Gehaltssprüngen und einer breiten geographischen Orientierung. Die Effektivität der Suche wird von Bewerbern zwar als niedriger als das persönliche Networking, aber als höher als die Suche durch Zeitungsinserate oder Direktanrufe eingeschätzt.

Es gibt zahlreiche Unterschiede zu konventioneller Diagnostik. Zum einen sind interaktiven Möglichkeiten größer, es werden nach und nach neue Testformate entwickelt und weiterenwickelt (vgl. Montel & Wottawa, 2001), zum anderen unterliegt die Gewinnung von Teststichproben veränderten Bedingungen: zum Teil können zielspezifisch repräsentative Stichproben für das interessierenden Anwendungsgebiet einfacher gewonnen werden, während die Qualität der Daten und die Motivation der Teilnehmer auf gesonderte Weise zu testen sind. Für computergestützte Diagnostik ist es wie für die herkömmliche Diagnostik unverzichtbar, zwei Facetten in der Entscheidungsprozedur eindeutig zu bestimmen: Zum einen muss eine Auswahl von Skalen erfolgen, die für die Testung herangezogen werden, zum anderen müssen auf der Grundlage dieser Auswahl Entscheidungsregeln definiert werden, die von der Erhebung und Auswertung der einzelnen Tests zur Diagnose führen. Die Bestimmung der notwendigen Skalen ist in hohem Maße abhängig von der erfolgten Erstellung eines Anforderungsprofils:

> Hervorzuheben ist [...] die inhaltliche Bedeutung von Anforderungsprofilen. Nur bei transparenten Anforderungen lässt sich die Passung von Person zu Tätigkeit beurteilen und Über- oder Unterforderung vermeiden. (Heider-Friedel, Strobel & Westhoff, 2006, S. 24)

Was wie ein Truismus anmutet, stellt sich in der unternehmerischen Praxis allerdings als potenzielle Barriere heraus, da der Aufwand der Erstellung von Profilen in vielen Fällen vermieden werden soll:

> Oft wird auf wissenschaftlich entwickelte Verfahren verzichtet [...]. Anforderungen werden stattdessen durch mehr oder weniger erfahrene Personen subjektiv und unsystematisch abgeschätzt. Entscheidungen, die zu einem bestimmten Anforderungsprofil führen, sind damit nicht immer nachvollziehbar [...] und gerade im Auswertungsprozess herrscht häufig kein gemeinsames Verständnis darüber, was unter einer Anforderung zu verstehen ist, da konkrete Verhaltensbeschreibungen fehlen. (Heider-Friedel et al., 2006, S. 24)

Neben der Auswahl geeigneter Skalen, die auf die Anforderungen abgestimmt sein sollten, rückt damit als nächstes der Umgang mit Skalenwerten für Diagnosefälle in den Vordergrund. Für viele Anwendungsfälle ist damit der Bezug zur Regelinduktion augenfällig, gerade im Hinblick auf die Schritte der Filterung von Teilnehmern und der Auswahl von Daten:

> **Eine ganz besonders wichtige Herausforderung liegt in der sachgerechten Erstellung solcher Filter und „Entscheidungsregeln“,** gleichgültig ob diese für die Auswahl von Bewerbern durch Recruiter oder für die „Beratung“ von Bewerbern unmittelbar herangezogen werden. **Die klassischen Kennziffern der Testkonstruktion** (zum Beispiel Validitätskoeffizienten) **sind dafür absolut nicht geeignet** (es gab noch nie einen Diagnostiker, der auf der Basis von Validitätskoeffizienten in der Lage war, Testprofile in diagnostische Entscheidungen für den Einzelfall zu übersetzen). Es müssen **statt dessen konfigurale Techniken** eingesetzt werden. (Wottawa & Woike, 2002, S. 38)

Für zeitgenössische diagnostische Maßnahmen gewinnt zudem die eingeführte Norm DIN 33430 an Bedeutung.

2.2.3 DIN 33430 und die Form von Entscheidungsregeln

Eine Präzisierung der Bedingungen für professionelle Diagnostik in Abgrenzung zu quasi-wissenschaftlichen Ersatzpraktiken soll mit der DIN 33430 (Ausgabe: 2002-06) für berufsbezogene Eignungsbeurteilungen etabliert werden[10]. Wottawa und Oenning (2002) betrachten die folgenden Aspekte der Norm als zentral:

[10] Die DIN 33430 ist nicht unumstritten. Die Norm wird von Kersting und Hornke (2003) bspw. hinsichtlich ihrer Sprache kritisiert, die vor allem für Nicht-Psychologen schwer verständlich bleibt und ebenso hinsichtlich der nationalen Ausrichtung bei gleichzeitiger Internationalisierung der Eignungsdiagnostik. In einer entscheidenden Hinsicht ist die Norm zudem für die Autoren unvollständig:

- Eine Anforderungsanalyse wird als Grundlage für diagnostische Verfahren eingefordert.
- Ein mit der entsprechenden Qualifikation ausgestatteter Auftragnehmer soll die Regeln der Interpretation von Ergebnissen festlegen und verantworten.
- Es wird eine explizite Festlegung der Regeln gefordert, nach denen die Eignungsbeurteilung aufgrund der eingehenden Informationen zustande kommt (auch für Vorauswahlentscheidungen).
- Die Urteilungsbildung bei dieser Eignungsbeurteilung muss zum einen nachvollziehbar sein, zum anderen auch an Laien kommunizierbar sein.
- Die Verantwortung für das Festlegen von Interpretationen und Beurteilungsregeln wird eindeutig dem qualifizierten Auftragnehmer zugeschrieben.

Im Kontext dieser Arbeit besonders relevant ist die eindeutige Forderung nach explizierten Auswertungsregeln, die auch an Laien kommunizierbar bleiben müssen. Dies erzeugt zunächst ein methodisches Dilemma: Ist eine Methode für die Klassifikation von Fällen nachweislich besser geeignet als eine zweite, ist sie aber im Gegensatz zur zweiten Methode nicht an Laien konmunizierbar, dürfte die erste Methode keine Verwendung finden. Dies ist insofern eine gute Entscheidung, als dass die Verwendung zweifelhafter Kriterien und möglicherweise diskriminierender Teilregeln offensichtlich wird und dass die Anwendung und Interpretation der Regeln für alle Beteiligten klar nachvollziehbar bleibt. Angenommen, es handelte sich hier aber um ein medizinisches Diagnoseverfahren, von dem Leben und Tod eines Patienten abhingen und ein neuronales Netzwerk mit zahlreichen Inputs und mehreren versteckten Neuronenschichten lieferte die besten Diagnoseergebnisse, dann wäre es unverantwortbar, ein nachweislich schlechteres Verfahren heranzuziehen, nur weil die Regeln, nach denen eine (Fehl-)Entscheidung generiert wurde, besser an die Verwandten des Patienten kommunizierbar sind. Im beruflichen Kontext ist die Situation glücklicherweise weniger lebensbedrohlich, trotzdem sollten Qualitätseinbußen nicht außer Acht gelassen werden. Glücklicherweise gibt es nicht notwendigerweise einen Konflikt zwischen Einfachheit der Regeln und der Genauigkeit der darauf fußenden Prädiktionen. Dieser Umstand gehört zu den zentralen Themen dieser Arbeit.

Die Form von Entscheidungsregeln ist abhängig von der Art der erforderlichen Eignungsbeurteilung (Wottawa & Oenning, 2002). Für eine absolute Beurteilung von Personen ist eine direkte Eignungsaussage über den Einzelfall erforderlich, so dass eine Person klassifiziert werden kann. Steht hingegen eine relative Eignungsaussage im Vordergrund, da aus einer Gruppe von Personen ausgewählt werden soll, müssen zumindest ordinalskalierte Prädiktionen erstellt werden.

Von besonderer Bedeutung ist weiterhin die Form der Verarbeitung gewonnener Testwerte: Bei einem einzigen Prädiktor muss bestimmt werden, wie Personengruppen mit bestimmten Prädiktorwerten zu behandeln sind, im einfachsten Fall kann ein Cutoff angegeben werden, der zwei unterschiedlich zu klassifzierende Gruppen voneinander trennt, wobei die Festlegung von Cutoff-Werten eben nicht durch Korrelationen ermittelt werden kann, sondern eine detaillierte Betrachtung von Verteilungen erfordert.

Bei mehreren Prädiktorwerten werden die Möglichkeiten für Regeln komplexer, ein mehrdimensionaler Input-Raum muss auf eine eindimensionale Entscheidungsdimension abgebildet werden. In vielen Fällen werden relativ schlichte Versuche der Vereinfachung dieser Aufgabe angewandt:

> Für Reihenfolgen-Entscheidungen werden in der Praxis oft „gewichtete Summen“ verwendet, wobei die „Gewichte“ für die Indikatoren meist durch Einschätzungen von Experten (viel seltener durch multiple Regressionen mit Daten aus Validierungsstudien) bestimmt werden.

Das schwerwiegendste Problem besteht darin, dass die DIN 33430 Regelungsbedarf schafft, der nicht durch die DIN selbst befriedigt wird. So wird von den Verantwortlichen und Mitwirkenden im eignungsdiagnostischen Prozess beispielsweise zwar der Nachweis definierter Kenntnisse und Erfahrungen verlangt, die Frage, wie dieser Nachweis zu erbringen ist, bleibt aber ausgespart. Eine große Gefahr geht schließlich von unseriösen Zertifizierungen von Verfahren und Weiterbildungen aus. Dadurch könnte minderwertige Qualität mit dem Mantel der DIN-Kompatibilität beschönigt werden. (Kersting & Hornke, 2003, S. 177)

Für eine produktive Wirkung der Norm sehen die Autoren weiterhin die Notwendigkeit einer postgraduierten eignungsdiagnostischen Weiterbildung von Psychologen und der entsprechenden Qualifizierung von Nichtpsychologen, einer Präsenz der DGP im Zertifizierungsverfahren, des Ausbaus von Review- und Informationssystemen, sowie der Einrichtung von Publikationsorganen und akkreditierten Beratungsstellen.

> Dieses Konzept ist mit dem Problem verbunden, dass es eine nahezu unbegrenzte Kompensationsmöglichkeit von Defiziten zulässt. [...] Da dies in nahezu allen Fällen den Ergebnissen einer vernünftigen Anforderungsanalyse widersprechen dürfte (so kann man z.B. massive Defizite im „Können“ nicht ohne weiteres durch besonderes „Wollen“ ersetzen, und umgekehrt), sind Ansätze dieser Art strukturell kaum geeignet, die Entscheidungsregeln für die Eignungsbeurteilungen, wie sie von der DIN 33430 gefordert werden, abzubilden. (Wottawa & Oenning, 2002, S. 46)

Eine gewichtete Summe von Prädiktorwerten $\{x_1, x_2, \ldots, x_k\}$ entspricht für ein Objekt j der Linearkombination:

$$s_j = \sum_{i=1}^{k} w_i x_{i,j} = w_1 x_{1,j} + w_2 x_{2,j} + \ldots + w_k x_{k,j} \tag{2.5}$$

Je nach Wertebereich von Gewichten und Prädiktorwerten sind unter der Annahme positiver Gewichte niedrige Werte auf einem Prädiktor durch hohe Werte auf einem oder mehreren anderen Prädiktoren in dem Sinne kompensierbar, dass ein Fall mit dieser Wertkombination ununterscheidbar ist in der Prädiktionssumme von einem Fall mit durchschnittlichen Werten auf allen Prädiktoren. Man nennt Modelle mit der beschriebenen Eigenschaft der Kompensierbarkeit auch *kompensatorische Modelle*.

Die beschriebene Verrechnung der Werte erfordert zudem das Intervallskalenniveau der Prädiktorwerte sowie die Existenz einer gemeinsamen Dimension der Vergleichbarkeit für eine sinnvolle Bestimmung von Gewichten. Die multiple Regression erzwingt - ebenso wie eine willkürliche Festlegung von Gewichten - eine solche Dimension als zunächst lediglich mathematisches Konstrukt. Allerdings führt dieses Vorgehen zu einem Konflikt, wenn Grenzwerte für derart berechnete Summenscores bestimmt werden sollen, etwa wenn die gewichtete Körpergröße zur gewichteten Soziabilität und zur gewichteten Nervenleitgeschwindigkeit addiert wird. Das Intervallskalenniveau kann für viele Skalen nicht sinnvoll angenommen werden, insbesondere eine Veränderung von Itemschwierigkeiten hat bei Verwendung der klassischen Testtheorie nicht-lineare Veränderungen der zu erwartenden Testreihe zur Folge, so dass nur bei einer vollständig operationalen Definition von Eigenschaften, die auch an einen festen Itemsatz geknüpft wird, die Skalenproblematik vermieden wird (Wottawa & Oenning, 2002). Die Verwendung linearer Modelle hat weiterhin auch Konsequenzen für den Umgang mit Modellergebnissen:

> Criterion-related validity has consistently assumed a linear model - a model that says that if some level of score is good, the next higher level is better, and that how much better is the same at any score level. When linearity is assumed in test validation, it is also assumed that selection will follow a top-down system, choosing first the person with the highest score, then choosing the one with the next highest score, and so on. (Guion, 1999, S. 359)

Diese Art der Entscheidungsfindung kann allerdings für viele Entscheidungsprozeduren nicht als angemessen angesehen werden:

> Except in the public sector, this is not how selection decisions are typically made. [...] In many real-life settings, cut scores are set and people are considered for employment every day with daily numbers to hire for specific jobs. As soon as enough people who „pass“ the test (and who have no disqualifying characteristic in the opinions of the decision makers) are found, hiring for tha job, for that day, is over, others coming later , even if more qualified, are not considered. (Guion, 1999, S. 359)

Bei der beschriebenen Verwendung wären dann auch andere beliebte Maße wie Validitätskoeffizienten irreführend. Als Begründung für die Verbreitung der Methode verweist Guion (1991) auf die Trägheit von Veränderungsprozessen in Praktikergemeinschaften:

> One tends to use the linear additive model, because, like Mount Everest, it's there. (S. 370)

Die breite Verfügbarkeit von Software zum Einsatz von Standardmethoden steigert dabei diesen Trend weiter, da auch die Verbreitung der Implementierungen als Qualitätsaussage fehlgedeutet werden kann. Deshalb sieht Guion (1999) den Bereich der Verwertung von diagnostischen Ergebnissen als evaluationsbedürftiger an als den Bereich der Erhebung, entgegen einer Fantasie der Letztbegündbarkeit der eigenen Entscheidungspraxis:

> [...] the Biblical Gideon did not need to evaluate the effectiveness of his tests because they were given by God, and too many people today seem to believe that their favorite HR applications are God-given and therefore need no further evaluation. (Guion, 1999, S. 363)

Weinert (1998) bezeichnet lineare Modelle zudem als Bestandteile des „traditionellen Selektionsmodells“, das in seiner Sicht auch an kurvilinearen Beziehungen zwischen Prädiktoren und Kriterium kranken kann. Bei einem sequenziellen Vorgehen ohne die Einräumung von Kompensationsmöglichkeiten kann eine einfache Cutoff-Strategie verfolgt werden: Nach jedem Test wird das Ergebnis mit einem Grenzwert verglichen und abhängig von dem Ergebnis des Vergleichs wird eine Entscheidung getroffen oder der nächste Test vorgegeben. Weinert (1998) bezeichnet die Methode als das Multiple Hürdensystem des sequenziellen Entscheidungsfällens. Eine solche Strategie kann natürlich auch nach Vorgabe aller Tests bei allen Kandidaten angewandt werden. Als Problem betrachtet Weinert (1998) die Festlegung von Cutoff-Werten:

> Die Problematik ist darin zu sehen, dass es schwierig sein wird, Minimalwerte (Cutoff-Werte) festzulegen, die dem Vorwurf der Willkür, der Subjektivität und Relativität entgehen. (S. 282)

Einen Lösungsvorschlag für eben dieses Problem zu unterbreiten, ist aber gerade Gegenstand der vorliegenden Arbeit. Guion (1991) bezeichnet ein Cutoff-Modell auch als *nicht-kompensatorisches Modell*. Er betrachtet dabei zwei Formen, die er als *conjunctive noncompensatory model* und *disjunctive noncompensatory model* bezeichnet: Das erste Modell trifft die Entscheidung für die Auswahl dann, wenn alle Cutoff-Werte überschritten sind, das zweite Modell trifft die Entscheidung, wenn mindestens ein Cutoff-Wert überschritten wird. Die in der Praxis eingesetzten Entscheidungsregeln weisen zum Teil eine recht illustre Geschichte auf, viele Prozeduren sind auf eher nicht-wissenschaftlichem Wege entstanden. Wottawa und Oenning (2002) zählen folgende Prinzipien zu Erklärungen für eine beobachtete Entscheidungspraxis:

- Das Lernen am Modell: Vom Vorgänger eingearbeitet, hat der Entscheider durch Nachahmung ein implizites Regelmodell von diesem übernommen.
- Die Übernahme institutioneller Normen: Liegen organisationale Regeln für die Auswahl vor, wird mit der Akzeptanz dieser Bedingungen entweder das Arbeitsergebnis eines Vorgängers übernommen (dessen Vorgehen wiederum ebenfalls in einem der Erklärungsansätze gespiegelt werden könnte) oder das Ergebnis eines unternehmensinternen Interessensausgleichs, der nicht notwendigerweise entscheidungsbezogen ausgehandelt worden sein muss.
- Gibt es kein Regelsystem zur Übernahme, so muss ein eigenes Regelsystem aufgebaut werden - oft ohne die Information über den Erfolg der gewählten Regeln - und ein Wiederholungslernen kann zum Ausblenden von Möglichkeiten der Validierung der eigenen Praxis führen, die sich zunehmend verfestigen kann und auf illusorischer Basis ein Vertrauen in die diagnostische Kompetenz des Entscheiders herausbilden lässt: Er kann schließlich Entscheidungserfahrung vorweisen, die er seinem Nachfolger vererben wird.

Mit der Forderung der Explikation von Expertenwissen durch die DIN 33430 wird bei Folgeleistung eine solche Praxis unterbunden. Die Explikation von Expertenwissen in Interaktion mit den Experten und die nachfolgende Verbesserung von explizierten Regelsystemen hat sich als erfolgreiche Maßnahme erwiesen (siehe zum Beispiel Wottawa & Echterhoff, 1982). Dabei fordern Wottawa und Oenning (2002) die Analyse und Modifikation von Einzelregeln bei Bereitstellung von Feedback zur Bewährung von Grenzwerten und Intervallen.

Was die Forderung nach Verständlichkeit von Regeln angeht, kann kompensatorischen Modellen vorgeworfen werden, dass sie keinen intuitiven Zugang zum Verständnis des Zustandekommens von Entscheidungsergebnissen ermöglichen, da mehrdimensionale Regressionsmodelle nicht nur in der Berechnung, sondern auch in der Ergebnisform als Rechenformel für nichtspezialisierte Anwender ein Hemmnis darstellen. Gleiches gilt für verschiedene Varianten der Regressionsrechnung, Diskriminanzanalysen, neuronale Netzwerke - hier ist der „Black Box“-Status der Modelle ein Hauptproblem (vgl. Woike, 2000) -, Verfahren der Fuzzy-Logik und andere komplexe Berechnungsverfahren. Wünschenswert sind hier hingegen

Regelsysteme, die aus Antezedenzbedingungen und Konsequenzen bestehen, die für den Laien einsichtig und nachvollziehbar sind (Wottawa & Oenning, 2002).

Für die Gewinnung von Regeln der beschriebenen Art kommen unter anderem sowohl konfigurale Verfahren als auch Entscheidungsbäume in Frage. Eine Induktion von Regeln bei kleinen Stichproben erweist sich hier aber in hohem Maße als unzuverlässig, so dass bei Nichtverfügbarkeit entsprechender Datenmengen die Norm nur durch die Explizierung von Expertenurteilen mit anschließender Validierung umgesetzt werden könnte. Langfristig ist aber eine Intensivierung der Maßnahmen zu befürworten, die die Entwicklung von Entscheidungsprozeduren durch die Erhebung empirischer Validierungsdaten begleiten. Wottawa und Oenning (2002) halten fest:

> Die Möglichkeiten der universitären psychologischen Forschung zum Bereich der Eignungsdiagnostik reichen bei weitem nicht aus, um den durch die Forderung der DIN 33430 nach Explikation und empirische [sic] Begründung der Regeln für die Eignungsbeurteilung auch nur annähernd zu entsprechen. Die dadurch geschaffenen Probleme sollten aber keineswegs dazu führen, die Forderungen der DIN 33430 abzulehnen. Nicht die Norm, sondern die bedauerliche Gewöhnung vieler im Bereich der Eignungsbeurteilung Tätigen, sich in vielen Bereichen mit einer letztlich nur sehr schwach fundierten empirischen Grundlage ihrer Eignungsbeurteilungen zufrieden zu geben, ist das Problem. Die DIN 33430 erzeugt nicht einen Missstand, sondern macht auf diese Problemlage mit ihren Forderungen aufmerksam. (S. 56)

Ein weiteres mögliches Argument gegen die Akzeptanz der Norm ist die Problematik der Explizierung intuitiver Entscheidungen: Ein veräußerlichter Entscheidungsalgorithmus kann nicht mehr durch Prozesse der Eindrucksbildung des Entscheiders beeinflusst werden. Die Frage, ob durch die Substitution von menschlichen Entscheidern durch statistische Regelsysteme substanzielle Einbußen in der Entscheidungsqualität zu erwarten sind, wird seit den Publikationen von Mehl in den Fünfziger Jahren diskutiert (Mehl, 1954, 1957), der in den meisten Studien eher einen Gewinn an Entscheidungsgenauigkeit fand. Problematisch ist insbesondere die Behandlung von Sonderfällen. Im Durchschnitt entscheidet ein statistisches Modell trotzdem in vielen Fällen besser als der Experte, wenn der Sonderfall nicht eine extreme Form annimmt: Das Ergebnis eines Weitsprungversuches lässt sich nach einem Beinbruch des Athleten vom Modell nur dann korrekt vorhersagen, wenn der Status seiner Knochen und Bewegunsfähigkeit in der Prädiktionsprozedur berücksichtigt wird. Eine Kontrolle der Ergebnisse von Entscheidungsprozeduren scheint also auf jeden Fall angebracht, wenngleich sich das Vertrauen auf das subjektive Gefühl von Unstimmigkeit ohne Präzisierungsmöglichkeit negativ auswirken kann. Kleinmuntz (1990) empfiehlt die Integration von statistischen und intuitiven Entscheidungsverfahren und stellt die Beobachtung an, dass aufgrund versteckter Kosten von Entscheidungshilfen oft kein anderer Ausweg bleibt, als den Kopf zu verwenden.

Gegenstand der vorliegenden Arbeit ist die Unterstützung der Verbindung zwischen empirischen Daten und der Explikation von Entscheidungsregeln. Vor der formalen Spezifizierung des statistischen Problems sollen einige konkrete Einsatzgebiete für Regelinduktionsverfahren in der Berufseignungsdiagnostik vorgestellt werden.

2.2.4 Beispiele für Entscheidungsregeln in computergestützten Testverfahren

Im Folgenden werden einige Beispiele für computergestützte Recruitingverfahren und die Notwendigkeit von Regelinduktion gegeben. Einen Überblick über frühere Beispiele der Computerdiagnostik mit komplexen Szenarios gibt Funke (1993), einen aktuelleren Überblick über wirtschaftsbezogene Varianten geben Sarges und Wottawa (2004a).

2.2.4.1 Anforderungsprofile und regelgestützte Auswahl mit ELIGO

Miesen et al. (1999) diagnostizieren eine „Technologie-Lücke“ zwischen Messinstrumenten und darauf basierenden Entscheidungen. Zusammen mit der Bindung der Durchführung psychologischer Tests an den Besitz eines Psychologie-Diploms erhalten lediglich Unternehmen auch nur die Möglichkeit einer auf fundierten Testverfahren beruhenden qualifizierten Praxis der Einstellungsentscheidungen, die sich die Einstellung eines Psychologen für Auswahlfragen leisten können. Das Computerprogramm ELIGO wurde mit dem Ziel einer Schließung dieser Lücke entwickelt. Es ermöglicht zunächst die Auswahl von

Testinstrumenten, die für die anstehende Entscheidung als relevant beurteilt werden. Es ist primär für mittlere Unternehmen (200 bis 500 Mitarbeiter) konzipiert, vor allem als Zwischenauswahlverfahren, und stellt 42 Dimensionen zur Verfügung, darunter Leistungs- und Persönlichkeitstests. Alle Testergebnisse werden als Perzentilwerte ausgegeben, so dass eine intuitive Bewertung der Resultate möglich wird. Perzentilwerte sind ordinalskaliert - die oben diskutierte Skalenniveauproblematik stellt sich bei sachgerechter Verwendung nicht - und die Ergebnisse sind im Unterschied zu Standardwerten, T-Werten oder IQ-Werten unabhängig von der Verteilungsform und vielleicht nur aus historischen Gründen (Probleme der Berechnung) bisher weniger verwendet worden (Wottawa & Oenning, 2002). Damit können sie einfacher interpretiert und im Sinne der DIN 33430 auch an Laien kommuniziert werden.

Für jede Stelle kann ein Anforderungsprofil im Sinne einer Intervallanagabe für akzeptable Werte auf den einzelnen Skalen angegeben werden. Da Perzentilwerte verwendet werden, entspricht die Angabe dementsprechend gleichzeitig auch einem Anteil von Personen der Normpopulation. Für die Setzung der Grenzwerte steht allerdings kein Werkzeug zur Verfügung:

> Die Grenzwerte können natürlich derzeit nur in wenigen Sonderfällen wissenschaftlich fundiert werden. Das ist kein ELIGO-spezifisches Problem, es fällt hier nur durch die Skalen besonders auf. Das eigentliche Problem ist ganz einfach: Es fehlen empirische Daten. Zunächst müssen Personalfachmann und Führungskraft sich darum auf ihre Erfahrung stützen. Sobald sie aber eine genügend große Zahl von Bewerbern mit ELIGO getestet haben, können sie ihr Anforderungsprofil an der Empirie ausrichten. (Miesen et al., 1999, S. 20)

Sarges und Wottawa (2004c) weisen auf eine positive Akzeptanz auch bei Betriebsräten hin.

2.2.4.2 Potenzialanalyse mit „pro facts“

Ein weiteres Testsystem zur Potenzialanalyse wird von Etzel und Küppers (2002) unter dem Namen „pro facts“ vorgestellt (vgl. Sarges & Wottawa, 2004b). Als Anwendungsbereiche werden die folgenden genannt:

- Personalmarketing und Vorauswahl, Personalauswahl
- Ergänzung in Assessment Centern, Potenzialanalyse
- Ermittlung des Trainingsbedarfs, Karriereberatung
- Unterstützung der systematischen Personal- und Organisationsentwicklung

Das System besteht aus 40 Bausteinen zur Erfassung persönlicher, sozialer und beruflicher Kompetenzen, die in ein fiktives aber realistisches Szenario eingebettet sind. Die Ergebnisse werden in sechstufigen Ergebnisprofilen präsentiert und können mit einem Sollprofil verglichen werden, das aus einer expliziten Angabe von Sollwerten für die einzelnen Ergebnisskalen besteht. Weiterhin kann eine Ergebnisdarstellung gewählt werden, die für jede Skala die relative Position des Kandidaten relativ zum Sollprofil sowie die Bedeutung der Skala für das Unternehmen aufträgt. Auf diese Weise können Trainingsbedarf, Potenziale und Kompetenzen ermittelt werden.

Bei der Darstellung von „pro facts“ verzichten die Autoren auf eine Explikation der Regeln für die Erstellung von Sollprofilen. Die sechsstufigen Skalen erscheinen weniger intuitiv als Perzentilwerte, und Sollprofile können nicht als Intervalle definiert werden.

2.2.4.3 Internet-Recruiting mit PERLS

PERLS ist ein Internet-Recruiting-Werkzeug, das auf dem ELIGO-System aufbaut und aus mehreren Komponenten besteht (vgl. De la Fontaine, Glas, Schafsteller & Wottawa, 2001; Kirbach & Montel, 2002):

- einer Datenbank aus ca. 100 Testskalen
- einer Möglichkeit der Rückmeldung von Testergebnissen mit optionalen Interpretationen und Empfehlungen

- einem integrierten Bewerbertracking bspw. zur Kontrolle von Mehrfachbewerbungen
- der Bereitstellung einer Datenbank mit Testergebnissen und Suchfiltern
- integrierten EDV-Werkzeugen zur Durchführung und Auswertung von unternehmensspezifischen Wettbewerbsspielen (*Recruitainment*)
- einem Inventar für Kunden- und Mitarbeiterbefragungen („Modulares Inventar zur Organisationsdiagnose (MIO)“).

Die Testverfahren sind dabei unterteilt in „freie“ Testverfahren, die unter wenig oder nicht kontrollierten Bedingungen im Netz verfügbar gemacht werden können, „halbgeschützten“ Verfahren, für die dem Bewerber ein Zugang zur Verfügung gestellt werden muss und „geschützten“ Testverfahren, die an die Einrichtung von Testräumen mit Standardbedingungen gebunden sind. Mit diesen drei Klassen von Verfahren werden unterschiedliche Ziele verfolgt: Die freien Verfahren ermöglichen das Üben oder geben Möglichkeiten für das Self-Assessment, weiterhin können Informationen über Unternehmen oder Vakanzen mitgeteilt werden. Gleichzeitig wird - etwa bei einer Bereitstellung im Intranet - auch die Möglichkeit einer Bewerber-Selbstselektion ohne die Gefahr des Gesichtsverlusts geboten (Kirbach & Montel, 2002). Die halbgeschützten Verfahren können durch Fehlen der Möglichkeit von Teilnahmewiederholung sinnvolle Daten generieren, die fehlende Möglichkeit der Identitätsfeststellung lässt aber lediglich ein Negativ-Screening aufgrund der Ergebnisse zu. Wird mit der Testung mehr intendiert, ist die Vorgabe unter Standardbedingungen unverzichtbar. Ein integrierter Kommunikationsmanager ermöglicht zudem Analyse, Ergebniszusammenfassung und selektive Rückmeldungen via Email (Sarges & Wottawa, 2004d).

Im Jahr 2000 wurde im Rahmen des von der Firma Siemens angebotenen Wettbewerbsspiel „Challenge Unlimited“ eine erste große Internet-Recruiting-Aktion mit PERLS durchgeführt (vgl. Wild, De la Fontaine & Schafsteller, 2001), bei dem 13 000 Teilnehmer Spielelemente und Testkomponenten absolvierten, die zu einem großen Anteil aufgrund von Alter und Hintergrund für das Unternehmen interessant waren:

> Mit diesem Projekt wurde somit **sowohl die technische Machbarkeit** psychologiegestützter Internet-Recruiting-Tools **als auch die Akzeptanz** solcher Instrumente bei den Bewerbern nachgewiesen. (Kirbach & Montel, 2002, S. 40)

Ein ähnliches Tool wie „Challenge Unlimited“ stellt der mit dem „Monster Award“ ausgezeichnete „Allianz Voyager“ dar, bei dem Teilnehmer verschiedene Events im Umfeld des Formel-1-Sponsorings der Allianz planen sollten, womit das Arbeitsumfeld der Zielpositionen erlebbar gemacht werden konnte (Wottawa et al., 2005).

Weitere Einsatzgebiete aufgrund der Integration der enthaltenen Komponenten sind die Umberatung von Bewerbern, wenn etwa festgestellt wird, dass ein Profil für die avisierte Stelle nicht passend erscheint, aber für alternative Optionen geeignet wirkt, sowie die Einrichtung von direkten Chats mit Recruitern und einer adäquaten, psychologie-gestützten Ergebnisrückmeldung mit eventuell folgender Nachbetreuung.

Dies soll den kurzen Überblick über ein Anwendungsfeld von Entscheidungsregeln abschließen, so dass im Folgenden das in dieser Arbeit bearbeitete Problem erläutert werden kann.

2.3 Charakterisierung des statistischen Regelinduktions-Problems

2.3.1 Die erste Form des statistischen Problems

Das von der Software $\textsc{Gier}_{[1.0]}$bearbeitete statistische Problem lässt sich in folgender Art und Weise charakterisieren. Gegeben sei eine Menge $O = \{O_1, O_2, \ldots O_n\}$ von Objekten, für die jeweils k Prädiktorwerte auf k Prädiktorvariablen $X_1, X_2, \ldots X_k$ vorliegen, mit $X_{i,j} \in \mathbb{N}$ und $0 \leq X_{i,j} \leq 100$ für $1 \leq i \leq k$ und $1 \leq j \leq n$. Zudem sei für jedes Objekt die Ausprägung der Kriteriumsvariablen Y bekannt mit $Y_j \in \{0, 1\}$ für $1 \leq j \leq n$. Gesucht ist eine Menge hierarchischer Regeln $R = \{R_1, R_2, \ldots, R_m\}$ mit folgenden Eigenschaften:

1. Die Antezedenzbedingung jeder Regel besteht aus einer Konjunktion von bis zu k Relationen der Form $u_i \leq X_{i,j} \leq o_i$ mit $1 \leq i \leq k$ und $o_i, u_i \in [0; 100]$.

2. Ist die Antezedenzbedingung erfüllt, wird das Objekt i als Konsequenz in eine der beiden Vorhersageklassen 0 bzw. 1 eingeordnet, also $O_i \rightarrow \{0,1\}$ bzw. $P_i = c_h$ mit $c_h \in \{0,1\}$ und $1 \leq h \leq m$.

Für jedes Objekt i tritt die Regel mit dem kleinsten Index in Kraft, für die die Antezedenzbedingung erfüllt ist. Ist keine der Antezedenzbedingungen erfüllt, wird dem Objekt die Vorhersage $P_i = -1$ zugeordnet und es gilt als nicht klassifiziert, in jedem anderen Fall gilt es als der Klasse P_i zugeordnet. Das Regelsystem soll nun so gewählt werden, dass die Vorhersage möglichst gut der tatsächlichen Klassenzugehörigkeit entspricht, also möglichst wenig Klassifikationsfehler gemacht werden:

$\min \sum_{j=1}^{n} t_j$, mit $t_j = \begin{cases} 0 & P_{i,j} = Y_j \\ 1 & P_{i,j} \neq Y_j \end{cases}$.

Damit ist das statistische Problem charakterisiert.

2.3.2 Modifikation des statistischen Problems

Dieses Problem hat zunächst einmal eine optimale Lösung: Für eine Menge von Objekten können maximal so viele Regeln sinnvoll aufgestellt werden, wie Objekte vorhanden sind. Werden mehr Regeln aufgestellt, so würden die zusätzlichen Regeln niemals zur Anwendung kommen. Sind allerdings Profile von Objekten identisch, ist also für zwei Objekte i und j die Bedingung $(\forall h)(1 \leq h \leq k \rightarrow X_{h,i} = X_{h,j})$ erfüllt, so erfüllen beide Objekte die gleiche Teilmenge von Antezedenzbedingungen, sie sind für ein Regelsystem der beschriebenen Art ununterscheidbar. Dementsprechend reduziert sich die Anzahl potenziell sinnvoller Regeln auf die Anzahl der distinkten Prädiktorprofile. Wird jetzt für jedes distinkte Prädiktorprofil eine Regel aufgestellt mit einer Antezedenzbedingung, die für ein Objekt j für jedes Merkmal i sowohl $u_i = X_{i,j}$ als auch $o_i = X_{i,j}$ setzt, so wird für jedes Objekt genau eine Antezedenzbedingung erfüllt. Wird jetzt also für jede Gruppe von Objekten mit ununterscheidbaren Prädiktorprofilen in den Regeln mit erfüllter Antezedenzbedingung der Modus der Kriteriumswerte der Objekte dieser Gruppe zugeordnet (bzw. eine zufällige oder beliebige Klasse bei gleichen Häufigkeiten), so kann kein hierarchisches Regelsystem das so bestimmte Regelsystem übertreffen. Jedes Verfahren, das die Entscheidung aufgrund der Prädiktorinformationen deterministisch trifft, bleibt durch die gleiche Grenze gebunden. Bei jedem Regelsystem dieser Art wäre die Reihenfolge der Regeln irrelevant, da jedes Objekt nur eine der Regeln erfüllt. Das Regelsystem ist allerdings weder notwendigerweise eindeutig (manche Intervalle könnten ohne Konsequenz vergrößert werden) noch hat es die minimale Regelanzahl für die optimale Klassifikationsleistung, unter Umständen gibt es allerdings Regelsysteme mit weniger Regeln und gleicher Klassifikationsgleichung, da Regeln mit gleicher Konsequenz zusammengefasst werden könnten. Ein Beispiel soll diese optimale Zuordnung veranschaulichen:

Gegeben seien die Objekte o_1, o_2, o_3, o_4 und o_5 bei drei Merkmalen mit $X_1 = (22, 98, 22, 22, 22)$, $X_2 = (35, 14, 35, 97, 35)$ und $X_3 = (11, 57, 11, 88, 11)$ und den Kriteriumswerten $Y = (1, 0, 1, 1, 0)$. In diesem Fall würde ein Regelsystem aus drei Regeln zur optimalen Klassifikation konstruiert werden:

$$\begin{aligned}
R_1: &\quad (22 \leq X_{1,j} \leq 22) \wedge (35 \leq X_{2,j} \leq 35) \wedge (11 \leq X_{3,j} \leq 11) \rightarrow P_j = 1 \\
R_2: &\quad (98 \leq X_{1,j} \leq 98) \wedge (14 \leq X_{2,j} \leq 14) \wedge (57 \leq X_{3,j} \leq 57) \rightarrow P_j = 0 \\
R_3: &\quad (22 \leq X_{1,j} \leq 22) \wedge (97 \leq X_{2,j} \leq 97) \wedge (88 \leq X_{3,j} \leq 88) \rightarrow P_j = 1.
\end{aligned}$$

Die erste Regel betrifft hier drei Objekte, von denen zwei der Kriteriumsklasse $Y_j = 1$ angehören, deshalb wird für alle drei Objekte diese Vorhersage gemacht, ein Objekt wird dabei also falsch klassifiziert. Die gleiche Klassifikation wie durch die drei Regeln oben könnte aber auch durch die folgenden beiden Regeln erreicht werden:

$$\begin{aligned}
R_1: &\quad (22 \leq X_{1,j} \leq 22) \wedge (35 \leq X_{2,j} \leq 97) \wedge (11 \leq X_{3,j} \leq 88) \rightarrow P_j = 1 \\
R_2: &\quad (98 \leq X_{1,j} \leq 98) \wedge (14 \leq X_{2,j} \leq 14) \wedge (57 \leq X_{3,j} \leq 57) \rightarrow P_j = 0.
\end{aligned}$$

Dieses System ist wiederum für die betrachtete Objektmenge im Ergebnis äquivalent zu folgendem System:

$$R_1: \quad (0 \leq X_{1,j} \leq 97) \wedge (0 \leq X_{2,j} \leq 100) \wedge (0 \leq X_{3,j} \leq 100) \rightarrow P_j = 1$$

$R_2: \quad (98 \leq X_{1,j} \leq 98) \wedge (14 \leq X_{2,j} \leq 14) \wedge (57 \leq X_{3,j} \leq 57) \rightarrow P_j = 0.$

Formal gibt es zunächst keinen Grund, eine der Lösungen zu präferieren, da weder Effizienz noch Intervallgrößen im Zielkriterium enthalten sind. Es gibt an dieser Stelle mindestens drei Argumente, um der beschriebenen optimalen Lösung kritisch gegenüber zu stehen und das Attribut „optimal“ zu qualifizieren.

1. Das beschriebene Regelsystem ist optimal in der Vorhersage von Klassenzugehörigkeiten, wenn ein zu klassifizierendes Objekt aus der Menge gewählt wird, die zur Erstellung des Regelsystems genutzt wurde (der Trainingsmenge). Optimal ist also die *Fitting*-Leistung des Systems. Über den Erfolg bei Objekten, die nicht in der Trainingsmenge waren, kann nichts gesagt werden. Das Attribut „optimal“ muss also hinsichtlich der Aufgabe (*Fitting*) qualifiziert werden, da es Systeme geben könnte, die bei schlechterer oder äquivalenter *Fitting*-Leistung bessere Ergebnisse bei neuen Objekten erzielen, also die eine bessere Generalisierungs-Leistung demonstrieren können.

2. Neue Objekte können nur dann klassifiziert werden, wenn ihr Profil exakt dem Profil eines der Objekte aus der Trainingsmenge entspricht. Das vollständige Regelsystem in der beschriebenen Form kann nicht extrapolieren. Die reduzierten Regelsysteme hingegen sind dazu in der Lage, allerdings mit ungeklärtem Erfolg. Das beschriebene System ist dementsprechend hinsichtlich seiner Anwendbarkeit bei Generalisierungen möglicherweise suboptimal, da über große Bereiche des Profilraums überhaupt keine Aussagen gemacht werden können. Eine zusätzliche Regel für Restfälle, also mit einer immer erfüllten Antezedenzbedingung, kann dieses Problem nur unbefriedigend lösen. „Optimal“ muss also auch in dieser Hinsicht spezifiziert werden.

3. Das Konstruktionsprinzip erzeugt eine Menge von Regeln, die mit jedem neuen Objektprofil wächst. Dementsprechend wird zum einen der Speicheraufwand in der Konstruktion sehr hoch, zum anderen bedeutet dies für die Implementierung des Systems, dass bei n Objekten bis zu n Antezedenzbedingungen überprüft werden müssen, um eine Entscheidung zu treffen. Bei d distinkten Prädiktorprofilen werden für jedes Objekt durchschnittlich $\frac{d+1}{2}$ Abfragen notwendig, wobei dieser Wert bei identischen Objekten verkleinert werden kann, wenn Regeln nach vorn gestellt würden, die mehr Objekte betreffen als andere. Trotzdem ist das beschriebene System in der beschriebenen Hinsicht ebenfalls nicht als „optimal“ zu bezeichnen.

Es bietet sich hinsichtlich der vorgestellten Argumente also an, dem vorgestellten „optimalen“ System skeptisch gegenüber zu stehen, auch wenn es die formalen Anforderungen „optimal“ erfüllt. Genauer gesagt sind es also die formalen Anforderungen, die nicht zufrieden stellen. Es ist für das Zielsystem durchaus gewünscht, sowohl in der Anzahl der Regeln überschaubar zu bleiben als auch in der Lage zu sein, Vorhersagen für zukünftig beobachtete Objekte zu machen. Gleichwohl ist über die Struktur zukünftiger Anforderungen im Allgemeinen wenig bekannt, so dass sich das Dilemma ergibt, dass für die gewünschte Anwendung prinzipiell keine optimale Lösung existiert, wenn nicht deterministische Annahmen über zukünftige Beziehungen getroffen werden sollen. Dementsprechend wird bereits an dieser Stelle eine heuristische Methode verwendet, wenn die formale Spezifikation um folgende Bedingung ergänzt wird:

Bedingung der begrenzten Regelzahl: Die Anzahl der Regeln im System wird auf eine obere Grenze m_g festgelegt, die kleiner ist als die Anzahl der distinkten Profile in der Trainingsstichprobe.

Diese Einschränkung hat zur Folge, dass die im System verbleibenden Regeln genereller sein müssen, damit sie mehr Prädiktorprofile klassifizieren können. Zudem wird das entstehende System handhabbarer. Eine unangenehme Konsequenz dieser Entscheidung besteht allerdings in der beträchtlichen Erweiterung des Suchraums für legitime Regeln: Wird in den Antezedenzbedingungen die untere Grenze auf den Wert der oberen gesetzt, so gibt es bei 101 möglichen Werten pro Merkmal 101 mögliche Festlegungen und damit pro Regel bei k Merkmalen 101^k Kombinationen. Für m Regeln gibt es somit $101^{m^k} = 101^{mk}$ Kom-

binationen[11]. Lässt man hingegen alle erfüllbaren Antezedenzbedingungen zu, so muss lediglich gelten, dass die untere Grenze kleiner oder gleich der oberen ist. Bei 101 Werten gibt es dafür $\frac{101}{2}(101+1) = 5151$ Möglichkeiten, also das 51-fache der Möglichkeiten pro Regel pro Merkmal. Dementsprechend erhöht sich die Anzahl der zu berücksichtigenden Gesamtmöglichkeiten um den Faktor 51^{mk}. Bei nur zwei Merkmalen und drei Regeln gäbe es bspw. im ersten Fall bereits 1 061 520 150 601 Gesamtmöglichkeiten, im zweiten Fall hingegen 1 867 881 4076 536 059 118 401. Bei einem realistischen Anwendungsfall von acht Merkmalen und zehn Regeln sind die Verhältnisse ungefähr $2,217 \cdot 10^{160}$ zu $8,940 \cdot 10^{296}$. Schon an dieser exemplarischen Betrachtung wird plastisch deutlich, dass selbst Rechnersysteme mit zeitgenössischer Geschwindigkeit nicht alle möglichen Kombinationen untersuchen können. Da aber kein analytisches Verfahren zur Verfügung steht, das das Problem in ähnlicher Weise lösen könnte wie es für das ursprüngliche Problem möglich war, müssen auch für die Lösung der neu gestellten Aufgabe heuristische Methoden zum Einsatz kommen, um den Suchraum effizient bearbeiten zu können, da keine weiteren Vereinfachungen der Aufgabenstellung intendiert sind. Die Software GIER$_{[1.0]}$stellt zu diesem Zweck einige Algorithmen zur Verfügung.

2.3.3 Abgrenzung des statistischen Problems von verwandten Problemen

Das vorgestellte formale Problem unterscheidet sich unter anderem in zweierlei Hinsicht von verwandten Problemen der Klassifikation:

1. Unterschiede hinsichtlich der Segmentierung des Suchraums
2. Unterschiede hinsichtlich des Klassifikationskriteriums

Diese beiden Abweichungsmöglichkeiten sollen nacheinander vorgestellt werden.

2.3.3.1 Segmentierung des Prädiktorprofilraums

Zur Veranschaulichung der Segmentierung des Suchraums durch Regelsysteme der vorgeschlagenen Art hilft folgende Überlegung. Jeder Teil der Konjunktion in erfüllbaren Antezedenzbedingungen legt ein Intervall auf einer Prädiktorvariablen fest, das im Extremfall zu einem Punkt zusammenfällt. Stellt man sich den Prädiktorprofilraum als einen k-dimensionalen euklidischen Raum vor, in dem jedes Prädiktorprofil als ein Punkt abgebildet wird, dessen Koordinaten den k Variablenausprägungen zugeordnet wird, so wird durch jede Konjunktion ein maximal k-dimensionaler Raum festgelegt, in dem alle Prädiktorprofile die Antezedenzbedingung erfüllen. Da nachfolgende Regeln frühere Entscheidungen nicht revidieren können, kann der durch eine Regel abgedeckte Raum keiner nachfolgenden Regel mehr zugeordnet werden. Für den zweidimensionalen Fall könnte damit die in Abbildung 2.3 dargestellte Zuteilung durch die folgenden vier Regeln vorgenommen werden:

$$
\begin{aligned}
R_1: &\quad (10 \leq X_{1,j} \leq 90) \wedge (40 \leq X_{2,j} \leq 70) \rightarrow P_j = 1 \\
R_2: &\quad (43 \leq X_{1,j} \leq 59) \wedge (67 \leq X_{2,j} \leq 89) \rightarrow P_j = 0 \\
R_3: &\quad (22 \leq X_{1,j} \leq 49) \wedge (16 \leq X_{2,j} \leq 52) \rightarrow P_j = 0 \\
R_4: &\quad (03 \leq X_{1,j} \leq 85) \wedge (04 \leq X_{2,j} \leq 98) \rightarrow P_j = 1.
\end{aligned}
$$

Dieses Beispiel zeigt, dass Segmentierungen so vorgenommen werden können, dass Areale, die der gleichen Klasse zugeordnet werden, komplett von Arealen unterbrochen werden können, die einer anderen Klasse zugeordnet werden können, d.h. es gibt Regelsysteme, die auch dann Objekte erfolgreich zuordnen können, wenn die Klassen nicht linear separierbar sind. Das unterscheidet die beschriebene Segmentierung von denen, die durch die lineare Diskriminanzanalyse oder einige Formen neuronaler Netzwerke wie einfache Perzeptronen gebildet werden können. In dieser Hinsicht zeigt sich aber eine Ähnlichkeit zu Entscheidungsbäumen, die aufgrund einfacher Ungleichungen Entscheidungen treffen. Das dargestellte

[11] Hier soll berücksichtigt werden, dass Perzentilwerte auf verschiedene Weise kodiert werden können: dementsprechend ist das Intervall von 0 bis 100 vollständig zugelassen. Streng genommen gibt es 99 Perzentilwerte, die damit 100 Gruppen voneinander trennen. Wird die Überschreitung betont, so bieten sich für die Gruppen die Werte von 0 bis 99 an, wird die Unterschreitung betont, die Werte von 1 bis 100.

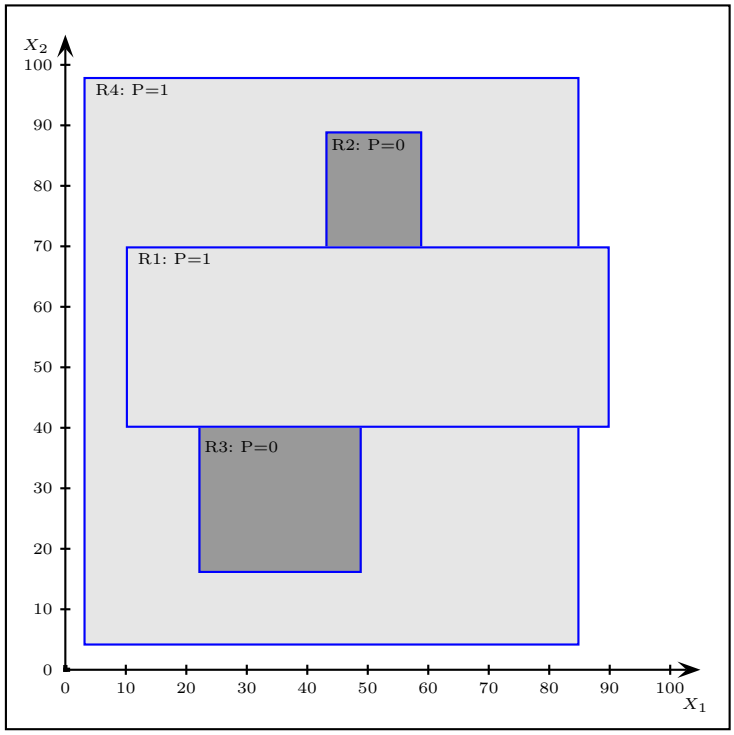


Abbildung 2.3: Beispielsegmentierung durch vier Regeln: Nachfolgende Areale werden von vorangehenden Regeln überdeckt

Regelsystem ist in den Ergebnissen äquivalent zu dem in Abbildung 2.4 dargestellten Entscheidungsbaum. Da durch hierarchische Regelsysteme prinzipiell für jedes Prädiktorprofil unabhängig Voraussagen getroffen werden können, gibt es dementsprechend keine Segmentierung, die nicht als hierarchisches Regelsystem repräsentiert werden könnte. Gleiches gilt für Entscheidungsbäume, da durch entsprechende Abfragen Endknoten genauso spezifisch für einzelne Prädiktorprofile gemacht werden könnten. In diesem Punkt unterscheiden sich die Ansätze auch von zahlreichen kompensatorischen Verfahren, die Prädiktorwerte zur Gewinnung der Vorhersage integrieren. In der bereits erwähnten Diskriminanzanalyse können Werte eines Prädiktors durch Kombinationen von Werten anderer Prädiktoren in den meisten Fällen ausgeglichen werden, Regelsysteme hingegen bieten die Möglichkeit, aufgrund eines einzelnen Wertes bereits Entscheidungen zu treffen. Zudem sind Interaktionen zwischen Variablen (wie im Beispiel) einfach zu repräsentieren. Faktisch werden hingegen nicht alle möglichen Segmentierungen zu erreichen sein, da dies aufgrund der Regelzahlbeschränkung und der Verwendung eines heuristischen Ansatzes verhindert werden kann. Auch bei Entscheidungsbäumen gibt das gewählte Konstruktionsverfahren (CART, CHAID, QUEST, ID3, etc.) Grenzen vor.

Hierarchische Regelsysteme unterscheiden sich in ihrer Struktur auch vom *SOMC*-Ansatz, der von Montel (2005) im Rahmen der Anwendung von ELIGO-Testverfahren vorgestellt wird. Hier werden für jede Variable getrennt Intervalle vorgegeben, dabei wird für jedes Objekt die Anzahl der Abweichungen von Prädiktorwerten hinsichtlich dieser Intervalle über alle Prädiktoren gezählt, was bei einheitlich gewichteten k Prädiktoren zu Ergebniswerten zwischen 0 und k führt. Zur Klassifizierung können so unterschiedliche Cut-Offs bezüglich der Summenwerte angesetzt werden, je nach den Gegebenheiten der Situation. Zur Schätzung der Intervallgrenzen stellt Montel mit dem Verfahren zur *simultanen Optimierung multipler Cut-Offs (SOMC)* ebenfalls einen genetischen Algorithmus vor, dessen Klassifikationsleistung hinsichtlich dreier Stichproben mit klassischen Auswahlverfahren verglichen wird (s.u.).

2.3.3.2 Klassifikationsgütekriterien

Konventionell wird in der Bewertung von Klassifikationsergebnissen vom Einzelfall abstrahiert, die beiden Gruppen der Partitionierung durch das Kriterium werden nicht weiter binnendifferenziert, so dass Mitglieder beider Gruppen als für die Bewertung des Ergebnisses als äquivalent angesehen werden. Dies ist keine offensichtliche Entscheidung und entspricht nicht unbedingt der gängigen Praxis in allen Anwendungs-

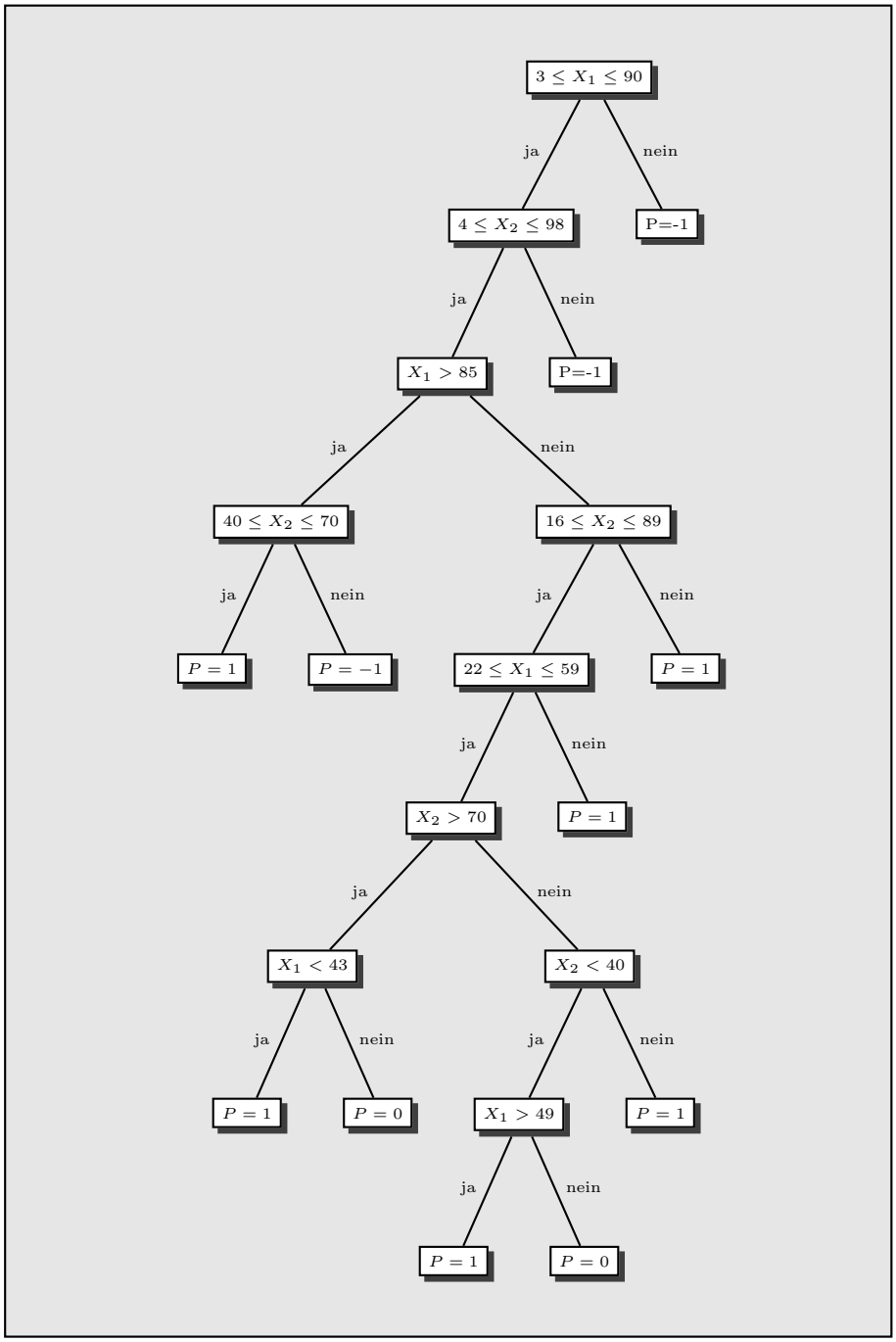


Abbildung 2.4: Ergebnisäquivalenter Entscheidungsbaum zum Beispielregelsystem: Der Baum segmentiert den zweidimensionalen Prädiktorprofilraum in gleicher Weise wie das Regelsystem

Tabelle 2.1: Konfusionsmatrix: Benennug der vier Klassifikationsmöglichkeiten

		P_j	
		1	0
Y_j	1	n_{11}	n_{10}
	0	n_{01}	n_{00}

dabei:

$$n_1 = n_{11} + n_{10}$$
$$n_0 = n_{01} + n_{00}$$

feldern. Hier kann durchaus beobachtet werden, dass Fehlentscheidungen bei bestimmten Teilgruppen als gravierender betrachtet werden oder die Zuordnung von Teilgruppen zu bestimmten Kategorien als bedeutender angesehen wird. Im Rahmen dieser Arbeit soll die Betrachtung allerdings auf Anwendungen beschränkt werden, in denen makro- oder mikropolitische Absichten oder individuelle Vorlieben und Differenzierungen auf die Auswahl der Prädiktoren, Benennung der Kriteriumsklassen und vorgegebene Zuordnung von Fällen der Stichprobe zu Kriteriumsklassen beschränkt bleiben. Das Thema möglicher Diskriminierung durch Klassifikation soll nicht im Fokus der Untersuchung stehen. Dementsprechend soll die (bereits von den Einzelentscheidungen abstrahierte) Konfusionsmatrix den Ausgangspunkt für die Beurteilung von Klassifikationsleistungen darstellen. Das gewählte Klassifikationsgütekriterium besteht aus einem einfachen Abzählen der Klassifikationsfehler. Wählt man die in Tabelle 2.1 dargestellten Benennungen, so läuft das gewählte Kriterium auf eine Minimierung der Summe $n_{10}+n_{01}$ hinaus. Diese Summe ist proportional zur Fehlerquote und negativ proportional zur Klassifikationsgenauigkeit, wenn alle Fälle klassifiziert werden können: Als Klassifikationsgenauigkeit A, die auch als Trefferquote bezeichnet wird, soll folgender Quotient definiert werden:

$$A = \frac{n_{11} + n_{00}}{n_{11} + n_{10} + n_{01} + n_{00}} \tag{2.6}$$

Die Minimierung der Fehleranzahl entspricht bei vollständiger Klassifikation der Maximierung der Klassifikationsgenauigkeit. Dabei wird kein Unterschied gemacht hinsichtlich der Art der Fehler. Falsche Zuweisungen zu den beiden Kategorien werden gleich behandelt, die relative Größe der beiden Klassen bleibt unberücksichtigt.

Tabelle 2.2: Beispielklassifikationsergebnis 1

		P_j	
		1	0
Y_j	1	75	25
	0	50	25

Tabelle 2.3: Beispielklassifikationsergebnis 2

		P_j	
		1	0
Y_j	1	100	0
	0	75	0

Dies hat die Konsequenz, dass die beiden in den Tabellen 2.2 und 2.3 dargestellten Ergebnisse zweier fiktiver Klassifikationsverfahren als gleich gut bezeichnet werden müssen:

$$A_1 = \frac{75 + 25}{75 + 25 + 50 + 25} = \frac{4}{7} = \frac{100 + 0}{100 + 0 + 75 + 0} = A_2 \tag{2.7}$$

Im zweiten Fall wurde jedes Objekt der gleichen Klasse zugeordnet, unabhängig von der Ausprägung der Prädiktorwerte, im ersten Fall wurde stärker differenziert. Ein Ausweg aus dem Dilemma der Gleichbehandlung von Ungleichem besteht in der Aufspaltung der Klassifikationsleistung in zwei verschiedene Kennwerte. Dabei beschreibt der eine Kennwert die Genauigkeit der Vorhersage für Objekte mit einem Kriteriumswert von 1, der andere Kennwert die Genauigkeit für Objekte mit einem Kriteriumswert von 0 (vgl. Hollmann, 1991). Es ist eine Konvention, die mit 1 kodierte Klasse als Zielgruppe zu betrachten, aus diesem Grund wird der erste Kennwert auch *Sensitivität* genannt, was also die Fähigkeit bezeichnet, Objekte mit „positivem“ Kriteriumswert zu identifizieren. Der zweite Kennwert wird auch *Spezifität* genannt, was dementsprechend die Eigenschaft bezeichnet, nicht fälschlicherweise Objekte mit „negativem“ Kriteriumswert der positiven Klasse zuzuordnen. Besteht in der Art der Gruppen kein qualitativer Unterschied, so läuft es auf eine offene Entscheidung hinaus, welche Gruppe den positiven Kriteriumswert erhält. Dementsprechend soll hier die Sensitivität definiert werden als:

$$S^+ = \frac{n_{11}}{n_{11} + n_{10}}, \tag{2.8}$$

die Spezifität als

$$S^- = \frac{n_{00}}{n_{00} + n_{01}}. \tag{2.9}$$

Für die Beispieltabellen gilt hinsichtlich der Sensitivität $S_1^+ = \frac{75}{75+25} = 0,75$ bzw. $S_2^+ = \frac{100}{100+0} = 1$, hinsichtlich der Spezifität $S_1^- = \frac{25}{25+50} = \frac{1}{3}$ bzw. $S_2^- = \frac{0}{0+75} = 0$. In Abschnitt 2.4 soll eine Möglichkeit vorgestellt werden, wie die Kennwerte visuell miteinander in Bezug gesetzt werden können.

Die Angabe von zwei Kennwerten für die Güte einer Zuordnung löst allerdings nicht das Problem der differenziellen Bewertbarkeit von vorgeschlagenen Lösungen: Das erste Verfahren im Beispiel hat eine schlechtere Sensitivität, dafür eine höhere Spezifität als das zweite Verfahren. Zwar ist die Differenz bezüglich der Spezifität größer, aber es wurde bereits berechnet, dass die Genauigkeit der Zuordnungen gleich ist. Eindeutige Entscheidungen lassen sich im Vergleich zweier Verfahren allerdings für den Fall treffen, dass sowohl die Sensitivität als auch die Spezifität eines Verfahrens beide nicht schlechter sind als die eines Vergleichverfahrens und mindestens einer der Kennwerte besser ist (vgl. Amelang & Zielinski, 2002). In diesem Fall könnte man davon sprechen, dass das erste Verfahren das zweite in seiner Klassifikationsleistung dominiert, formal wäre das durch folgende Bedingung erfüllt:

$$(S_1^+ > S_2^+ \vee S_1^- > S_2^-) \wedge \neg(S_1^+ < S_2^+ \vee S_1^- < S_2^-) \tag{2.10}$$

.

Für den Fall $S_1^+ = S_2^+ \wedge S_1^- = S_2^-$ bei der gleichen Klassifikationsaufgabe muss die Klassifikationstabelle für beide Verfahren gleich sein, da n_0 und n_1 in beiden Fällen gleich sind und aus Sensitivität und Spezifität alle vier Werte der Matrix rekonstruiert werden können. Dies würde allerdings nicht beinhalten, dass jedes Objekt gleich klassifiziert wird, denn zwei gegenläufige Abweichungen gleichen sich in der Matrix aus.

Eine analoge Betrachtung ergibt sich, wenn als Basis für die Aufteilung der Fehlerarten nicht die tatsächlichen Kriteriumswerte, sondern die Vorhersagen gewählt werden. Ein Wert, der Auskunft über die Güte einer positiven Vorhersage gibt, ist der Kennwert

$$V^+ = \frac{n_{11}}{n_{11} + n_{01}}, \tag{2.11}$$

also der Anteil der korrekt positiv klassifizierten Objekte an den positiv klassifizierten Objekten. Dieser Kennwert soll in dieser Arbeit als *positive Validität* definiert werden. Andere Benennungen für diesen Wert sind bspw. „selektiver Eignungsquotient“ (Hollmann, 1991; Amelang & Zielinski, 2002), „Eignungsquotient“ (Montel, 2005) oder „prädiktiver Wert der positiven Zuordnung“. Ein entsprechender Wert für negative Vorhersagen ergibt sich aus

$$V^- = \frac{n_{00}}{n_{00} + n_{10}}, \tag{2.12}$$

der hier als *negative Validität* bezeichnet werden soll und auch als „prädiktiver Wert der negativen Zuordnung“ diskutiert wird. Positive und negative Validität werden in diesem Kapitel zugunsten von Sensitivität und Spezifität vernachlässigt, sie werden aber in den Verfahren in Kapitel 5 eine Bedeutung haben.

In der Literatur werden verschiedene Kennwerte diskutiert, um eine Entscheidbarkeit hinsichtlich der Präferenz zwischen Klassifikationslösungen zu erzwingen. (Montel, 2005) entscheidet sich für den Δ_{UAPV}-Score (UAPV steht hier für „*U*npassende *a*bgewiesen *P*assende *v*erloren“). Dieser Wert berechnet sich als

$$\Delta_{UAPV} = \frac{n_{11}n_{00} - n_{10}n_{01}}{(n_{11} + n_{10})(n_{01} + n_{00})} \tag{2.13}$$

.

Dabei ergibt sich die Beziehung

$$\begin{aligned}\Delta_{UAPV} =& \frac{n_{11}n_{00} - n_{10}n_{01}}{(n_{11} + n_{10})(n_{01} + n_{00})} = \frac{n_{00}}{n_{01} + n_{00}} - \frac{n_{10}}{n_{11} + n_{10}} = \frac{n_{00}}{n_{01} + n_{00}} - (1 - \frac{n_{11}}{n_{11} + n_{10}}) \\ =& \frac{n_{00}}{n_{01} + n_{00}} + \frac{n_{11}}{n_{11} + n_{10}} - 1 = S^- + S^+ - 1.\end{aligned} \tag{2.14}$$

.

Der Vergleich zweier Algorithmen läuft also auf den Vergleich der jeweiligen Summen von Sensitivität und Spezifität hinaus. Die Kennzahl wird durch die Subtraktion von 1 auf einen Wertebereich zwischen -1

Tabelle 2.4: Beispielklassifikationsergebnis 3

		P_j	
		1	0
Y_j	1	750000	250000
	0	1	4

Tabelle 2.5: Beispielklassifikationsergebnis 4

		P_j	
		1	0
Y_j	1	949999	50001
	0	2	3

und 1 festgelegt. Im Beispiel ist $\Delta_{UAPV1} = \frac{75 \cdot 25 - 25 \cdot 50}{(75+25) \cdot (50+25)} = \frac{625}{7500} = S_1^+ + S_1^- - 1 \approx 0,75 + 0,33 - 1 = 0,083$ und $\Delta_{UAPV1} = 1 + 0 - 1 = 0$. Verwendet man dieses Kriterium, so wird die erste Strategie als besser eingestuft.

Ein Problem mit diesem Koeffizienten tritt allerdings auf, wenn man sich vergegenwärtigt, dass die Leistung des Klassifikationsalgorithmus mit dem Ergebnis in Tabelle 2.4 als besser eingestuft wird als die Leistung des Klassifikationsalgorithmus mit dem Ergebnis in Tabelle 2.5: $\Delta_{UAPV3} = \frac{750000}{750000+250000} + \frac{4}{4+1} - 1 = 0,75 + 0,8 - 1 = 0,55 > 0,549999 = 0,949999 + 0,6 - 1 = \frac{949999}{949999+50001} + \frac{3}{3+2} - 1 = \Delta_{UAPV4}$. Das eine Objekt mit negativem Kriteriumswert mehr, das durch den dritten Algorithmus korrekt klassifiziert wird, wird durch 200000 Objekte mit positivem Kriteriumswert aufgewogen, die falsch klassifiziert werden. Die Genauigkeitswerte liegen hingegen bei $A_3 = \frac{750004}{1000000} = 0,750004$ und $A_4 = \frac{950002}{1000000} = 0,950002$. Es bleibt hier festzuhalten, dass eine Präferenz für eines der beiden Kriterien nicht formal begründet werden kann. Statt dessen bedarf es der Abwägung der Konsequenzen der Klassifikationsfehler im Einzelfall, was den Vorschlag für eine Erweiterung der Aufgabenstellung motivieren soll.

Büttner (1997) bezeichnet den Kennwert Δ_{UAPV} als *Youden Index J* im Verweis auf Youden (1950) und setzt ihn zur sogenannten *Diagnostic effectiveness* ins Verhältnis, die sich als einfache Summe von Sensitivität und Spezifität berechnet. Neben diesen Kriterien sind weitere Kriterien in Gebrauch, an dieser Stelle sollte noch der *ϕ-Koeffizient* genannt werden, der sich als

$$\phi = \frac{n_{11}n_{00} - n_{10}n_{01}}{\sqrt{(n_{11} + n_{01})(n_{10} + n_{00})(n_{11} + n_{10})(n_{01} + n_{00})}} \tag{2.15}$$

berechnet und in direkter Beziehung zur Korrelation zwischen Kriteriumswerten und vorhergesagten Werten steht. Hollmann (1991) kommt allerdings bezüglich dieses Kriteriums zu dem Schluss:

> Zusammenfassend lässt sich festhalten, dass der üblicherweise berechnete Phi-Koeffizient zwar bei einem dichotomen Prädiktor und einem dichotomen Kriterium durchaus ein Zusammenhangsmaß im Sinne der Kriteriumsvalidität darstellt, jedoch in der Anwendung nur von begrenzter praktischer Relevanz ist. (Hollmann, 1991, S. 37)

Amelang und Zielinski (2002) verweisen zudem auf das *Minmax-Kriterium*, das auf eine Maximierung des Minimums von Sensitivität und Spezifität hinausläuft, also

$$max!(min(S^+, S^-)). \tag{2.16}$$

Das *Minimum-Loss*-Kriterium minimiert Zuordnungsfehler über alle Klassen hinweg und fällt damit einer Maximierung von A zusammen, da

$$L = n(1 - A) \tag{2.17}$$

.

Ein weiteres Kriterium, das gerade für die Klassifikation durch Regelmengen relevant ist, ist die Anzahl der Regeln in der Regelmenge (Gür Ali & Wallace, 1997):

Tabelle 2.6: Nutzentabelle zur Bewertung von Klassifikationsergebnissen

		P_j	
		1	0
Y_j	1	o_{11}	o_{10}
	0	o_{01}	o_{00}

> More rules for the same dataset mean smaller sample size per rule. Therefore, the variability in the correct classification rates of individual rules will be larger for the same overall sample size if we have more rules. The variance of the correct classification proportion changes inversely with the rule sample size. (Gür Ali & Wallace, 1997, S. 5)

Bei gleicher Klassifikationsgüte sind im Allgemeinen also Lösungen mit weniger Regeln vorzuziehen. Zudem wird für Regelmengen auch die Anzahl unklassifizierter Fälle von Bedeutung sein. Neben Kriterien, die sich auf die Klassifikationsgüte im engen Sinne beziehen, werden weitere Gütekriterien für Regeln diskutiert, die einen subjektiveren Charakter aufweisen: Dazu gehört beispielsweise die Einschätzung des Interesses, das eine Regel verdient: Eine Regel, die den Anwender überrascht, aber trotzdem genau ist, wird nach dem Kriterium der *Rule Interestingness* höher bewertet als eine eher erwartete Regel. Auch eher objektive Varianten dieses Kriteriums werden eingesetzt, indem die Überraschung definiert wird als die Heranziehung von Attributen mit lediglich geringem Informationsgewinn (vgl. Freitas, 2002). Da in der vorliegenden Arbeit der Fokus auf der Erstellung von Klassifikationsregeln zur Vorhersage und nicht vorerst auf der Erklärung von Zusammenhängen liegen soll, werden Maße der *Rule Interestingness* hier nicht weiter verfolgt.

2.3.4 Vorschlag für eine zweite Modifikation der Zielaufgabe

2.3.4.1 Definition des U_c-Kriteriums und des KNMK

Ein Ausweg aus den im vorangegangenen Abschnitt geschilderten Problemen besteht in der Explizierung des quantitativen Nutzens der vier möglichen Klassifikationsfälle (der jeweils auch negativ beurteilt werden kann). Dabei soll erneut betont werden, dass die Entscheidung für die Gewichtung im Allgemeinen nicht formal getroffen werden kann. Eine verantwortbare Festlegung sollte sich auf ethische und professionelle Kriterien stützen, die in der Klassifikationssituation relevant werden (vgl. Amelang & Zielinski, 2002). In der psychologischen Diagnostik sollten etwa die Konsequenzen der Klassifikation beurteilt werden, wobei neben den expliziten Interessen des Auftraggebers eine professionelle Haltung des Diagnostikers auch ethische Erwägungen hinsichtlich der Konsequenzen für die Betroffenen von Entscheidungen verlangt.

Bei einer Personalauswahlentscheidung sollte etwa berücksichtigt werden, als wie groß der Nutzen einer Auswahl von für eine Position geeigneten Bewerbern beurteilt werden sollte im Vergleich zum Nutzen eines ungeeigneten Bewerbers und dem Nutzen der Abweisung von geeigneten und ungeeigneten Bewerbern. Bei der Abweisung von geeigneten Bewerbern könnten etwa Auswirkungen auf das Image des Auftraggebers, aber auch Auswirkungen auf das Selbstkonzept und den Lebensweg der Beurteilten berücksichtigt werden; Aspekte, die natürlich auch bei der Aufnahme ungeeigneter Bewerber bedeutsam sind. Bei Entscheidungen über die Zuteilung zu therapeutischen Interventionsprogrammen sollten analog die Konsequenzen falscher Zuweisungen sowohl aus der Perspektive von Auftraggebern als auch aus der Sicht der Betroffenen überdacht werden.

Ob ein Klassifikationsverfahren als konservativ gewünscht wird, d.h. als mit einer hohen Spezifität ausgezeichnet auch auf Kosten einer niedrigen Sensitivität, oder als permissiv, d.h. als mit einer hohen Sensitivität ausgezeichnet auch auf Kosten einer niedrigen Spezifität, hängt von der Gewichtung der Alternativen ab. Von daher wird folgendes alternatives Zielkriterium für die Klassifikation durch GIER$_{[1.0]}$vorgeschlagen:

Klassifikationsnutzenmaximierung: Sind die Nutzenwerte für die vier Klassifikationsmöglichkeiten quantifiziert und wie in Tabelle 2.6 benannt, dann ist es die Aufgabe der Klassifikationsverfahren, den

Wert N_c zu maximieren, also:

$$\max U_c \text{ , mit } U_c = \sum_{j=1}^{n} u_j \text{ , wobei } u_j = \begin{cases} o_{11} & P_i = 1 \wedge Y_i = 1 \\ o_{10} & P_i = 0 \wedge Y_i = 1 \\ o_{01} & P_i = 1 \wedge Y_i = 0 \\ o_{00} & P_i = 0 \wedge Y_i = 0 \end{cases} \tag{2.18}$$

Das Kriterium wird im Folgenden als *Klassifikationsnutzenmaximierungskriterium* (KNMK) bezeichnet.

Aus der Abhängigkeit, dass $n_1 = n_{10} + n_{11}$ und $n_0 = n_{00} + n_{01}$ unabhängig vom Klassifikatonsverfahren vorgegeben sind, ergibt sich eine Äquivalenz der Vergleichsergebnisse (nicht der U_c-Werte selbst) zwischen Algorithmen, wenn der Quotient aus $\delta_1 = o_{11} - o_{10}$ und $\delta_0 = o_{00} - o_{01}$, also aus den Differenzen der Nutzenwerte bei korrekter und fehlerhafter Entscheidung für Objekte mit gleichem Kriteriumswert bei zwei Zuteilungen identisch ist. Die δ-Werte entsprechen der Gewichtung der beiden möglichen Fehler. Die Addition eines Wertes ω zu beiden Nutzenwerten für eine Kriteriumsklasse vergrößert U_c um einen konstanten Wert $\omega \cdot n_1$ bzw. $\omega \cdot n_0$ ohne Konsequenz für den Vergleich zwischen Klassifikationsergebnissen. Die Multiplikation aller Nutzenwerte mit einem Faktor κ führt zu einer Multiplikation aller U_c-Werte um den gleichen Faktor, dementsprechend kann sich die Differenz der δ-Werte ebenfalls ohne Konsequenz für den Vergleich zwischen Klassifikationsergebnissen verändern. Die Veränderung eines δ-Wertes ohne proportionale Veränderung des anderen Wertes hingegen verändert das Verhältnis der Konsequenzen von Fehlern.

2.3.4.2 Vergleichbarkeit zu anderen Kriterien

Zum Vergleich zwischen Kriterien reicht es aus, den Quotienten der δ-Werte zu betrachten. Folgendes Theorem soll das Kriterium in seinem Verhältnis zu alternativen Kriterien kennzeichnen:

Satz 2.2 (Kriterienvergleichbarkeit) ***i)*** *Gegeben sei eine Klassifikationsaufgabe und eine Nutzenwertverteilung mit $\delta_1 > 0$ und $\delta_0 > 0$. Ist $\frac{\delta_0}{\delta_1} = 1$, dann ist das KNMK bezüglich der Entscheidung zwischen Klassifikationsverfahren für diese Aufgabe identisch mit dem Kriterium, die Klassifikationsgenauigkeit zu maximieren.* ***ii)*** *Ist $\frac{\delta_0}{\delta_1} = \frac{n_1}{n_0}$, dann ist das Kriterium identisch mit dem Kriterium, den Δ_{UAPV}-Score zu maximieren.* ***iii)*** *Ist $\frac{\delta_0}{\delta_1} > \frac{n_1}{n_0}$, wird die Spezifität höher gewichtet als die Sensitivität eines Verfahrens und ist $\frac{\delta_0}{\delta_1} < \frac{n_1}{n_0}$, wird die Sensitivität höher gewichtet als die Spezifität.*

Beweis: i) Ist $\frac{\delta_0}{\delta_1} = 1$, werden zwei Aufteilungen als identisch betrachtet, wenn die Summe der korrekten Entscheidungen identisch ist, da durch Fehler beider Art der gleiche Betrag von U_c subtrahiert wird ($\delta_0 = \delta_1$). Da $n = n_{00} + n_{01} + n_{10} + n_{11}$ unabhängig vom Klassifikationsalgorithmus durch die Klassifikationsaufgabe vorgegeben wird, ist n für alle Aufteilungen konstant, dementsprechend ist $U_c = A\delta_0 + n_0 o_{01} + n_1 o_{10}$ und da $n_0 o_{01} + n_1 o_{10}$ für alle Algorithmen konstant ist, ist jede Differenz von U_c-Werten allein abhängig von der Differenz der A-Werte. Solange der Nutzen von Fehlern geringer bewertet wird als der Nutzen von korrekten Entscheidungen ($\delta_0 > 0$) sind damit Vergleichsentscheidungen auf der Basis des Klassifikationsgenauigkeitskriteriums identisch mit denen des KNMK.

ii) Angenommen $\frac{\delta_0}{\delta_1} = \frac{n_1}{n_0}$: Zerlegt man U_c in die Summanden U_{c0} und U_{c1}, wobei U_{c0} die Nutzensumme für Objekte mit negativem Kriterium, U_{c1} die Nutzensumme für Objekte mit positivem Kriterium bezeichnen soll, dann ist $U_{c0} = S^- n_0 \delta_0 + n_0 o_{01}$ und $U_{c1} = S^+ n_1 \delta_1 + n_1 o_{10}$. Dementsprechend ist $U_c = U_{c0} + U_{c1} = S^+ n_0 \delta_0 + S^- n_1 \delta_1 + n_0 o_{01} + n_1 o_{10}$. Der Term $n_0 o_{01} + n_1 o_{10}$ ist für jede Klassifikationslösung konstant, da n_0 und n_1 vorgegeben sind und o_{01} und o_{10} zwischen Bewertungen nicht variieren. Dementsprechend ist die Nutzenwertdifferenz zwischen zwei Lösungen darstellbar als: $U_{ca} - U_c b = S_a^+ n_1 \delta_1 + S_a^- n_0 \delta_0 - S_b^+ n_1 \delta_1 - S_b^- n_0 \delta_0 = n_1 \delta_1 (S_a^+ - S_b^+) + n_0 \delta_0 (S_a^- - S_b^-)$. Aus der Annahme ergibt sich $\delta_0 n_0 = \delta_1 n_1$, also ist $U_{ca} - U_{cb} = n_1 \delta_1 (S_a^+ - S_a^+ - S_b^+ - S_b^-)$. Die Differenz zweier Δ_{UAPV}-Scores ist wiederum darstellbar als $\Delta_{UAPVa} - \Delta_{UAPVb} = S_a^+ + S_a^- - 1 - (S_b^+ + S_b^- - 1) = S_a^+ + S_b^- - S_b^+ - S_b^-$. Dementsprechend ist also $n_1 \delta_1 (\Delta_{UAPVa} - \Delta_{UAPVb}) = U_{ca} - U_{cb}$. Wenn $n_1 \delta_1 > 0$, also eine Fehlentscheidung nicht als nützlicher bewertet wird als eine korrekte Entscheidung, sind damit die Algorithmenvergleichsergebnisse identisch.

iii) Oben wurde demonstriert, dass $U_{ca} - U_{cb} = n_1 \delta_1 (S_a^+ - S_b^+) + n_0 \delta_0 (S_a^- - S_b^-)$. Ist $\frac{\delta_0}{\delta_1} > \frac{n_1}{n_0}$, dann ist, gegeben $\delta_1 > 0$, $n_0 \delta_0 > n_1 \delta_1$. Deshalb wird bei der Bestimmung der Nutzenwertdifferenz die Spezifi-

tätsdifferenz höher gewichtet als die Sensitivitätsdifferenz. Aus $\frac{\delta_0}{\delta_1} < \frac{n_1}{n_0}$ hingegen folgt bei positivem δ_1 der umgekehrte Zusammenhang, da $n_0\delta_0 < n_1\delta_1$ ■.

Eine weitere Vergleichsmöglichkeit ergibt sich aus der Überlegung, dass eine Verdoppelung der Differenz des Nutzens zwischen falscher und richtiger Zuordnung im Klassifikationsnutzen ununterscheidbar ist von einem Verfahren mit verdoppelter Stichprobe für diese Klasse und unveränderter Sensitivität und Spezifität.

Adams und Hand (2000) problematisieren die Annahme gleicher Kosten für die beiden möglichen Fehlentscheidungen. Ihrer Meinung nach ist es gerade in medizinischen Fragen sehr selten, dass die beiden Kosten einander entsprechen. Die Fehldiagnose einer Krankheit wird als weniger gravierend betrachtet als die Nichtdiagnose einer tatsächlich vorhandenen Krankheit. Die Autoren sind ähnlich kritisch gegenüber der Festsetzung expliziter Kosten:

> The costs depend on profitability, which includes issues such as marketing and advertising costs. Moreover, the relative misclassification costs are likely to change over time as the economy evolves. Although it is probable that a range of likely costs can be given, it is improbable that exact ones can be given. (S. 307)

Die von ihnen vorgeschlagene Variante der Angabe von oberen und unteren Grenzen für Kostenverhältnisse und Angabe von Erwartungswerten für das Eintreten dieser Werte würde allerdings nicht weniger, sondern mehr Präzision und Wissen erfordern.

Da in der vorliegenden Arbeit virtuelle Entscheidungsmodelle diskutiert werden, die nicht zu zukünftigen Konsequenzen für die Betroffenen führen, wird auf eine Anpassung des Kriteriums an den Anwendungsfall verzichtet und statt dessen soll die einfache Klassifikationsgenauigkeit zum Einsatz kommen, also zum Beispiel durch $o_{00} = 1, o_{11} = 1, o_{01} = 0$, und $o_{10} = 0$. Damit werden alle Algorithmen auf die gleiche Stufe gestellt und können hinsichtlich dieses Kriteriums verglichen werden. Gleichwohl sollen Spezifität und Sensitivität dabei nicht aus den Augen verloren werden. Trotzdem ist die Möglichkeit für abweichende Nutzenbeurteilungen aus den diskutierten Gründen in GIER$_{[1.0]}$implementiert worden.

2.3.5 Fitting, Vorhersage und Kreuzvalidierung

2.3.5.1 Unterscheidung zwischen Aufgabentypen

Selbst bei identischen Klassifikationsaufgaben und einer identischen Menge von Algorithmen kann ein Algorithmenvergleich verschiedenen Formen annehmen, denn Datensätze an sich geben die Aufgabe nicht vor. Die nahe liegende Variante, die Modelle so gut an den Datensatz anzupassen, dass die Kriteriumswerte möglichst gut vorhergesagt werden können, legt den Schwerpunkt auf eine statistische Vorhersage im Sinne einer Anpassung. Eine solche Aufgabe wird deshalb auch als *Data Fitting* bezeichnet. Der Datensatz wird vollständig herangezogen zur Schätzung von Parametern des Modells und zur Bestimmung der Güte des Modells. Als *Fitting*-Klassifikationsgenauigkeit wird die Klassifikationsgenauigkeit für die Fälle der Trainingsstichprobe bezeichnet. Dieser Wert ist vor allem dann relevant, wenn ein Klassifikationsalgorithmus eine ökonomischere Beschreibung eines Datensatzes liefern soll, wenn die Parameter für eine bekannte Struktur quantifiziert werden sollen, oder wenn das Ziel explorative Hypothesen über mögliche Zusammenhänge zwischen Prädiktoren und Kriterium sind. Ihn als Maß für die Leistungsfähigkeit des Algorithmus zu bezeichnen, ist in einem bestimmten Sinne irreführend, denn es kommt hier schließlich zu einer Resubstitution der Trainingsstichprobe als Teststichprobe. Bereits bei der Diskussion der „optimalen“ Klassifikationsmethode oben wurde auf die Beschränkungen eines Ansatzes hingewiesen, der sich optimal an eine Stichprobe anpasst. Letztendlich sind die Kriteriumswerte für die Stichprobe bekannt, deshalb ist es von beschränktem Nutzen, eben diese akkurat vorhersagen zu können. Nichtsdestotrotz wird von einem guten Klassifikationsverfahren natürlich erwartet, nicht bereits bei der Vorhersage bekannter Klassenzugehörigkeiten zu scheitern.

Bei der Bestimmung der Vorhersagefähigkeit eines Klassifikationsmodells für Fälle, die nicht zur Schätzung des Modells herangezogen wurden, entsteht ein Dilemma: Sowohl zur Bestimmung der Parameter als auch zur Prüfung der Modellgüte ist es notwendig, die Kriteriumswerte der Objekte zu kennen. Alle Objekte mit bekannten Kriteriumswerten befinden sich aber üblicherweise bereits in der Stichprobe. Für die bestmögliche Bestimmung der Parameter ist es deshalb im Allgemeinen vorzuziehen, die gesamte Stichprobe heranzuziehen. Das verbietet aber streng genommen die Verwendung der *Fitting*-Klassifikationsgenauigkeit (A_F) als Maß für die Modellgüte bei unvertrauten Fällen. Es gibt verschiedene

Ansätze, diesen Konflikt zu lösen. Es ist zunächst nahe liegend, einen Teil der Fälle aus der Stichprobe zu nehmen und diesen exklusiv für die Modellüberprüfung zu verwenden. Der Stichprobenteil zur Schätzung der Parameter wird als *Trainingsstichprobe*, der verbleibende Teil als *Teststichprobe* bezeichnet. Wird ein mit Hilfe der Trainingsstichprobe entwickeltes Modell an der Teststichprobe getestet, so ist dies ein Beispiel für die Überprüfung der Generalisierbarkeit eines Modells auf Fälle außerhalb der bekannten Menge, was auch als *out-of-sample prediction* bezeichnet wird. Gegen diese Methode gibt es allerdings Einwände: Ein Modell, das nur an einem Teil der Stichprobe entwickelt wird, ist unter Umständen schlechter als das Modell, dem die gesamte Stichprobe zur Verfügung steht; damit würde durch die Verwendung der out-of-sample Klassifikationsgenauigkeit (A_O) die Genauigkeit des an der vollständigen Stichprobe geschätzten Modells unterschätzt. Je mehr Objekte in die Teststichprobe sortiert werden, desto gravierender wird das Defizit. Je weniger Objekte sich aber in der Teststichprobe befinden, desto schlechter wird die Zuverlässigkeit der Schätzung von (A_O). Die Zuteilung in Test- und Trainingsstichprobe erfolgt zudem zufällig, so dass bei einer Wiederholung der Prozedur ein unterschiedliches Ergebnis zu erwarten ist. Eine einmalige Durchführung der Prozedur scheint nur bei ausreichend großen Fallzahlen in Trainings- und Teststichprobe gerechtfertigt. Dementsprechend muss die Aufteilung wiederholt durchgeführt werden, was je nach Rechenzeit der Algorithmen einen hohen Zeitaufwand bedeuten kann. Enthalten die Algorithmen nicht-deterministische Elemente, so würde selbst bei gleicher Aufteilung nicht ein identisches Ergebnis erwartet werden können, und streng genommen existiert das Modell, dessen Generalisierbarkeit überprüft werden kann, in diesen Fällen lediglich als Wahrscheinlichkeitsverteilung seiner Parameterwerte gegeben die Stichprobe. Unterschiedliche Trainingsstichprobengrößen können sich auf die Leistungsfähigkeit verschiedener Algorithmen unterschiedlich auswirken, so dass nicht nur die Gütewerte selbst, sondern auch ihre Reihenfolge von diesen Entscheidungen beeinflusst werden.

Eine Alternative zu dieser Methode stellt die *k-fache Kreuzvalidierung* dar. Dieses Verfahren versucht das Dilemma insofern zu umgehen, als dass die gesamte Stichprobe als Teststichprobe verwendet wird, aber gleichzeitig auch zur Bildung von Trainingsstichproben herangezogen wird. Das gelingt durch ein sequenzielles Vorgehen: Die Stichprobe wird zufällig in k Partitionen geteilt. Der Reihe nach wird jede dieser Partitionen als Teststichprobe aussortiert, während das Modell mit den verbleibenden $k-1$ Partitionen geschätzt wird. Als kreuzvalidierte Klassifikationsgenauigkeit (A_K) wird nun die Genauigkeit bezeichnet, mit der Fälle klassifiziert werden konnten, wenn sie in der Trainingsstichprobe waren. Dabei sind so viele Modellschätzungen wie Partitionen notwendig. Maximal sind bei n Objekten $k = n$ Partitionen möglich. In diesem Fall würde die Teststichprobe aus nur einem Fall bestehen und nahezu die gesamte Stichprobe könnte für die Modellschätzung verwendet werden. Dieses Vorgehen wird auch als leave-one-out-Kreuzvalidierung bezeichnet. Für alle anderen Fälle sind unterschiedliche Ergebnisse je nach Aufteilung der Objekte zu erwarten. Auch bei nicht-deterministischen Algorithmen sollte das Verfahren mehr als einmal zum Einsatz kommen, um einen Schätzwert zu erhalten. Kleinere Werte von k bieten sich an, wenn der Rechenaufwand der Algorithmen hoch ist.

Witten und Frank (2001) weisen darauf hin, dass keine der Methoden verzerrungsfrei arbeitet. Entgegen mancher überzogener Erwartung trifft das auch auf die Methode der leave-one-out-Kreuzvalidierung zu (vgl. Zhu & Rohwer, 1996): Als einfaches Gegenbeispiel kann ein hypothetischer Datensatz dienen, der sich in exakt gleichen Anteilen auf zwei Klassen aufteilt. Die Vorhersagegenauigkeit einer Methode, die für jeden Fall die Majoritätsklasse vorhersagt und bei gleichen Anteilen eine der beiden Klassen auswählt, liegt bei $\frac{1}{2}$, da unabhängig von der einmaligen Entscheidung jeweils die Hälfte der Fälle der bestimmten Klasse angehören. Nutzt man aber die Methode der leave-one-out-Kreuzvalidierung, so wird die Genauigkeit der Prozedur auf 0 geschätzt, da in der reduzierten Stichprobe genau jene Klasse die Mehrheit der Fälle stellt, der der ausgelassene Fall eben nicht angehört. Somit wird jeder Fall falsch klassifiziert.

Der Fall, in dem die *Fitting*-Genauigkeit eines Modells zu einer Überschätzung seiner Leistungsfähigkeit führt und sich beispielsweise die Leistungsreihenfolge zwischen Modellen zwischen Trainingsset und Testset umkehrt, wird auch als *Overfitting* oder als Überanpassung bezeichnet:

> A model is said to overfit the data if there is another model which performs less well on the training set but better on the test set. (Martignon & Hoffrage, 2002, S. 41)

Im Folgenden sollen einige Beispiele für *Overfitting* gegeben werden.

2.3.5.2 Beispiele für Overfitting und Vergleich mit eigenen Simulationsergebnissen

Als ein Beispiel für *Overfitting* lässt sich die Situation der Rekonstruktion einer zweidimensionalen Funktion heranziehen aufgrund einer Anzahl von Punkten in der Stichprobe. Unter der Annahme, dass die Punkte einer polynomialen Funktion entstammen, allerdings mit einem Messfehler behaftet sind, ergibt sich in der Konsequenz, dass bei der Anpassung von Polynomen aufsteigender Ordnung trotzdem der Punkt erreicht werden wird, an dem die Punkte der Stichprobe perfekt „vorhergesagt“ werden können, bei n Punkten mit unterschiedlichen x-Werten ist dies spätestens durch ein Polynom der Ordnung $n-1$ möglich. In der Tat würde die Summe der Abweichungen für Polynome höherer Ordnung bis zu diesem Punkt monoton fallen.

Wird anschließend eine zweite Stichprobe von Punkten aus der ursprünglichen Funktion gezogen, werden diese sehr wahrscheinlich nicht auf der ermittelten Kurve liegen. Quantifiziert man die Abweichungen der unterschiedlichen Modelle, so ist zu erwarten, dass bis zu einem Grad in der Rekonstruktion Polynome höherer Ordnung erfolgreicher sein werden als Polynome niedrigerer Ordnung, sich der Fittingvorteil ab diesem Punkt aber in einen Nachteil in der Generalisierungsleistung umschlägt. Ab diesem Punkt kann dementsprechend davon gesprochen werden, dass bei den Vorhersagemodellen ein *Overfitting* vorliegt.

Als Beispiel für die Relevanz dieser Überlegungen liefert Einhorn (1972) zum einen das Beispiel einer Studie, bei der 84 Prädiktoren bei 136 Versuchspersonen zum Einsatz kamen. Es mag kaum überraschen, dass der ermittelte multiple Korrelationskoeffizient bei einer linearen Modellierung von $R = 0,73$ bei der Kreuzvalidierung auf $R = 0,04$ zurückging. Auch Korrekturformeln für die erwartete Schrumpfung können diese Schwäche nicht immer vorhersagen. Einhorn (1972) nennt ein weiteres Beispiel, bei dem ein ermitteltes und korrigiertes R von $R = 0,92$ sich in der Kreuzvalidierung in $R = -0,21$ verwandelte. Wäre der erste Wert als Indikator für eine hervorragende Erfassung einer Dimension gewertet worden mit einer Nutzung in praktischen Anwendungen, so wäre beträchtlicher Schaden angerichtet worden, da unter Umständen auf dieser Grundlage Entscheidungen schlechter als Zufallsentscheidungen gewesen wären. Ein ähnliches Problem affligiert auch schrittweise Regressionen: Bei 193 Versuchspersonen wurde bei zehn Prädiktoren $R = 0,430$ erreicht, bei einer schrittweisen Regression und einer Auswahl von zehn aus vierzig Prädiktoren $R = 0,549$. In der Kreuzvalidierung kehrte sich das Verhältnis um, der erste Koeffizient sank auf $R = 0,380$, der zweite auf $R = 0,363$. Thompson (1995) spricht von einem „statistical welfare system“, wenn vier bis fünf Prädiktoren ausgewählt werden, indem bei fünfzig Variablen bei jedem Schritt nachgesehen wird, welche Hinzufügung das beste Ergebnis bringt:

> [...] the computer packages are erroneously not charging us any degrees of freedom for consulting our data in this manner. This statistical welfare system may cause us to radically overestimate the atypicality of our results [...]. (Thompson, 1995, S. 527)

Freedman (1999) betont, dass der Ursprung der Regressionsmethode am Anfang des neunzehnten Jahrhunderts durch Legendre und Gauss mit der Zielsetzung verbunden war, die Umlaufbahnen astronomischer Objekte vorherzusagen. Messungen waren hier mit hoher Präzision und in großer Anzahl möglich, die Art der Fehler in Messung und Gleichungen war bekannt, und es bestanden reichhaltige Möglichkeiten zum Abgleich zwischen Modell und Beobachtung. Bei modernen Anwendungen hingegen sind die Annahmen über Fehlerterme selten durch wissenschaftliche Vorarbeiten diktiert, so dass ein großes Potenzial für Fehler besteht, erst recht, wenn nicht nur Anpassungen berechnet, sondern Interventionsergebnisse vorhergesagt werden sollen:

> Fitting an equation to an existing data set is one activity, predicting the results of an intervention is quite another, and the crucial issue is getting from here to there. (Freedman, 1999, S. 255)

Es bietet sich hier der Vergleich mit Ergebnissen einer Simulation des Verfassers zum Nutzen konnektionistischer Modelle für die Diagnostik an (Woike, 2000). In einem Vergleich der Klassifikationsleistung von *Multilayer Perceptrons*, *Radial Basis Function Networks* mit Entscheidungsbaumverfahren (CART und CHAID) und der Diskriminanzanalyse in einfacher und schrittweiser Variante wurden Variationen der Verfahren auf das - künstliche - Problem angesetzt, das Geschlecht eines - realen - Teilnehmers an einer Fragebogenstudie zum Bochumer Inventar zur berufsbezogenenen Persönlichkeitsbeschreibung (Hossiep & Paschen, 1998) aufgrund seiner Skalen- bzw. Itemwerte in einem Persönlichkeitstest vorherzusagen. Dabei kamen einmal drei Skalenwerte als Prädiktoren zum Einsatz, einmal alle 14 Skalenwerte und zuletzt

Tabelle 2.7: Generalisierungsleistung von Diskriminanzanalysen, Entscheidungsbäumen und konnektionistischen Klassifikationsverfahren nach Woike (2000)

		Klassifikationsfehler 196 Items		Klassifikationsfehler 14 Skalen		Klassifikationsfehler 3 Skalen	
Algorithmus	Anzahl Varianten	beste Variante	schlechteste Variante	beste Variante	schlechteste Variante	beste Variante	schlechteste Variante
DA	4	0,1728	0,1995	0,2255	0,2838	0,2574	0,3949
DA (stepw.)	8	0,1833	0,2468	0,2270	0,2873	0,2583	0,3983
CART	24	0,2510	0,3544	0,2600	0,3701	0,2605	0,3826
CHAID	8	0,2578	0,3010	0,2574	0,2721	0,2625	0,2672
Exh. CHAID	8	0,2544	0,2875	0,2613	0,2694	0,2625	0,2625
MLP	12	-	-	0,2243	0,2368	0,2566	0,2637
RBF	24	-	-	0,2301	0,2625	02591	0,2701

alle 196 Itemwerte. Es kam die Methode der dreifachen Kreuzvalidierung zum Einsatz, die Stichprobe ($n = 3965$) wurde in drei Teile gespalten, wobei an jeweils zwei Teilen das Modell geschätzt wurde, das an der dritten überprüft wurde. Dementsprechend sind die angegebenen Ergebnisse Generalisierungsergebnisse, die aufgrund der (damals rechenzeitbedingten) geringen Trialzahl nur unter großem Vorbehalt interpretiert werden können.

Es fällt bei einem Vergleich der Ergebnisse auf, dass das beste Ergebnis bei jeder der drei Aufgaben von einer einfachen linearen Diskriminanzanalyse erzielt wurde, während das schrittweise Verfahren - das oft gerade zur Verhinderung von *Overfitting* eingesetzt wird - jeweils schlechtere Ergebnisse erbringt. Auch die deutlich komplexeren konnektionistischen Modelle können trotz einer - hier nicht dargestellten - besseren *Fitting*-Leistung die Ergebnisse des schlichtesten Modells nicht übertreffen, was genauso für die Entscheidungsbäume zutrifft. Eine Detailanalyse der Ergebnisse zeigt zudem, dass die Berücksichtigung von Interaktionstermen zu einer Verschlechterung der Ergebnisse des linearen Modells führte. Dies deutet darauf hin, dass im Datensatz keine einfach auffindbaren Interaktionen zu finden waren, weil entweder von Linearität ausgegangen werden kann oder das Interaktionsmuster komplexer ist, als dass es durch die Verfahren gefunden werden konnte. Diese Ergebnisse mögen auf jeden Fall als Mahnung dienen, auch die einfachen Möglichkeiten in einem Methodenvergleich zu berücksichtigen und *alle* Modelle in einem Methodenvergleich auf ihre Generalisierungsleistung zu prüfen.

Eine Begutachtung der gewonnenen Modelle zur Feststellung von *Overfitting* scheint also angeraten. In dieser Arbeit soll eine Variante der beschriebenen Vorgehensweisen zur Kontrolle von *Overfitting* zum Einsatz kommen, da nicht die Erstellung eines spezifischen Vorhersagemodells im Vordergrund steht, sondern die Leistungsfähigkeit der Algorithmen miteinander verglichen werden soll. Aus diesem Grund wird die Stichprobe als Population betrachtet, aus der wiederholt Zufallsstichproben festgelegter Größe gezogen werden, um als Trainingsstichproben für die Schätzung eines Modells zu dienen, das an den jeweils verbleibenden Fällen geprüft wird. Dieses Vorgehen wird nun systematisch für verschiedene Stichprobengrößen in einer für eine Bestimmung der durchschnittlichen Leistungsfähigkeit ausreichenden Anzahl wiederholt[12]. Ziel bei diesem Vorgehen ist die Gewinnung eines Überblick über die Leistungsfähigkeit von Algorithmen bei unterschiedlich großen Stichproben. Damit können Fragen untersucht werden, die minimale Stichprobengrößen für verlässliche Generalisierbarkeit von Ergebnissen bei verschiedenen Algorithmen betreffen. Die Abhängigkeit von der Größe der Trainingsstichprobe ist damit auch außerhalb des Algorithmenvergleichs interessant. Die Stichprobenziehung soll dabei so oft wiederholt werden, dass auf eine Anwendung von Signifikanztests zur Prüfung von Unterschieden in der gewählten Auflösung verzichtet werden kann.

2.4 Visualisierung von Klassifikationsergebnissen

2.4.1 Das ROCS-Diagramm

Zur Veranschaulichung der Klassifikationsleistung von Algorithmen ist die Darstellung als ROC-Kurve (Receiver Operation Characteristic) üblich. Die Namensgebung hat ihre Wurzeln in der Signalentdeckungstheorie (vgl. Amelang & Zielinski, 2002). Dabei wird in einem zweidimensionalen Diagramm die Sensitivität eines Klassifikationsalgorithmus gegen die von 1 subtrahierte Spezifität aufgetragen. Der Begriff der „Kurve“ ist streng genommen nur in den Fällen anwendbar, wenn bspw. aufgrund eines wählbaren Cut-Off-Wertes ein Verfahren in seiner Strenge bzw. Permissivität angepasst werden kann. In diesem Fall kann das Diagramm veranschaulichen, wie sich eine Verschärfung des Auswahlkriteriums auf Sensitivität (im Allgemeinen negativ) und Spezifität (im Allgemeinen positiv) auswirkt, um so eine Strategie wählen zu können, die den Bedürfnissen der Entscheidungssituation am besten entspricht. Wird der Cut-Off-Wert hingegen vom Verfahren implizit gesetzt, d.h. liegt lediglich das Klassifikationsergebnis für Fälle vor und kein Wahrscheinlichkeits- oder Distanzwert für Klassenzugehörigkeiten, reduziert sich die Darstellung eines Algorithmus im ROC-Diagramm auf einen Punkt. Nichtsdestotrotz können auf diese Weise aber verschiedene Algorithmen übersichtlich miteinander verglichen werden. In diesem Abschnitt soll eine Modifikation der ROC-Darstellungsform eingeführt werden, das ROC-*Shading*-Diagramm, oder kurz

[12] Witten und Frank (2001) nennen diese Vorgehensweise eine „wiederholte Holdout-Methode“.

ROCS-Diagramm. Abbildung 2.5 zeigt ein vereinfachtes ROCS-Diagramm für den Fall, dass $n_0 = n_1$, also für den Fall, dass in der Stichprobe die beiden Kriteriumsklassen gleich verteilt sind.

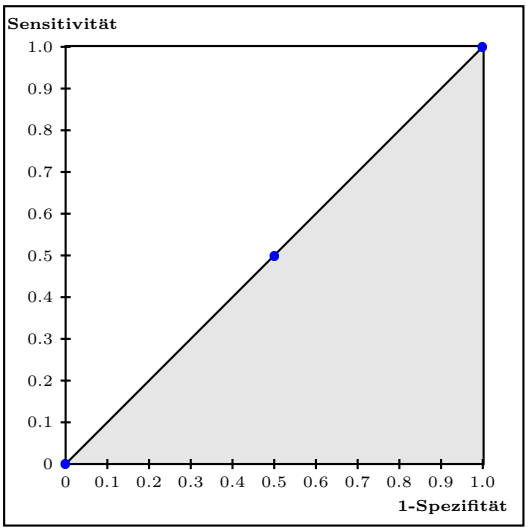


Abbildung 2.5: Vereinfachtes ROC-Shading-Diagramm für $n_1 = n_0$

2.4.2 Iso-Klassifikationsgenauigkeitslinien für $n_1 = n_0$

In das Diagramm sind drei Punkte eingetragen, die das Ergebnis dreier unterschiedlicher Auswahlstrategien repräsentieren. Der Punkt $(0,5; 0,5)$ entspricht dem Erwartungswert einer zufälligen Zuteilung der Objekte zu den beiden Kategorien. In diesem Fall würde sowohl für die Sensitivität als auch für die Spezifität ein Erwartungswert von $0,5$ angesetzt werden können. Der Punkt $(1,1)$ entspricht dem Ergebnis der Strategie, für alle Objekte eine positive Vorhersage zu treffen. In diesem Fall würde jedes Objekt mit positivem Kriteriumswert als solches erkannt, allerdings würde kein Objekt mit negativem Kriteriumswert als solches identifiziert werden können. Die gegenteilige Strategie, nämlich alle Objekte in die negative Kategorie einzuordnen, entspricht dem Punkt $(0,0)$. Hier ist die Sensitivität bei Null, dafür die Spezifität auf dem Maximalwert von 1. Für alle drei Strategien gilt, dass die Genauigkeit des Algorithmus bzw. im Fall der Zufallsauswahl die erwartete Genauigkeit der Strategie bei 0,5 liegt. Analog wird für den Δ_{UAPV}-Score bei jeder Strategie ein Wert von 0 erwartet: $E(\Delta_{UAPVZ}) = 0,5 + 0,5 - 1 = 0, \Delta_{UAPVN} = 1 + 0 - 1 = 0, \Delta_{UAPVN} = 0 + 1 - 1 = 0$. Als Antwort auf die Frage, für welche anderen Punkte im Diagramm die gleichen Ergebnisse erhalten werden, ergibt sich die im ROCS-Diagramm eingetragene Gerade. Dies kann in folgender Weise demonstriert werden:

Angenommen $A = v$, dann ergibt sich:

$$\frac{n_{00} + n11}{n_{00} + n_{10} + n01 + n_{11}} = v \tag{2.19}$$

bzw.

$$\frac{n_{00} + n11}{n_0 + n_1} = v \tag{2.20}$$

und deshalb

$$n_0 \frac{n_{00}}{n_0} + n_1 \frac{n_{11}}{n_1} = v(n_0 + n_1) \tag{2.21}$$

Mit $S^+ = \frac{n_{11}}{n_1}$ und $S^- = \frac{n_{00}}{n_0}$ ergibt sich

$$n_0 S^- + n_1 S^+ = v(n_0 + n_1) \tag{2.22}$$

Da $n_0 = n - n_1$ dementsprechend

$$nS^- + n_1(S^+ - S^-) = nv \tag{2.23}$$

Für den Fall, dass $n_0 = n_1$ und damit auch $n_1 = \frac{n}{2}$ vereinfacht sich Gleichung 2.23 zu

$$nS^- + \frac{n}{2}(S^+ - S^-) = nv \tag{2.24}$$

oder

$$S^+ + S^- = 2v \tag{2.25}$$

Also ist schließlich

$$S^+ = 2v - 1 + (1 - S^-) \tag{2.26}$$

Für den gewünschten Fall $v = 0,5$ liegen dementsprechend alle gesuchten Punkte auf der Geraden, die durch $S^+ = (1-S^-)$ definiert wird, was der eingezeichneten Winkelhalbierenden im Diagramm entspricht. Das gleiche Ergebnis erhält man direkter, wenn $\Delta_{UAPV} = 0$ gesetzt wird. In diesem Fall ist

$$S^+ + S^- - 1 = 0 \tag{2.27}$$

und damit erneut

$$S^+ = (1 - S^-). \tag{2.28}$$

Weiterhin ist es der Fall, dass alle Algorithmen, deren Ergebnis in den Bereich oberhalb der Winkelhalbierenden fällt, eine Klassifikationsgenauigkeit größer 0,5 und einen Δ_{UAPV}-Score größer als Null aufweisen müssen, alle Algorithmen, deren Ergebnis in den Bereich unterhalb der Winkelhalbierenden fällt, eine Klassifikationsgenauigkeit kleiner als 0,5 und einen Δ_{UAPV}-Score kleiner als Null. Jedem Ergebnis, das oberhalb der Winkelhalbierenden liegt, entspricht ein Ergebnis, das auf der Winkelhalbierenden liegt und die gleiche Spezifität repräsentiert, d.h. es unterscheidet sich im Ergebnis von diesem Verfahren durch eine höhere Sensitivität. Damit ist der Δ_{UAPV}-Score definitionsgemäß größer als der des Verfahrens auf der Winkelhalbierenden, und die Accuracy muss größer sein aus der Überlegung, dass bei konstantem n_1 gilt: $S_a^+ > S_b^+ \rightarrow n_{11_a} > n_{11_b}$ und sich damit der Zähler bei der Berechnung der Genauigkeit erhöht, während der Nenner konstant bleibt. Aus diesem Grund wird der Bereich unterhalb der Winkelhalbierenden schattiert, um zu markieren, dass Algorithmen, deren Ergebnisse in diesen Bereich fallen, eine schlechtere Genauigkeit aufweisen als eine Zufallszuordnung der Objekte zu den beiden Klassen.

Aus dem Diagramm ist allerdings noch mehr zu entnehmen. Nennt man eine Gerade, deren Punkte alle die gleiche Klassifikationsgenauigkeit implizieren eine Iso-Genauigkeits-Gerade, so gibt es unendlich viele Geraden dieser Art: Aus Gleichung 2.26 wird ersichtlich, dass die Funktionsschar (für $x = 1 - S^-$)

$$f(x, v) = 2v - 1 + x \tag{2.29}$$

als Wertebereich die reellen Zahlen hat. Der zulässige Definitionsbereich für x als Differenz zwischen 1 und der Spezifität ist das Intervall $[0; 1]$, der zulässige Definitionsbereich für v als Klassifikationsgenauigkeit besteht aus dem gleichen Intervall. Für jedes $x, y \in [0; 1]$ gibt es zudem ein $v \in [0; 1]$, so dass $f(x, v) = y$, dass sich als $v = \frac{1}{2}(y + 1 - x)$ berechnen lässt. Da die Iso-Genauigkeitsgeraden parallel zur Winkelhalbierenden sind, lässt sich damit der Vergleich zwischen zwei Ergebnissen vornehmen, in dem der Abstand zur Winkelhalbierenden verglichen wird. Liegen beide Ergebnisse oberhalb der Winkelhalbierenden, entspricht dem Ergebnis eine höhere Klassifikationsgenauigkeit, das den größeren Abstand aufweist; unterhalb der Winkelhalbierenden gilt das Umgekehrte. In Abbildung 2.6 sind alle Klassifikationsgeraden in $\Delta A = 0,05$ Schritten zwischen $A = 0$ und $A = 1$ abgebildet.

Alternativ lassen sich die Linien auch als Iso-Δ_{UAPV}-Linien verstehen, in diesem Fall mit einer Schrittgröße von $0,1$.

2.4.3 Iso-Klassifikationsgenauigkeitslinien für $n_1 \neq n_0$

Die Lage der Klassifikationsgenauigkeitslinien verändert sich, wenn $n_1 \neq n_0$. Lediglich für den Δ_{UAPV}-Score gibt es in der Ablesbarkeit im Diagramm keinen Unterschied zur Situation mit $n_1 = n_0$, da er sich direkt aus Sensitivität und Spezifität berechnen lässt ohne Berücksichtigung der Gruppengrößen. Die Iso-Linien bleiben für diesen Score die gleichen. In diesem Fall kann aber Gleichung 2.23 nicht wie oben vereinfacht werden, statt dessen ergibt sich

$$nS^- + n_1(S^+ - S^-) = nv \tag{2.30}$$

Nach Umstellung erhält man

$$\frac{n_1}{n}S^+ = v - S^-(1 - \frac{n_1}{n}) \tag{2.31}$$

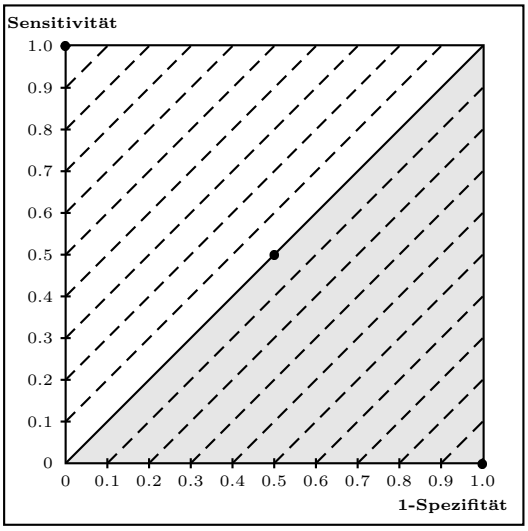


Abbildung 2.6: ROCS-Diagramm für $n_1 = n_0$ mit Iso-Genauigkeitslinien von A=0 (rechts unten) bis A=1 (links oben) in Schritten mit $\Delta A = 0,05$

Da aber in jedem Fall $1 - \frac{n_1}{n} = \frac{n_0}{n}$ ergibt sich

$$\frac{n_1}{n} S^+ = v - \frac{n_0}{n}(S^-) \tag{2.32}$$

$$S^+ = \frac{n}{n_1} v - \frac{n_0}{n_1}(-1 + 1 - S^-) \tag{2.33}$$

und letztendlich

$$S^+ = \frac{n_0}{n_1}(1 - S^-) + \frac{vn - n_0}{n_1}. \tag{2.34}$$

Zur Prüfung der Plausibilität dieser Gleichung sollen erneut einige Strategien getestet werden: Für die Zufallsauswahl liegt die Genauigkeit unverändert bei $A_Z = 0,5$ mit $S^- = 0,5$ und $S^+ = 0,5$. Die Iso-Genauigkeitslinie ergibt sich dementsprechend als

$$f(1 - S^-; 0,5) = \frac{n_0}{n_1}(1 - S^-) + \frac{n_1 - n_0}{2n_1} \tag{2.35}$$

Der vertikale Abstand für $\Delta A = d$ ist zudem:

$$f(1 - S^-; v) - f(1 - S^-; v - d) = \frac{vn - n_0}{n_1} - \frac{(v - d)n - n_0}{n_1} = d\left(1 + \frac{n_0}{n_1}\right). \tag{2.36}$$

Damit lassen sich für ein Beispiel mit $n_1 = 200$ und $n_0 = 100$ die Iso-Linien mit $\Delta A = 0,05$ bestimmen, wie in Abbildung 2.7 eingetragen. Die Winkelhalbierende ist als gepunktete Linie eingetragen, und im Vergleich zeigt sich, dass alle Linien die Winkelhalbierende im gleichen Winkel ($\alpha = \arctan(\frac{n_0}{n_1} - 1)$) schneiden. Die Linien gelten für alle Beispiele, in denen $\frac{n_0}{n1} = \frac{1}{2}$ bzw. $n_1 = 2n_0$. Im Beispiel mit $n_1 = n_0$ lagen die Punkte, die dem Ergebnis der Klassifikationsregel entsprachen, alle Objekte in eine Kategorie einzusortieren, ebenfalls auf der Iso-Linie mit $A = 0,5$. Bei ungleichen Gruppengrößen muss man jetzt aber zwei Strategien unterscheiden: Eine Zuordnung aller Objekte zu einer Kategorie führt zu einer Genauigkeit von $A_M = \frac{n_0}{n}$ bei Zuordnung zur Klasse 0 und zu einer Genauigkeit von $A_M = \frac{n_1}{n}$ bei Zuordnung zur Klasse 1. Die entsprechenden Berechnungsvorschriften für die beiden Iso-Linien sind deshalb:

$$f(1 - S^-; \frac{n_0}{n}) = \frac{n_0}{n_1}(1 - S^-) \tag{2.37}$$

und

$$f(1 - S^-; \frac{n_1}{n}) = \frac{n_0}{n_1}(1 - S^-) + 1 - \frac{n_0}{n_1}. \tag{2.38}$$

Ist das Ziel eine möglichst hohe Klassifikationsgenauigkeit, dann ist in jedem Fall außer für $n_0 = n_1$ genau eine der beiden Strategien eine sinnvolle Strategie. Ist $n_0 < n_1$, dann ist nur die Strategie bezüglich

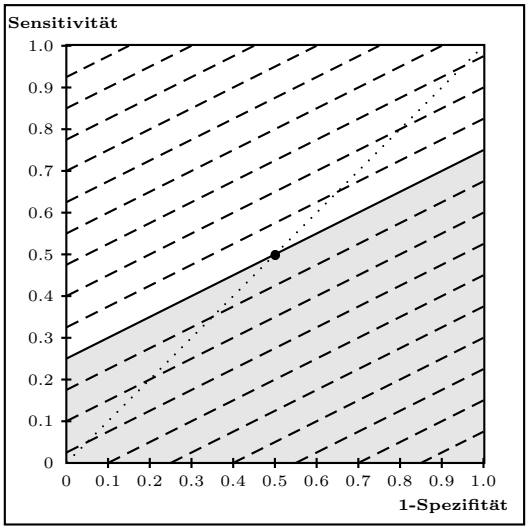


Abbildung 2.7: ROCS-Diagramm für $n_1 = 2n_0$ mit Iso-Genauigkeitslinien von A=0 (rechts unten) bis A=1 (links oben) in Schritten mit $\Delta A = 0,05$

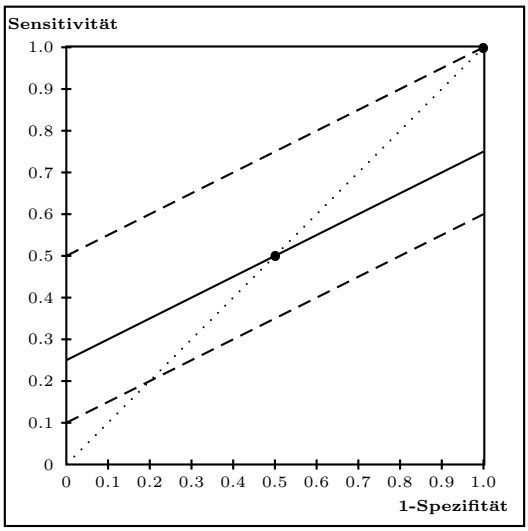


Abbildung 2.8: ROCS-Diagramm für $n_1 = 2n_0$ mit Standard-Iso-Genauigkeitslinien für Zufallsauswahlstrategie Z und Majoritätsregel M, sowie für $A = 0,4$

des Kriteriums sinnvoll, alle Objekte in die Klasse 1 einzuordnen und umgekehrt. In Abbildung 2.8 sind die Iso-Genauigkeitslinien für die Zufallsauswahl und die Majoritätsregel eingezeichnet. Zur besseren Orientierung wird ebenfalls die Iso-Linie für $A = 0,4$ eingezeichnet. Der Abstand zwischen den beiden unteren Linien entspricht damit genau einer Klassifikationsgenauigkeitsdifferenz von $\Delta A = 0,1$. Vielfache des Abstandes entsprechen damit auch Vielfachen von $0,1$.

2.4.4 Beispieldiagramme

Ein Beispielergebnis für sechs fiktive Klassifikationsverfahren soll den Nutzen dieser Diagrammform verdeutlichen. Tabelle 2.8 enthält die Ergebnisse der sechs Verfahren und in Abbildung 2.9 sind die Ergebnisse in ein ROCS-Diagramm eingetragen. Es gilt, anders als im Beispiel zuvor, $\frac{n_0}{n1} = \frac{1}{3}$ mit $n_1 = 300$ und $n_0 = 100$.

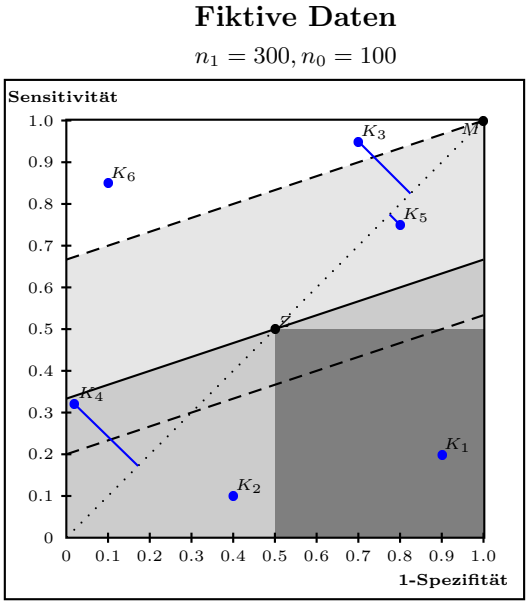


Abbildung 2.9: ROCS-Diagramm für sechs fiktive Klassifikationsalgorithmen ($n_1 = 300$, $n_0 = 100$)

Aus diesem Diagramm lässt sich die Reihenfolge der Verfahren bezüglich verschiedener Kriterien erkennen: Das Ablesen der y-Koordinaten ergibt: $S_3^+ > S_6^+ > S_5^+ > S_4^+ > S_1^+ > S_2^+$. Die x-Koordinate ergibt unter Berücksichtigung der Invertierung $S_4^- > S_6^- > S_2^- > S_3^- > S_5^- > S_1^-$. Die Lage der Iso-Genauigkeitslinie, die den Punkt schneiden würde ergibt sich die Reihenfolge bezüglich der Klassifikationsgenauigkeit: $A_6 > A_3 > A_5 > A_4 > A_2 > A_1$. Schließlich resultiert die Reihenfolge der Δ_{UAPV}-Scores bei Berücksichtigung der Abstände von der Winkelhalbierenden in der Form $\Delta_{UAPV6} > \Delta_{UAPV4} > \Delta_{UAPV3} > \Delta_{UAPV5} > \Delta_{UAPV2} > \Delta_{UAPV1}$.

Tabelle 2.8: Ergebnisse für sechs fiktive Klassifikationsalgorithmen ($n_1 = 300$, $n_0 = 100$)

Algorithmus	n_{11}	n_{10}	n_{01}	n_{00}	S^+	S^-	Δ_{UAPV}	A
K_1	60	240	90	10	0,20	0,10	-0,70	0,175
K_2	30	270	40	60	0,10	0,60	-0,30	0,225
K_3	285	15	70	30	0,95	0,30	0,25	0,788
K_4	96	204	2	98	0,32	0,98	0,30	0,485
K_5	225	75	80	20	0,75	0,20	-0,05	0,613
K_6	255	45	10	90	0,85	0,90	0,75	0,863
Z	150	150	50	50	0,50	0,50	0,00	0,500
M	300	0	100	0	1,00	0,00	0,00	0,75

Im Diagramm sind drei verschiedene Graustufen verwendet worden. Der Bereich von Ergebnissen, der sowohl hinsichtlich Sensitivität als auch hinsichtlich der Spezifität unter $0,5$ liegt, so dass Strategien mit einem Ergebnis in diesem Bereich von der Zufallsauswahl-Strategie dominiert werden, ist im dunkelsten

Ton schattiert. Für den Klassifikationsalgorithmus 1 trifft dieser Umstand zu. Es ist schwer, eine Situation zu konzipieren, in der ein Algorithmus mit vergleichbarem Ergebnis nützlich sein kann. Allerdings ließe sich ein guter Klassifikationsalgorithmus bilden, der jeweils die entgegengesetzte Vorhersage zu K_1 macht. Dieser Algorithmus mit einer Spezifität von $S_{\bar{1}}^{-} = 1 - (1 - 0,9) = 0,9$ und einer Sensitivität von $S_{\bar{1}}^{+} = 1 - 0,2 = 0,8$ würde sogar sehr gut gegen die anderen Algorithmen abschneiden können. Trotzdem ist das Ergebnis möglich, wenn beispielsweise die Verhältnisse in der Trainingsstichprobe denen in der Teststichprobe entgegengesetzt sind. Das Ergebnis wäre dementsprechend lediglich als *Fitting*-Ergebnis gleichbedeutend mit der Diagnose der Unsinnigkeit der dahinter liegenden Strategie und einer Schwachstelle des Klassifikationslalgorithmus. Da dieser Bereich im Allgemeinen nicht besetzt sein sollte, wird er in zukünftigen Diagrammen nicht schattiert.

Amelang und Zielinski (2002) würden Entscheidungsregeln in diesem Quadranten als unzulässige Regeln bezeichnen, da es mit der Zufallsauswahl mindestens eine Regel gibt, die besser ist:

> „Besser“ bedeutet mindestens so gute Trefferraten in jeder der Klassen und Überlegenheit in mindestens einer weiteren. (Amelang & Zielinski, 2002, S.445)

Im Bereich der mittleren Schattierung liegen Ergebnisse, deren Genauigkeit kleiner ist als die der Zufallsauswahl, die aber nicht von der Strategie dominiert werden, da entweder eine höhere Sensitivität oder eine höhere Spezifität vorliegt. Dies trifft für die Klassifikatoren K_2 und K_4 zu. Im hell schattierten Bereich liegen Strategien, deren Klassifikationsgenauigkeit höher liegt als die der Zufallsauswahl, deren Klassifikationsgenauigkeit aber nicht die Genauigkeit der Majoritätsregel erreicht. Dies trifft im Beispiel für den Klassifikationsalgorithmus 5 zu. Dabei sollte darauf hingewiesen werden, dass die Majoritätsregel nur Strategien dominieren könnte, deren Spezifität bei Null liegt und die Sensitivitäten kleiner als 1 aufweisen. Dementsprechend wird keine der sechs Strategien von der Majoritätsregel dominiert, auch wenn zwei Drittel der Strategien eine schlechtere Genauigkeit aufweisen. Im nicht schattierten Bereich schließlich liegen Strategien, deren Genauigkeit größer ist als die von Zufallsregel und Majoritätsregel, wie die von Klassifikationsalgorithmen 3 und 6. Für Klassifikationsalgorithmus 6 trifft zusätzlich zu, dass diese Strategie die Zufallsauswahlstrategie dominiert, da sie sich im linken oberen Quadranten befindet mit $S^{+} > 0,5 \wedge S^{-} > 0,5$.

Das Beispiel hilft ebenfalls zu veranschaulichen, dass die Verwendung von Klassifikationsgenauigkeit und Δ_{UAPV}-Score eine unterschiedliche Rangreihe der Verfahren erzeugen. Dies kann an den Ergebnissen von K_3, K_4 und K_5 nachvollzogen werden, für die die Abstände zur Winkelhalbierenden ins Diagramm eingetragen wurden. Obwohl die Klassifikationsgenauigkeit von K_4 unterhalb der Klassifikationsgenauigkeit der Zufallsauswahl liegt und die Genauigkeit von K_3 oberhalb der Genauigkeit der Majoritätsregel, hat K_3 den höheren Δ_{UAPV}-Score. Umgekehrt wird K_5 hinsichtlich Klassifikationsgenauigkeit höher bewertet als K_4, obwohl der Δ_{UAPV}-Score von K_5 sogar unter dem der Zufallsauswahl liegt.

Tabelle 2.9: Ergebnisse für sechs fiktive Klassifikationsalgorithmen ($n_1 = 100$, $n_0 = 300$)

Algorithmus	n_{11}	n_{10}	n_{01}	n_{00}	S^{+}	S^{-}	Δ_{UAPV}	A
K_1	20	80	270	30	0,20	0,10	-0,70	0,125
K_2	10	90	120	180	0,10	0,60	-0,30	0,475
K_3	95	5	210	90	0,95	0,30	0,25	0,463
K_4	32	68	6	294	0,32	0,98	0,30	0,815
K_5	75	25	240	60	0,75	0,20	-0,05	0,338
K_6	85	15	30	270	0,85	0,90	0,75	0,888
Z	50	50	150	150	0,50	0,50	0,00	0,500
M	0	100	0	300	0,00	1,00	0,00	0,750

Die Verhältnisse verändern sich, wenn die umgekehrte Situation vorliegt und $n_1 = 100$ und $n_0 = 300$. Angenommen, sechs Klassifikationsalgorithmen hätten jetzt die gleiche Sensitivität und Spezifität wie zuvor. Das entspräche den Ergebnissen in Tabelle 2.9 und der Visualisierung in Abbildung 2.10. In diesem Fall würde die Majoritätsregel alle Fälle der Klasse 0 zuordnen. Von den Gütekriterien hat sich hier lediglich die Klassifikationsgenauigkeit gegenüber der vorherigen Auswertung geändert; die neue Rangfolge ergibt sich analog zur Lage der Punkte relativ zu den neuen Iso-Linien als $A_6 > A_4 > A_2 >$

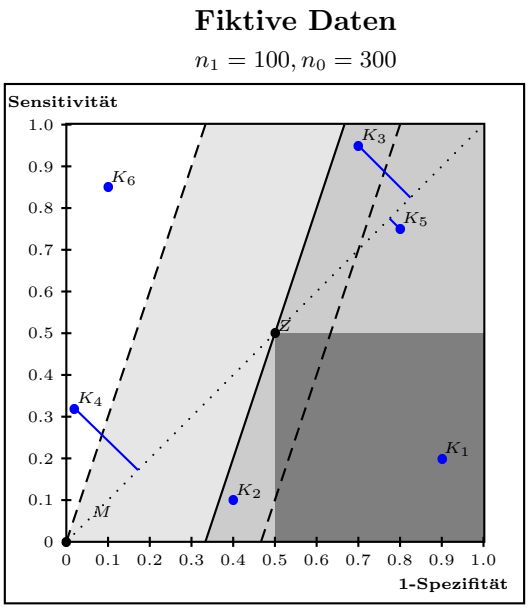


Abbildung 2.10: ROCS-Diagramm für sechs fiktive Klassifikationsalgorithmen ($n_1 = 100$, $n_0 = 300$)

$A_3 > A_5 > A_1$. Die Rangposition aller Strategien mit Ausnahme der besten und schlechtesten hat sich damit verschoben.

Weitere Beispiele für ROCS-Diagramme für verschiedene Klassengrößenverhältnisse sind in Abbildung 2.11 zusammengefasst. Dabei wird deutlich, dass das Verhältnis der Katheten des Dreiecks, das durch die Iso-Linie der Majoritätsregel und zwei Achsenabschnitte gebildet wird, dem Verhältnis der Klassengrößen entspricht. Dies lässt sich in einem Satz festhalten:

Satz 2.3 *ROCS-Streckenverhältnisse Bezeichnet man im ROCS-Diagramm den Schnittpunkt der Iso-Klassfikationsgenauigkeitsgeraden für $A = A_M$ als P, den Schnittpunkt dieser Geraden mit der durch $S^+ = 1$ definierten Geraden als Q und den Punkt $(1,1)$ als R, so gilt $\frac{l(\overline{PR})}{l(\overline{QR})} = \frac{n_0}{n_1}$.*

Beweis: Unterschieden werden die drei Fälle **(i)** $n_1 > n_0$, **(ii)** $n_1 = n_0$ und **(iii)** $n_1 < n_0$. **(i)** Angenommen $n_1 > n_0$ und ein Klassifikationsergebnis j mit einer Spezifität von 1 und $A_j = A_M$. Für alle Ergebnisse mit einer Spezifität von 1 gilt $n_{00} = n_0$. Aus der Definition von A folgt $n_{11} = nA - n_{00}$ und für diesen Fall also $n_{11} = nA - n_0$. Da $S_j^+ = \frac{n_{11}}{n_1}$, gilt also $n_1 S_j^+ = nA - n_0$. Für $A_M = \frac{\max(n_0,n_1)}{n} = \frac{n_1}{n}$ gilt dementsprechend $S_j^+ = \frac{n_1 - n_0}{n_1} = 1 - \frac{n_0}{n_1}$. P hat die Koordinaten $(0, S_j^+)$ und existiert für $n_1 > 0$. Die Iso-Linie der Majoritätsregel verläuft für $n_1 > n_0$ durch den Punkt $(1,1)$, der den Punkt Q darstellt. Das Verhältnis der Streckenlängen ist in diesem Fall $\frac{1-S_j^+}{1} = 1 - (1 - \frac{n_0}{n_1}) = \frac{n_0}{n_1}$.

(ii) Angenommen $n_1 = n_0$ und ein Ergebnis mit einer Spezifität von 1. In diesem Fall ist $A_M = A_Z$, da $n_{00} = n_{11} = \frac{1}{2}n$. In diesem Fall hat P die Koordinaten $(0,0)$ und Q die Koordinaten $(1,1)$, und der Quotient der Streckenlängen ist $\frac{1}{1} = 1$, was in diesem Fall dem Quotienten $\frac{n_0}{n_1} = 1$ entspricht.

(iii) Angenommen $n_1 < n_0$ und ein Ergebnis mit einer Sensitivität von 1. Für alle Ergebnisse mit einer Sensitivität von 1 gilt $n_{11} = n_1$. Aus der Definition von A folgt $n_{00} = nA - n_{11}$ und für diesen Fall also $n_{00} = nA - n_1$. Da $S_j^- = \frac{n_{00}}{n_0}$, gilt also $n_0 S_j^- = nA - n_1$. Für $A_M = \frac{\max(n_0,n_1)}{n} = \frac{n_0}{n}$ gilt dementsprechend $S_j^- = \frac{n_0 - n_1}{n_0} = 1 - \frac{n_1}{n_0}$. R hat die Koordinaten $(1 - S_j^-, 1)$, da Werte auf der x-Achse den Werten $1 - S^-$ entsprechen, und existiert für $n_0 > 0$. Die Iso-Linie der Majoritätsregel verläuft für $n_1 < n_0$ durch den Punkt $(0,0)$, der den Punkt P darstellt. Das Verhältnis der Streckenlängen ist in diesem Fall erneut $\frac{1}{1-S_j^-} = \frac{1}{1-(1-\frac{n_1}{n_0})} = \frac{n_0}{n_1}$. ■.

Das nicht eingefärbte Dreieck links oben im Diagramm kann also herangezogen werden, um die Stichprobenklassenverhältnisse (nicht aber die absoluten Größen) zu rekonstruieren. Soll an Stelle der Klassifikationsgenauigkeit die Position der Diagonalen durch das KNMK bestimmt werden, so hilft die Überlegung, dass das KNMK mit $\delta_0 = \delta_1$ und $n_0 = n_0^*$, $n_1 = n_1^*$ zur gleichen Reihenfolge der Verfahren führt wie ein KNMK mit $\delta_0 = g\delta_1$ und $n_0 = gn_1^*$, $n_1 = n_1^*$, wenn S^+ und S^- in beiden Fällen identisch sind. Allgemein lässt sich für ein Diagramm in diesem Fall für δ_0, δ_1, n_0 und n_1 die Einstellung $n_{0d} = \delta_0 \cdot n_0$ und $n_{1d} = \delta_1 \cdot n_1$ wählen. Setzt man n_{0d} und n_{1d} als Stichprobengrößen ein, so lässt sich am ROCS-Diagramm bei Standardberechnung von Sensitivität und Spezifität der Unterschied der Verfahren hinsichtlich des KNMK bestimmen. Das Dreieck im Diagramm enthält in diesem Fall allerdings nicht Informationen

über die Stichprobenverhältnisse, sondern über den Quotienten $\frac{\delta_0 n_0}{\delta_1 n_1}$. Da das Kriterium einer simplen Gewichtung der Stichproben entspricht, wird auf eine erneute Visualisierung an dieser Stelle verzichtet.

Alle Diagramme in diesem Abschnitt wurden mit GIER$_{[1.0]}$.ROCS erstellt, das ein Teilmodul der GIER$_{[1.0]}$-Software darstellt und ROCS-Diagramme nach Eingabe der Klassengrößen als LaTeX-Code ausgibt.

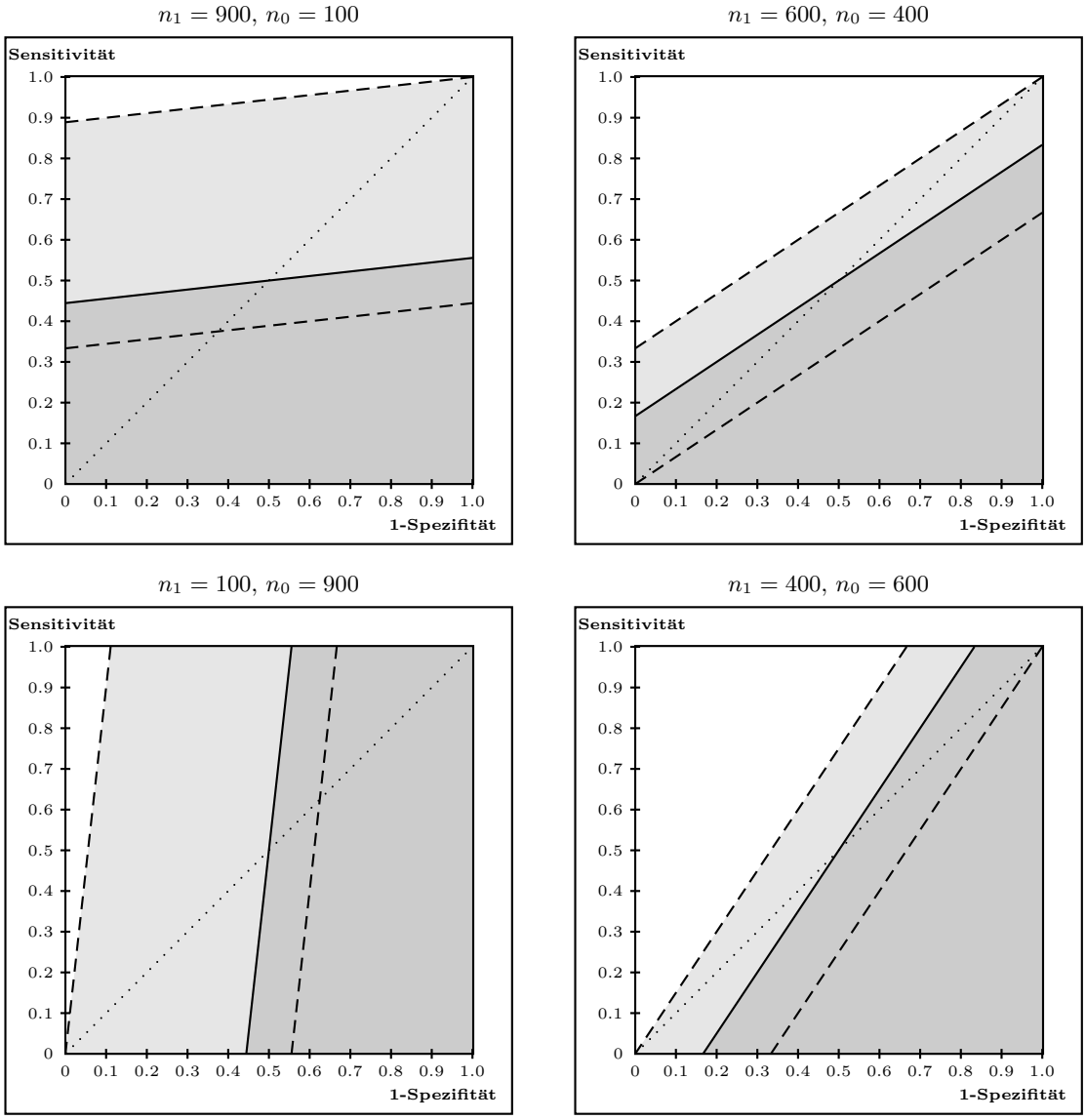


Abbildung 2.11: Beispiele für ROCS-Diagramme bei unterschiedlichen Klassengrößen: Diagramme für $n_1 = 900$, $n_0 = 100$ (links oben), $n_1 = 600$, $n_0 = 400$ (rechts oben), $n_1 = 100$, $n_0 = 900$ (links unten) und $n_1 = 400$, $n_0 = 600$ (rechts unten)

2.5 Die Ergebnisse des Algorithmenvergleichs bei Montel (2005)

An dieser Stelle sollen die Ergebnisse eines Algorithmenvergleichs bei Montel (2005) dargestellt werden, der aufgrund mehrerer Faktoren als relevant für die vorliegende Arbeit angesehen wird:

- Verwendung identischer Ausgangsdaten: Der Algorithmenvergleich wird an Datensätzen vorgenommen, die auch in dieser Arbeit zum Einsatz kommen.
- Verwendung genetischer Algorithmen: Auch bei Montel (2005) kommt ein genetischer Algorithmus zum Einsatz, der sich allerdings in seiner Zielsetzung und den verwendeten Repräsentationen von den in dieser Arbeit verwendeten Algorithmen unterscheidet.

- Vorgabe von Referenzwerten: In der Arbeit finden sich Hinweise auf die Leistungsfähigkeit von Logistischer Regression, linearer Diskriminanzanalyse und dem CART-Verfahren für die Klassifikation der Datensätze.

Allerdings gibt es auch einige wesentliche Differenzen, die eine Re-Analyse der Daten in der vorliegenden Arbeit als nicht-redundant erscheinen lassen:

- Ausweitung der Algorithmenmenge: Neu konstruierte Algorithmen und Heuristiken sollen an den Datensätzen überprüft werden.
- Änderung des Zielkriteriums: Wie oben geschildert, kommt in dieser Arbeit ein anderes Zielkriterium zur Anwendung.
- Verlegung des Schwerpunkts von Modellbildung auf Algorithmenvergleich: Nicht die Erstellung von Klassifikationsmodellen steht bei dieser Arbeit im Vordergrund, sondern die Beurteilung der differenziellen Leistungsfähigkeit von potenziell erstellbaren Modellen durch verschiedene Algorithmengruppen.
- Variation der Trainingsstichprobengröße: In dieser Arbeit wird die Trainingsstichprobengröße systematisch variiert, um die Abhängigkeit der Algorithmenleistung von Stichprobengrößen an Beispielen abzuschätzen. Diese Methode tritt an die Stelle der *leave-one-out*-Kreuzvalidierung.

2.5.1 Auswahl der Klassifikationsaufgaben und Algorithmen bei Montel

In der ursprünglichen Arbeit werden sechs Algorithmen miteinander verglichen: zwei Varianten der logistischen Regression (LR_s, LR_g), zwei Varianten der linearen Diskriminanzanalyse DA_s, DA_g), das CART-Verfahren (vgl. Breiman, H.Friedman, Olshen & Stone, 1984) und ein genetischer Algorithmus (SOMC). Die Varianten von logistischer Regression und Diskriminanzanalyse unterscheiden sich jeweils darin, dass in einem Fall A-Priori-Wahrscheinlichkeiten aus der Stichprobe geschätzt werden, im anderen Fall eine Gleichverteilung angenommen wird. Dabei ist Folgendes zu beachten:

> Einer Notiz bedürfen die Konsequenzen der unterschiedlichen Parametereinstellungen in Diskriminanzanalyse und logistischer Regression, die sich nicht auf die Modellbildung im Sinne der Parameterschätzung, sehr wohl jedoch auf die Klassifikationsergebnisse auswirken. Erhöht man im Rahmen der Klassifikation im Anschluss an eine Diskriminanzanalyse die Priorwahrscheinlichkeit einer betrachteten Gruppe, so hat dies zur Konsequenz, dass ein Datensatz mit höherer Wahrscheinlichkeit in diese Gruppe klassifiziert wird [...]. Im Gegensatz dazu führt innerhalb der logistischen Regression die Erhöhung [...] dazu, dass ein Fall mit geringerer Wahrscheinlichkeit in die betrachtete Gruppe klassifiziert wird [...]. (Montel, 2005, S. 147)

Aus den bei Montel (2005) angegebenen Gütekriterien für die verwendeten Algorithmen sollen hier zwei herausgegriffen werden: Klassifikationsgenauigkeit (A) und das Δ_{UAPV}-Kriterium.

2.5.2 Auswahl der Datensätze

Der Arbeit von Montel liegen drei Datensätze zugrunde, die im Einzelnen kurz vorgestellt werden sollen.

2.5.2.1 Der Datensatz Hochschule

Der Datensatz „Hochschule“ stammt aus den Jahren 2002 und 2003 und wurde im Rahmen der Traineeauswahl eines international agierenden Großkonzerns erhoben. Der Name des Datensatz leitet sich von der Zusammensetzung der Stichprobe aus Hochschulabsolventen ab, die sich für die Trainee-Stellen beworben haben.

Ziel der Datenerhebung war die Einrichtung einer internetbasierten Testbatterie zur Personalvorauswahl im Rahmen einer sequenziellen Diagnosestrategie (vgl. Woike, 2003b). Bisher wurde aufgrund einer Dokumentenanalyse entschieden, welche Bewerber zu einem teilstrukturierten Interview eingeladen wurden. Als Ergebnis sollte dementsprechend sowohl die Zusammenstellung einer Testbatterie als auch die

Festlegung eines Anforderungsprofils stehen, um eine Negativauswahl der anfänglichen Bewerber vornehmen zu können (vgl. Wottawa, 2000). Die Daten wurden dementsprechend als zusätzliche Informationen zur bestehenden Auswahlprozedur erhoben. Von 191 Bewerbern wurden 138 positiv (72,25%), 53 (27,75%) negativ bewertet. Grundlage für diese Bewertung war das sechsstufige Urteil eines Recruiters nach dem Interview, als negatives Kriterium wurde dabei die Vergabe einer der beiden schlechtesten Bewertungen gezählt. Die vierzehn ausgewählten Testskalen waren dabei die folgenden:

- Gestaltungsmotiv aufsuchend, Gestaltungsmotiv meidend
- Aktive Problemlösefähigkeit, Flexibilität im Denken, Kreativität mit Regelbeachtung
- Emotionale Konfliktreaktion,Vermeiden von Konflikten, Kooperationsorientierung
- Mitarbeiterorientierung, Perspektivenübernahme
- Aktives Innovationsstreben, Problemlösebereitschaft, Ehrgeiz/Belastungsbereitschaft, Lernbereitschaft

Alle Skalen wurden im Rahmen des Systems PERLS vorgegeben.

2.5.2.2 Der Datensatz Beratungsunternehmen

Beim zweiten Datensatz wird die Stichprobe durch die Testergebnisse von Personen gebildet, für die bereits eine positive Vorauswahlentscheidung getroffen wurde. In dem betrachteten Beratungsunternehmen fallen 40 000 Bewerber, die sich erneut hauptsächlich aus Hochschulabsolventen zusammensetzen, auf ca. 1 200 im deutschsprachigen Raum vakante Stellen.

Ziel der Klassifikation ist im Anwendungsbeispiel die Identifizierung von Bewerbern, die im Folgenden in einer beratenden Position einsetzbar sind. Die übrigen Stellen zeichnen sich dabei eher durch operative Tätigkeiten aus und sind dementsprechend geringer dotiert. Für 437 Bewerber wurden aufgrund der Analyse von biografischen Daten 318 (72,77%) Personen als für die Beratung geeignet eingestuft (positives Kriterium), während 119 (27,23%) für die eher operativen Tätigkeiten vorgesehen wurden (negatives Kriterium). Dabei kamen zwischen 2004 und 2005 die folgenden sechzehn Skalen zum Einsatz:

- Ideenreichtum, Problemlösungsbereitschaft, Flexibilität, Schlussfolgerndes Denken
- Kontaktfreude, Streben nach sozialer Akzeptanz, Einfühlungsvermögen
- Lernbereitschaft, Misserfolge meiden
- Führungsmotivation, Leistungsmotivation
- Kritikreaktion, Durchsetzungsvermögen, Konsensorientierung
- Ursachenzuschreibung (external vs. internal), Umgang mit unscharf definierten Situationen

2.5.2.3 Der Datensatz Schüler

Der dritte Datensatz wird durch die Testergebnisse von Schülern gebildet, die an dem Berufsorientierungstest eines Finanzdienstleisters teilgenommen haben, der als Papier-und-Bleistift-Verfahren in den Klassen durchgeführt und zentral ausgewertet wurde. Die Daten von 1993 Gymnasiasten der Jahrgangsstufe 10 wurden ebenfalls in den Jahren 2004 und 2005 erhoben. Aufgrund der parallel erhobenen Schulnoten wurden die Schüler und Schülerinnen zusammengefasst in zwei Gruppen unterteilt: Die ungefähr 71% leistungsstärksten Schüler wurden in die positive Kriteriumsgruppe, die 29% leistungsschwächsten Schüler in die negative Kriteriumsgruppe eingeteilt [13].

[13] Im Rahmen dieser Arbeit wurde eine weniger extreme Gruppenaufteilung in 1223 Schüler mit positivem Kriteriumswert und 770 Schüler mit negativem Kriteriumswert verwendet. Die Ergebnisse für diesen Datensatz sind also nur bedingt vergleichbar, wobei der Datensatz nur für eine Simulation verwendet wird.

Ziel der Klassifikation ist die grobe Identifizierung von Schülern mit Lernproblemen aufgrund von Testergebnissen. Die Aufstellung expliziter Entscheidungsregeln versucht der Schulsituation in Deutschland gerecht zu werden, in der ein einzelner Schulpsychologe mit der Betreuung von ca. 12 000 Schülern beauftragt ist (vgl. Wottawa & Hossiep, 1997).

Eingesetzt wurden die folgenden fünfzehn Testskalen:

- Verantwortungsbereitschaft, Teamfähigkeit, Zurückhaltung im Umgang mit anderen
- Situationsgerechtes Auftreten, Kundenorientierung
- Konflikt- und Kompromissfähigkeit, Effektiver Umgang mit Schwierigkeiten, Durchsetzungsfähigkeit
- Systematik des Lernverhaltens, Leistungsbereitschaft
- Technisches Verständnis, Räumliches Vorstellungsvermögen,Mathematisches Verständnis
- Textverständnis, Logische Denkfähigkeit

2.5.3 Klassifikationsergebnisse

Tabelle 2.10 zeigt die Ergebnisse der Algorithmen hinsichtlich der Klassifikationsgenauigkeit für Fitting (A_F) und Kreuzvalidierung (A_K), Tabelle 2.11 zeigt die Ergebnisse bezüglich des Δ_{UAPV}-Scores. Bei der Betrachtung der Klassifikationsgenauigkeit fällt auf, dass alle Verfahren bei der Kreuzvalidierung mehr oder weniger deutliche Einbußen hinnehmen müssen im Vergleich zur *Fitting*-Aufgabe. Die stärksten Verluste treten hier beim neu konstruierten Verfahren auf. Das relativ schlechte Ergebnis der Verfahrens selbst beim *Fitting* erklärt sich aber durch das verwendete Fitness-Maß: Die Ergebnisse bezüglich des Δ_{UAPV}-Scores sind dementsprechend besser. Die Unterschiede zwischen den beiden Varianten der Logistischen Regression und der Diskriminanzanalyse sind bezüglich beider Maße ausgeprägt.

In der Analyse der Ergebnisstruktur kann zudem beobachtet werden, dass sich die durch CART konstruierten Bäume sowohl durch einfache Strukturen (wenige Endknoten) als auch durch eine Berücksichtigung weniger Variablen auszeichnen: Für den Datensatz *Hochschule* wird ein Baum mit vier Ebenen und vier Endknoten konstruiert, in dem lediglich die Variablenwerte für Aktives Innovationsstreben und Lernbereitschaft abgefragt werden. Auch der Baum für den Datensatz *Beratungsunternehmen* zeigt eine identische Struktur bei Abfrage der Werte für Führungsmotivation, Umgang mit unscharf definierten Situationen und Schlussfolgerndes Denken. Der Baum für den Datensatz *Schüler* weist drei Ebenen und erneut vier Endknoten auf, abgefragt werden Leistungsbereitschaft, Technisches Verständnis und Textverständnis.

Das relativ gute Ergebnis der Entscheidungsbäume verdeutlicht, dass nur ein kleiner Teil der verfügbaren Informationen ausreicht, um darauf Entscheidungen basieren zu können, die von komplexeren Verfahren nicht übertroffen werden. Der Konstruktionsalgorithmus von CART-Bäumen ist kein einfacher Algorithmus und er führt nicht notwendigerweise zu einfachen Bäumen. Die Ergebnisbäume sind zudem das Ergebnis einer nachträglichen Verkleinerung der gefundenen Baumstrukturen (Pruning)[14].

Die CART-Ergebnisse sind einfach in hierarchische Regelsysteme überführbar, anders als die Ergebnisse von Logistischer Regression und der Diskriminanzanalyse. Montel (2005) schließt mit einer persönlichen Hoffnung für die Zukunft der psychologischen Diagnostik:

> Es bleibt zu hoffen, dass die Verfügbarkeit geeigneter Methoden einen bescheidenen Beitrag dazu leisten kann, die Wissensbestände der angewandten Diagnostik so zu erweitern, dass

[14] Allerdings muss eingewandt werden, dass für die Pruning-Entscheidung für die ersten beiden Datensätze die Stabilität der Lösung geprüft wurde (vgl. Montel (2005)), so dass das Kreuzvalidierungsergebnis streng genommen eine Überschätzung der Leistung von CART darstellt:

> Innerhalb der Praxisbeispiele werden Bäume für verschiedene Einstellungen von t erzeugt und letztlich schließlich die Parametereinstellung verwendet, die zu möglichst hoher Stabilität der Ergebnisse in der leave-one-out-Kreuzvalidierung bei einer möglichst hohen Anzahl terminaler Knoten führt. (Montel, 2005, S. 128)

Für eine exakte Schätzung müsste die Auswahl der Lösung unter mehreren Varianten allein aufgrund der zur Verfügung stehenden Trainingsstichprobe erfolgen, da ansonsten ein Overfitting auch auf der Ebene der Auswahl von Verfahren aufgrund von Generalisierungsergebnissen zu zu erwarten ist (vgl. Witten & Frank, 2001).

deren Leistungsfähigkeit steigt. Gerade der Einsatz von in der Psychologie weniger geläufigen Auswertungsmethoden, die einfach verbal explizierbare Ergebnisse liefern, kann hier einen Ansatzpunkt liefern. (Montel, 2005, S. 218)

2.5.4 Visualisierung der Ergebnisse von Montel (2005)

Abschließend soll das vorgestellte Visualisierungsverfahren zum Einsatz kommen, um die Ergebnisse von Montel (2005) in der neu entwickelten Form präsentieren zu können (s. Abbildung 2.12).

In dieser Ansicht werden vor allem zwei Beobachtungen augenfällig: Die meisten Verfahren verzeichnen bei den meisten Datensätzen beim kreuzvalidierten Ergebnis deutliche Einbußen gegenüber dem Fitting-Ergebnis, was die Stabilität der im fitting gefundenen Struktur ernsthaft hinterfragen lässt. Am stabilsten sind die Ergebnisse für den Datensatz *Schüler*, das $SOMC$-Verfahren weist die größten Unterschiede jeweils zwischen den beiden Ergebnissen auf. Zum anderen erreicht kaum ein Verfahren ein Klassifikationsgenauigkeitsergebnis, dass über dem der Majoritätsregel liegt. Beim Datensatz *Hochschule* trifft dies bei der Kreuzvalidierung nur noch für den $CART$-Algorithmus zu Das $SOMC$-Verfahren sinkt in der Kreuzvalidierung unter diese Grenze wie auch die übrigen Verfahren. Bei der Analyse des Datensatzes *Beratungsunternehmen* liegen nur zwei Verfahren knapp über der genannten Grenze, beim Datensatz *Schüler* liegen drei Verfahren knapp über der Grenze. Trotzdem bündeln sich die Punkte nicht an einer Stelle im Koordinatensystem, bspw. weist das $SOMC$-Verfahren in jedem Ergebnis eine höhere Spezifität, dafür eine niedrigere Sensitivität auf, LR_s und DA_g einerseits und LR_g und DA_s andererseits, liegen hingegen in jedem Ergebnis sehr dicht beieinander.

Die durch $\text{GIER}_{[1.0]}$.ROCS erstellte ROCS-Darstellung ermöglicht durch die gleichzeitige Präsentation einen ökonomischeren Vergleich der erzielten Ergebnisse und soll im weiteren Verlauf dieser Arbeit zur Präsentation von Ergebnissen herangezogen werden.

Tabelle 2.10: Algorithmenvergleich aus Montel (2005): Vier Algorithmen mit jeweils zwei Varianten für LDA und LR (gleiche Klassengrößen angenommen oder Schätzung aus Stichprobe) wurden an drei Datensätzen getestet und hinsichtlich Fittinggenauigkeit und kreuzvalidierter Genauigkeit verglichen

Datensatz	Hochschulabsolventen		Beratungsunternehmen		Schüler	
Klassifikationsgenauigkeit	Fitting	Kreuzvalidierung	Fitting	Kreuzvalidierung	Fitting	Kreuzvalidierung
Logistische Regression	75,92% - 68,59%	71,73% - 66,49%	76,20% - 68,65%	76,20% - 68,42%	73,36% - 67,74%	72,91% - 66,88%
Lineare Diskriminanzanalyse	70,16% - 74,87%	63,35% - 70,16%	69,11% - 76,20%	66,13% - 73,68%	67,79% - 73,61%	66,93% - 72,91%
Classification and Regression Trees	82,20%	80,63%	76,89%	67,96%	71,90%	71,70%
Simultane Optimierung von multiplen Cutoffs	80,10%	68,06%	77,80%	67,28%	66,38%	61,41%

Tabelle 2.11: Algorithmenvergleich aus Montel (2005): Vier Algorithmen mit jeweils zwei Varianten für LDA und LR (gleiche Klassengrößen angenommen oder Schätzung aus Stichprobe) wurden an drei Datensätzen getestet hinsichtlich Fitting- und kreuzvalidiertem Δ_{UAPV}-Score verglichen

Datensatz	Hochschulabsolventen		Beratungsunternehmen		Schüler	
Δ_{UAPV}-Score	Fitting	Kreuzvalidierung	Fitting	Kreuzvalidierung	Fitting	Kreuzvalidierung
Logistische Regression	22,50 - 32,12	15,55 - 25,73	25,75 - 37,99	26,80 - 37,67	20,51 - 36,03	19,47 - 33,40
Lineare Diskriminanzanalyse	33,13 - 23,38	17,90 - 11,05	39,14 - 27,33	30,32 - 21,24	36,11 - 21,69	33,57 - 19,88
Classification and Regression Trees	48,63	42,97	36,69	18,64	13,96	13,27
Simultane Optimierung von multiplen Cutoffs	68,98	37,20	62,19	29,26	33,21	25,29

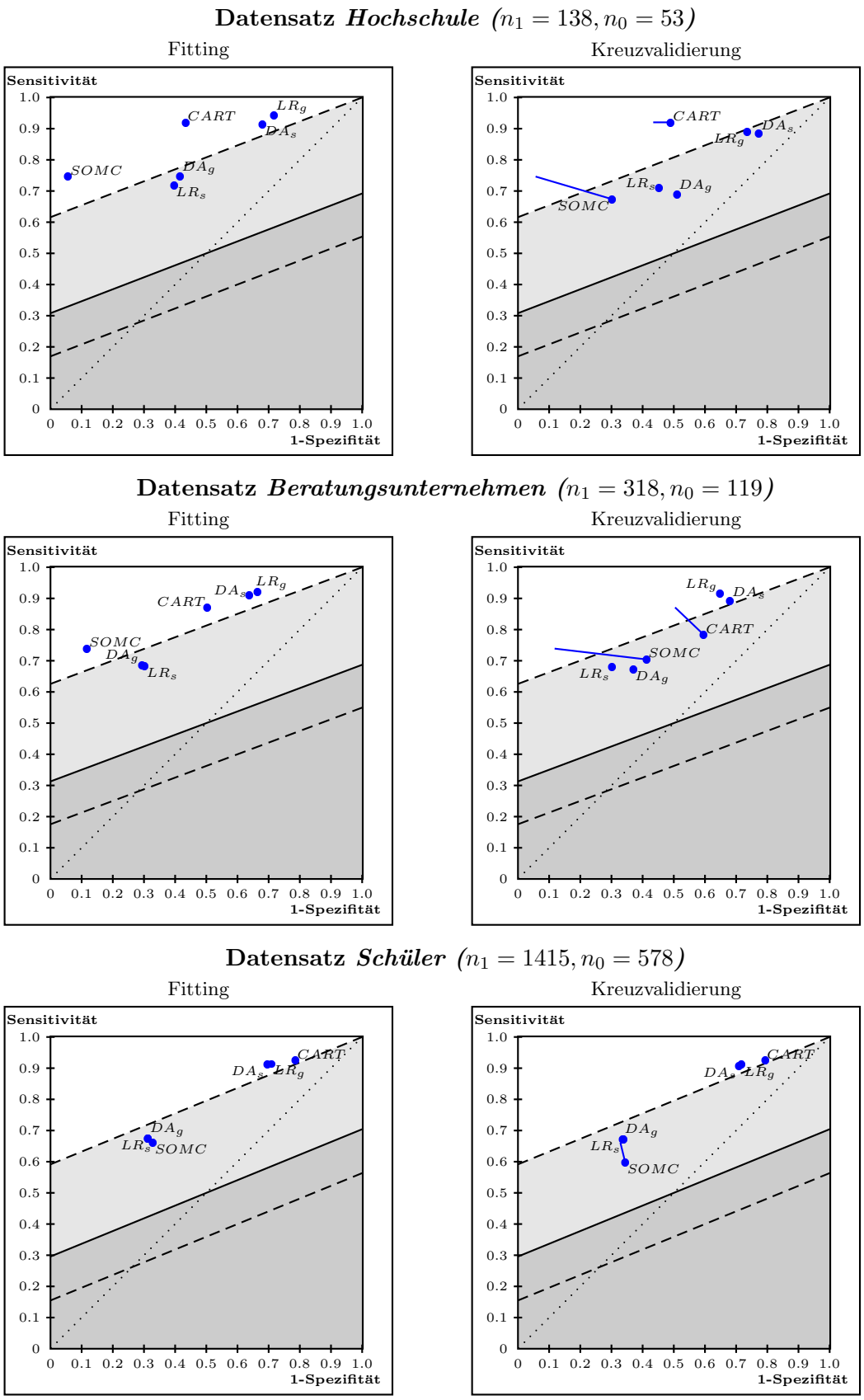


Abbildung 2.12: ROCS-Diagramm für die Ergebnisse von Montel (2005), Abkürzungen siehe Text: Für *CART* und *SOMC* ist im Kreuzvalidierungsdiagramm jeweils die Linie zwischen Fitting- und Kreuzvalidierungsergebnis eingezeichnet

The real cannot surpass the model, it is nothing but its alibi.

Baudrillard (1994)

3 Heuristiken des Künstlichen Lebens

In diesem Kapitel sollen zwei Algorithmenklassen vorgestellt werden, die im Forschungsbereich „Künstliches Leben“ entwickelt worden sind (vgl. Langton, 1989b, 1989a; Emmeche, 1994). Es sollen zunächst die Bausteine und die Funktionsweise von genetischen Algorithmen vorgestellt werden, die eine der Klassen von Heuristiken darstellen, die zur Lösung des Regelinduktionsproblems in die Software GIER$_{[1.0]}$integriert worden sind. Nach einer Beschreibung werden Komponenten und Operatoren erläutert, um dann eine spezifische Variante für das Regelinduktionsproblem vorzustellen. Weiterhin werden Lösungsmöglichkeiten für das Problem eingeschränkter Lösungsräume diskutiert.

Im Anschluss soll die Methode der Partikelschwarmoptimierung vorgestellt werden, die auch unter dem Namen der Schwarmintelligenz diskutiert wird (Kennedy & Eberhart, 2001). Der Grundalgorithmus mit seinen wesentlichen Varianten und seine Anwendung auf Optimierungsprobleme werden diskutiert. Zuletzt soll an Anwendungsbeispielen die Verbreitung der beiden Methoden demonstriert werden.

3.1 Grundstruktur von genetischen Algorithmen

Die unter dem Namen genetische Algorithmen verhandelten Objekte variieren sehr hinsichtlich Definition und Abstraktionsniveau. Zunächst einmal muss festgehalten werden, dass sich nicht unproblematisch von *dem* genetischen Algorithmus sprechen lässt, sondern dass eher von einer mehr oder weniger eng bestimmten Klasse von Algorithmen ausgegangen werden muss. Im Wesentlichen sollen hier mehrere Abstraktionsstufen unterschieden werden:

1. Ein programmiertes und lauffähiges Softwareprogramm, das keine weiteren Eingaben mehr benötigt
2. Ein programmierbares Softwareprogramm, das keine weiteren Eingaben mehr benötigt;
3. ein programmiertes und lauffähiges Softwareprogramm, das vom Anwender die Bestimmung eines oder mehrerer Parameter erfordert;
4. ein programmierbares Softwareprogramm, das vom Anwender die Bestimmung eines oder mehrerer Parameter erfordert;
5. eine programmierte Klasse von Algorithmen, deren Bestandteile und enthaltene Klassen zur Laufzeit zusammengefügt werden;
6. eine programmierbare Klasse von Algorithmen, deren Bestandteile und enthaltene Klassen zur Laufzeit zusammengefügt werden;
7. eine Sammlung programmierter Klassen von Algorithmen, deren Bestandteile und enthaltene Klassen zur Laufzeit zusammengefügt werden;

8. eine Sammlung programmierbarer Klassen von Algorithmen, deren Bestandteile und enthaltene Klassen zur Laufzeit zusammengefügt werden;

9. ein genereller Lösungsansatz zur Lösung bestimmter Problemklassen.

Des Weiteren lässt sich unterscheiden, ob die genannten Algorithmen oder Programme sich auf ein konkretes Problem, eine konkrete Problemklasse oder auf abstrakte Probleme beziehen, und im Fall des konkreten Problems die Anbindung an die Datengrundlage als Teil des Algorithmus gewertet wird. Da sich die Begriffsbestimmung zunächst als derart diffus darstellt, ist die Frage zu beantworten, welche hinreichenden oder notwendigen Eigenschaften Strukturen aufweisen müssen, um als genetischer Algorithmus klassifiziert werden können. Dabei sollte berücksichtigt werden, dass eine widerspruchsfreie Zusammenführung aller erfolgten Definitionsversuche aufgrund der Ebenenproblematik nicht gelingen kann. Als Verbindungsglied genetischer Algorithmen soll aber der Programmablauf betrachtet werden, der in Abbildung 3.1 visualisiert ist.

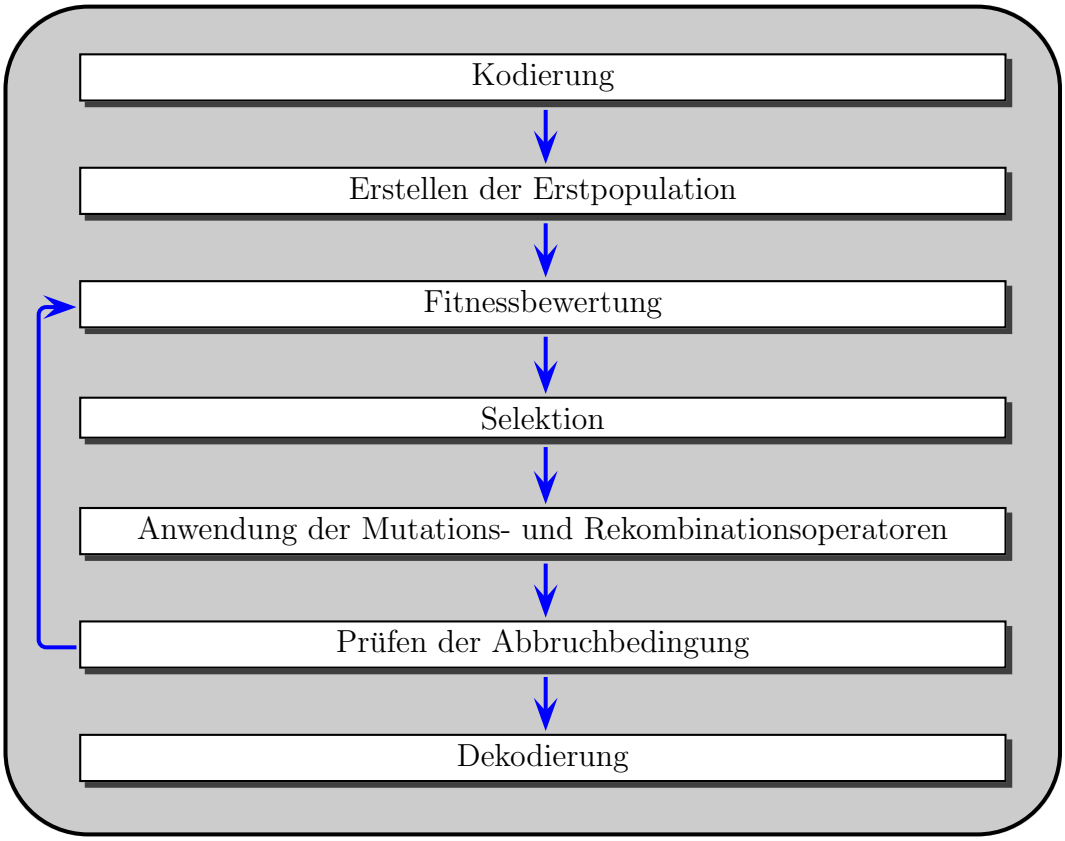


Abbildung 3.1: Schritte eines einfachen genetischen Algorithmus

Der Algorithmus wird gerahmt durch die Kodierung eines Problems in der Sprache des Algorithmus und die Dekodierung der Lösung zur Anwendung auf das Ursprungsproblem. Der Algorithmus verwendet eine künstliche Population mit einer virtuellen DNA, deren Mitglieder jeweils Eigenschaften einer Lösungsvariante in ihrer DNA gespeichert haben. Nach Erstellen einer Ausgangspopulation wird in mehreren Generationen die Qualität der individuellen Lösungen bewertet und durch einen Selektionsoperator eine Teilmenge der Individuen ausgewählt, die durch Mutations- und Rekombinationsoperatoren weiter verändert werden können. Der Algorithmus schreitet voran, bis eine Abbruchbedingung erfüllt ist. Die genannten Komponenten, die in zahlreichen Varianten auftreten, sollen im Folgenden einzeln vorgestellt werden.

3.2 Komponenten genetischer Algorithmen

Der Reihe nach soll zunächst die Repräsentation von Lösungen durch die virtuelle DNA diskutiert werden, dann die Erstellung von Populationen, der Einsatz von Fitnessfunktionen und Selektionsoperatoren,

Tabelle 3.1: Beispiel für eine binäre Kodierung: In diesem Fall werden natürliche Zahlen auf binäre Strings übertragen, die in den syntaktischen Operationen genetischer Algorithmen äquivalent sind zu einer künstlichen symbolischen DNA (kDNA)

Wert	Kodierung	kDNA	Wert	Kodierung	kDNA
0	0000	□□□□	8	1000	●□□□
1	0001	□□□●	9	1001	●□□●
2	0010	□□●□	10	1010	●□●□
3	0011	□□●●	11	1011	●□●●
4	0100	□●□□	12	1100	●●□□
5	0101	□●□●	13	1101	●●□●
6	0110	□●●□	14	1110	●●●□
7	0111	□●●●	15	1111	●●●●

Methoden der Fitnessskalierung und Abbruchkriterien. Die genetischen Operatoren der Mutation und Rekombination sollen erst im folgenden Abschnitt diskutiert werden.

3.2.1 Repräsentation

Dieser Abschnitt beschäftigt sich mit der Frage, in welcher Form potenzielle Lösungen so kodiert werden können, dass genetische Algorithmen mit ihnen operieren können. Die Form der Repräsentation bestimmt die Auswahlmöglichkeiten für genetische Operatoren und hat dementsprechend Konsequenzen für die Bestimmung des Fitnesswertes einer Lösung. Zunächst wird das Problem in allgemeiner Form repräsentiert, es folgen Angaben zu einzelnen, spezifischen Repräsentationsformen.

3.2.1.1 Lösungsraum und Enkodierung

Die Suche eines genetischen Algorithmus ist beschränkt auf die Anfangspunkte, die durch die erste Population bestimmt werden, sowie die Menge aller Punkte, die mit Hilfe der genetischen Operatoren von der Menge der Anfangspunkte aus erreicht werden können. Da sowohl die Anfangspunkte der Suche als auch die Operatoren in einer formalen Sprache kodiert sein müssen, bestimmt die Form dieser Kodierung den Suchraum in beträchtlicher Art und Weise mit. Wird ein Farbwert bspw. mit den Werten „grün“, „rot“, „blau“, „gelb“ kodiert, so befindet sich das Element „braun“ nicht im Suchraum, ein Unterschied zwischen „hellrot“ und „dunkelrot“ könnte ebenfalls nicht repräsentiert werden.

Der Code besteht aus einem Alphabet und einer Zuordnungsfunktion. Im allgemeinen stehen Kodierung und Kodiertes in einer annäherungsweise homomorphen Abbildungsbeziehung. Eine Enkodierungsfunktion γ^{-1} setzt eine Lösung in eine symbolische Form um, die ein genetischer Algorihmus syntaktisch bearbeiten kann, eine Dekodierungsfunktion γ leistet die Transformation von symbolischer Form in eine semantisch interpretierbare Form. Ist $\mathbb{M}$ ein problemspezifischer Raum, so bildet γ^{-1} diesen Raum auf den Kodierungsraum ab und γ den Kodierungsraum auf den problemspezifischen Raum. Nacheinander werden im Folgenden einige gebräuchliche Kodierungen für genetische Algorithmen mit ihren Anwendungsbedingungen sowie Vor- und Nachteilen diskutiert.

3.2.1.2 Kodierung mit binären Strings

Die Kodierung durch binäre Strings ist nicht die älteste Form der Kodierung für genetische Algorithmen, aber eine der am häufigsten verwendeten, was an der Akzeptanz des Schematheorems (Holland, 1975) zur Erklärung des Erfolgs von genetischen Algorithmen liegt. Demnach ermöglicht eine binäre Kodierung mit relativ längeren Zeichenketten als eine alternative Kodierung mehr impliziten Parallelismus (Mitchell, 1996).

Dementsprechend formuliert Goldberg (1989) die Empfehlung, dass der Anwender das kleinste Alphabet wählen sollte, das einen natürlichen Ausdruck des Problems ermöglicht. Gerade für Optimierungsprobleme mit reellen Zahlen ist der Wunsch nach Bearbeitung einer maximalen Schemaanzahl nicht

Tabelle 3.2: Beispiel für binäre Kodierung bei kontinuierlichen Zahlen: Die binären Strings und kDNA-Zeichenketten aus Tabelle 3.1 kodieren hier reelle Zahlen aus dem Intervall $[2; 2,3]$

Wert	Kodierung	kDNA	Wert	Kodierung	kDNA
2,00	0000	□□□□	2,16	1000	●□□□
2,02	0001	□□□●	2,18	1001	●□□●
2,04	0010	□□●□	2,20	1010	●□●□
2,06	0011	□□●●	2,22	1011	●□●●
2,08	0100	□●□□	2,24	1100	●●□□
2,10	0101	□●□●	2,26	1101	●●□●
2,12	0110	□●●□	2,28	1110	●●●□
2,14	0111	□●●●	2,30	1111	●●●●

empirisch begründbar (Fogel, 2000b). Mit dem Rückgang in der Unterstützung für das Schematheorem ging dementsprechend ein Rückgang in der Verwendung dieser Kodierungsform einher, da binäre Zeichenketten aufgrund ihrer Unhandlichkeit und Unnatürlichkeit auch Nachteile bieten.

Ein binärer String ist eine Zeichenkette beliebiger Länge unter Verwendung des binären Alphabets $(0, 1)$. Ein String entspricht damit so vielen Bits wie er Zeichen hat und ein m-bit-String ist im Raum $\mathbb{B}^m = \{0, 1\}^m$ positioniert (Deb, 2000a). Dementsprechend haben die Kodierungs- und Dekodierungsfunktion die Form $\gamma^{-1} : \mathbb{B}^m \to \mathbb{M}$ und $\gamma : \mathbb{M} \to \mathbb{B}^m$. Die meisten problemspezifischen Räume bestehen aus einer Menge von Parameterwerten oder lassen sich durch eine solche Menge beschreiben. In diesem Fall würde bei der binären Kodierung jeder dieser Werte in einen String angemessener Länge transformiert werden. Die Länge bestimmt sich dabei über den Wertebereich und den gewünschten Präzisionsgrad. Ein String der Länge m kann 2^m unterschiedliche Werte kodieren, also zum Beispiel den Bereich der natürlichen Zahlen in den Intervallen $[1; 2^m]$ oder $[1 + k; 2^m + k]$. Tabelle 3.1 stellt ursprüngliche Zahlen und Kodierungen im Intervall $[0; 15]$ einander gegenüber. Für dieses Intervall sind Strings der Länge 4 notwendig.

Es soll noch einmal darauf hingewiesen werden, dass Strings von den Algorithmen syntaktisch bearbeitet werden, und somit eine Darstellung durch ein binäres Symbolalphabet in der angegebenen kDNA-Form der binären Kodierung äquivalent ist, solange Kodierungs- und Dekodierungsfunktion aufeinander abgestimmt sind. Benennt man die binäre Zeichenkette für einen Parameter i als b_i mit der Länge m_i, dann besteht b_i aus den Zeichen $(b_{i1}, \ldots, b_{im_i})$. Die Gesamtlänge der Zeichenketten für die Kodierung setzt sich bei k kodierten Paramtern mit den Zeichenkettenlängen m_1 bis m_k als einfache Summe dieser Ketten zusammen:

$$m = \sum_{i=1}^{k} m_i. \tag{3.1}$$

Wird die Kodierung reeller Zahlen gewünscht, so ist die Länge der gewählten Kodierung direkt mit der möglichen Präzision der Repräsentation verbunden. Das Intervall zwischen kleinstem und höchstem Wert kann in maximal $2^m - 1$ Schritte zerlegt werden. Tabelle 3.2 zeigt, wie die gleichen Zeichenketten die in Tabelle 3.1 natürliche Zahlen aus dem Intervall $[0; 15]$ kodierten, auch reelle Zahlen aus dem Bereich $[2; 2,3]$ kodieren können, wobei die Präzision lediglich $0,02$ beträgt. Der Algorithmus führt eine Gittersuche durch, die bei zu geringer Auflösung das Optimum übersehen kann (Bäck, 2000). Es ist deshalb von großer Bedeutung, eine Präzision zu wählen, die für die Lösungen fein genug ist, aber nicht übermäßig präzise ist, da längere Zeichenketten zu längeren Such- und Laufzeiten und höherem Speicherbedarf führen. In diesem Fall ist also für einen Parameter i die Dekodierungsfunktion als $\gamma_i : \mathbb{B}^{m_i} \to [v_i, w_i]$ gegeben. Dabei gilt für die Rekonstruktion des kodierten Parameterwerts x_i aus der Teilzeichenkette b_i:

$$x_i = v_i + \frac{w_i - v_i}{2^{m_i} - 1} \left(\sum_{j=0}^{m_1 - 1} b_{i(m_i - j)} 2^j \right) \tag{3.2}$$

Tabelle 3.3: Beispiel für die Gray-Kodierung bei sechsstelligen binären Strings

n	binäre Kodierung	Gray Code
1	000001	000001
2	000010	000011
3	000011	000010
4	000100	000110
5	000101	000111
6	000110	000101
7	000111	000100
8	001000	001100
25	011001	010101
50	110010	101011
51	110011	101010
52	110100	101110

Die Abbildung ist dementsprechend linear mit konstanten Differenzen zwischen den der Größe nach geordneten möglichen Zahlenwerten. Damit können alle Parameterwerte aus den Teilzeichenketten rekonstruiert werden, um so den Lösungsvektor zu erhalten, der der Zeichenkette entspricht. Deb (2000a) verweist auf eine Variante der binären Kodierung, bei der die erste Position in der Zeichenkette das Vorzeichen kodiert. Sind die Bereiche im negativen und positiven Bereich aber symmetrisch, so besteht kein Unterschied zur üblichen Kodierung.

3.2.1.3 Gray-Kodierung

Eine Art der binären Kodierung, die sicher stellt, dass bei der Kodierung von natürlichen Nachbarzahlen sich die Strings genau in einem Bit unterscheiden und damit eine Hamming-Distanz von 1 haben, ist die so genannte *Gray-Kodierung*. Die Hamming-Distanz zwischen zwei Strings wird berechnet als die Anzahl der zwischen zwei Strings gleicher Länge verschiedenen Bits durch Verwendung des logischen XOR-Operators mit den Bits an gleicher Position und Addition der Ergebnisse. Zwei identische Strings der Länge k haben eine Hamming-Distanz von 0, zwei maximal verschiedene Strings eine Hamming-Distanz von k (Gelbard & Spiegler, 2000; Erlich, Gelbard & Spiegler, 2002).

Für die Transformation von binären Strings $(b_1, \ldots, b_k)$ der oben beschriebenen Form in so genannte Gray Codes $(g_1, \ldots, g_k)$ erfolgt über eine Abbildung $\gamma^{-}1 : \mathbb{B}^k \to \mathbb{B}^k$ der Form (Deb, 2000a):

$$g_i = \begin{cases} b_i & i = 1 \\ b_{i-1} \oplus b_i & i \neq 1. \end{cases} \tag{3.3}$$

Dabei steht der Operator $\oplus$ für die Addition modulo 2. Das erste Bit wird also übernommen, ab dem zweiten Bit wird das Bit an der betrachteten Stelle im ursprünglichen String zu dem Bit an der Vorgängerstelle addiert, um das Ergebnis 0 bei einer Summe von 0 oder 2 oder das Ergebnis 1 bei einer Summe von 1 zu erhalten.

Tabelle 3.3 zeigt einige Kodierungsbeispiele bei sechsstelligen Strings. Als Vorteil der Kodierung ergibt sich die gesicherte Existenz von Mutationen eines Strings mit geringen Auswirkungen auf die kodierten Variablen, ein Ansatz, der von Hollstien propagiert wurde (Goldberg, 1989).

3.2.1.4 Direkte Kodierung mit reellen Zahlen

Vielleicht am nahe liegendsten ist die Repräsentation von Werten durch Zahlen, die zu den kodierten Entitäten in einer Beziehung stehen, wie Messwerte zu Messobjekten. Dementsprechend besteht die Kodierung bei k Parameterwerten aus einem Vektor mit k reellzahligen Komponenten. Die Verwendung reeller Zahlen bedingt den Einsatz spezieller Operatoren, da eine rein syntaktische Verarbeitung keine günstigen Ergebnisse erbringt. An dieser Stelle zeigt sich das Auseinanderbrechen einer Fraktion von Anwendern, die dem biologischen Vorbild getreu die semantische Blindheit der Operatoren schätzt und einer stärker pragmatisch orientierten Strömung, die der Lösung von Optimierungsproblemen die höhere Priorität gibt.

3.2.2 Erstellung der Anfangspopulation

In der Theorie ist der Erfolg eines genetischen Algorithmus unabhängig von seinem Anfangszustand, jedenfalls was den langfristigen Erfolg angeht. Allerdings kann diese Frist in praktischen Anwendungen sehr lang werden, wenn eine ungünstige Anfangskonstellation gewählt wurde. Das größte Problem stellen dabei lokale Minima dar, in deren Nähe Anfangsindividuen nur mit geringer Wahrscheinlichkeit Nachwuchs erzeugen, der diesen Bereichen entkommt. Aus diesem Grund ist die Erstellung einer Anfangspopulation keine triviale Aufgabe, und gerade Divergenzen in der Anfangskonstellation führen zu der Praxis, sich selten mit einem einzigen Durchlauf eines genetischen Algorithmus zu begnügen. Die Aufgabenstellung ist nicht für genetische Algorithmen exklusiv. Analoge Probleme stellen sich etwa ein, wenn die anfänglichen Clusterzentren für clusteranalytische Verfahren gesucht werden müssen (vgl. Everitt, 2001). In diesem Abschnitt sollen einige Methoden vorgestellt werden, die zur Bearbeitung der Aufgabe in Anwendung sind.

3.2.2.1 Randomisierte Erzeugung

Um der Gefahr eines unerwünschten Bias zu entgehen, kann die erste Population ausschließlich zufallsbestimmt erstellt werden. Dabei ist allerdings exakt zu definieren, auf welche stochastische Verteilung sich diese Formulierung bezieht, aus der die erste Population eine Ziehung darstellen soll.

Diese Frage lässt sich nur dann unabhängig von der Repräsentation beantworten, wenn sich die Verteilung allein auf Genompositionen bezieht, indem etwa jede Position bei jedem Individuum mit den Wahrscheinlichkeiten $p_0 = 0,5$ und $p_1 = 0,5$ zufällig initialisiert werden. Bei kontinuierlichen Daten entspräche dies einer Ziehung aus einer über dem Intervall der möglichen Werte gleichverteilten Zufallsverteilung. Kodieren die Genomstellen hingegen Eigenschaften, für deren Verteilung bestimmte Annahmen getroffen werden können, und sollen die Individuen dementsprechend aus einem realistischen Wertebereich gezogen werden können, ist auch die Bevorzugung anderer Wahrscheinlichkeitsverteilungen denkbar, etwa der Normalverteilung für kontinuierliche Merkmale, Binomial- oder Poisson-Verteilungen für Häufigkeitsdaten, angenäherter Normalverteilungen für kategoriale Variablen etc. Die Forderung nach dieser Form von Anpassung ist allerdings nur in solchen Fällen von Vorteil, wenn angenommen werden kann, dass sich die Suche im Lösungsraum auf Werte konzentrieren sollte, die sich aus den verwendeten Verteilungen wahrscheinlich ergeben. Jede Konzentration auf einen Wertebereich geht mit einer Vernachlässigung der übrigen Bereiche einher, die von einem genetischen Algorithmus durch den Mutationsoperator zwar theoretisch langfristig ebenfalls exploriert werden, allerdings mit einem deutlich höheren Erwartungswert für die zu erwartende Zeitdauer.

3.2.2.2 Systematische Auswahl der ersten Individuen

Soll die frühzeitige Konvergenz auf einen kleinen Suchbereich verhindert werden, kann die Methode der systematischen Platzierung der ersten Individuen im Lösungsraum erfolgreich sein. Dazu ist es förderlich, wenn sichergestellt werden kann, dass für jede kodierte Eigenschaft entweder jeder mögliche Wert in der Population enthalten ist oder zumindest jeder Wertebereich bei einer gewählten Auflösungsstufe in der Population repräsentiert wird. Übersteigt die Anzahl der Kombinationen der Kategorien für Einzelwerte die Anzahl der handhabbaren Populationsmitglieder, kann eine Teilmenge der möglichen Individuen ausgewählt werden, etwa durch Methoden analog zu den Verfahren zur Erstellung unvollständiger Versuchspläne.

Vorteil dieser Verfahren ist eine große genetische Populationsvarianz und die Verfügbarkeit einer Vielzahl von konkurrierenden Lösungen gleich zu Beginn des Laufs. Nachteil ist die Irrelevanz dieser Eigenschaft bei Problemen mit hoher Interaktion zwischen den Einzelwerten, da es hier auf die spezifische Kombination von Werten ankommt und sich Lösungen, die Teilkomponenten enthalten, im ungünstigen Fall nicht oder sogar negativ von Lösungen ohne diese Teilkomponenten abheben können. Gleichzeitig bedeutet die Auswahl einer geeigneten Auflösung bei kontinuierlichen Variablen eine Wahl zwischen einem möglicherweise zu groben Raster und einem hohen Zeitbedarf durch eine möglicherweise übergroße Population.

3.2.2.3 Empiriegestützte Auswahl der ersten Individuen

Die Bestimmung einer problemadäquaten Wahrscheinlichkeitsverteilung für die Erstpopulation kann auch ein Beispiel für eine empiriegestützte Auswahl sein, wenn die Verteilungsannahmen aufgrund empirischer Beobachtungen getroffen wurden. Alternativ zu dieser Form gibt es die Praxis, Beispielexemplare für Variablenkombinationen im Sinne von beobachtbaren Problemlösungen oder Erfahrungswerten heranzuziehen. Die Form dieser Methode ist problemabhängig. Bei bestimmten Anwendungsgebieten gibt es eine bestehende Praxis oder die Verfügbarkeit von Experteneinschätzungen, bei anderen Fragestellungen, für die der Begriff 'Expertise' keinen Sinn ergibt, kann auch eine empirische Beobachtung an einer Zufallsstichprobe von relevanten Untersuchungseinheiten ein gewünschter Indikator sein.

3.2.2.4 Erstellung der ersten Population durch einen Hilfsalgorithmus

Im Prinzip kann jeder Algorithmus, durch den eine Population von möglichen Lösungen erstellt werden kann, einem genetischen Algorithmus vorangestellt werden. Diese Vorgehensweise führt zu einer Erhöhung der Abstraktionsstufe des Algorithmus, wenn dieser Algorithmus als genetischer Algorithmus bestimmt wird, dessen erste Population zur Verhinderung eines infiniten Regresses durch eine simplere Methode erstellt wurde. Gleichzeitig muss dieser Algorithmus sich strukturell vom Hauptalgorithmus unterscheiden, sonst wäre das Ergebnis des ersten Algorithmus auch das Ergebnis des zweiten Algorithmus. Normalerweise gehen diese Unterschiede auch über das Abbruchkriterium hinaus, denkbar wäre eine schnellere und oberflächliche Variante, die eine Art Vorarbeit leistet. Dies könnte auch erreicht werden, durch eine Reihe von Läufen dieses Algorithmus mit kleiner Populationszahl.

Erlaubt man die Möglichkeit der Hybridisierung des Algorithmus, kann auch jeder andere Suchalgorithmus mit den beschriebenen Eigenschaften vorgeschaltet werden, etwa ein einfacher *Hill Climber*. Vorteil dieser Methode ist eine mögliche Verkürzung der Laufzeit des Algorithmus, da nahe liegende Schritte zur Findung einer Lösung von anderen Algorithmen durchschnittlich schneller geleistet werden können als von genetischen Algorithmen. Einen gravierenden Nachteil hingegen stellt die Möglichkeit für einen falschen Zuschnitt des Suchraums dar, da bspw. *Hill Climbing* Algorithmen sehr anfällig für lokale Optima sind, und eine Lösungspopulation, die um ein lokales Optimum gruppiert ist, diesen Bereich nur bei unter Umständen sehr langer Laufzeit verlassen kann.

Zusammenfassend lässt sich sagen, dass keine Methode den übrigen ohne genauere Kenntnisse über den Suchraum überlegen ist. Häufig wird deshalb die Zufallsgenerierung als Initialisierungsmethode gewählt, da sie sehr zeitsparend und einfach zu implementieren ist.

3.2.3 Fitnessfunktion

3.2.3.1 Objective functions und Fitnessfunktionen

Die Zuordnungsfunktion von Qualitätsmaßen zu Eigenschaften nennen Biologen nach Goldberg (1989) *Fitnessfunktion*[1], er selbst bezeichnet sie als *Objective Function* oder *Bewertungsfunktion*. Im Folgenden soll zwischen Bewertungsfunktion und Fitnessfunktion differenziert werden, in dem Sinne, dass die Bewertungsfunktion als primäre Zuordnung von Qualitätsangaben zu Individuen, also als

$$o : A_x \rightarrow \mathbb{R} \tag{3.4}$$

bestimmt werden kann. A_x stellt dabei den Raum der Objektvariable dar, $a_i \in A_x$ repräsentiert dabei die Parameterkombination einer Beispiellösung, also ein mögliches Individuum der Population. Im Folgenden werden die Bezeichnungen „Individuum der Population“ und „Parameterkombination einer Beispiellösung“ bzw. „Struktur einer Beispiellösung“ in Bezug auf Funktionen synonym gebraucht, so dass gesagt werden kann, dass die Bewertungsfunktion Populationsmitgliedern einen Qualitätswert zuordnet.

Im einfachsten Fall entspricht die Fitnessfunktion der Bewertungsfunktion, allerdings kann die Bewertungsfunktion unter Umständen ungünstige Eigenschaften aufweisen. Hier soll unter Fitnessfunktion eine erste Transformation der Funktionswerte der Bewertungsfunktion verstanden werden, die die Erfüllung basaler Anforderungen an die Fitnessfunktion sicher stellt. In vielen Applikationen ist die Nichtnegativität der Fitnesswerte eine solche Anforderung, also ergibt sich die Fitnessfunktion in diesem Fall als:

[1] Der Begriff „Fitness“ sollte allerdings vorsichtiger gebraucht werden (vgl. Ariew & Lewontin, 2004)

$$f : A_x \rightarrow \mathbb{R}_+. \tag{3.5}$$

Die einfachste Methode der Transformation stellt dabei die Subtraktion des kleinsten aufgetretenen Wert der Bewertungsfunktion dar, sofern er negativ ist:

$$f(o(a_i(x)) = \begin{cases} o(a_i(x)) & , \min_{a_k \in A_x} o(a_k(x)) \geq 0 \\ o(a_i(x)) - \min_{a_k \in A_x} o(a_k(x)) & , \min_{a_k \in A_x} o(a_k(x)) < 0. \end{cases} \tag{3.6}$$

3.2.3.2 Fitnessfunktionen für Klassifikationsprobleme

Für Klassifikationsprobleme stellt sich die Frage der Überführung der Klassifikationsleistung in einen einzigen Zahlenwert. Dieses Problem ist bereits in Abschnitt 2.3.3.2 diskutiert worden. Jede der dort gefundenen Lösungen für die Beurteilung der Güte der Klassifikationsleitung stellt eine mögliche Zielfunktion für einen genetischen Algorithmus dar. Die Auswahl ist demenstprechend abhängig von der gewünschten Optimierungsrichtung.

3.2.4 Selektionsoperatoren

Im Selektionsschritt des genetischen Algorithmus kommt es positiv formuliert zu einer Auswahl von Individuen der aktuellen Generation für die nächste Generation, bzw. - negativ formuliert - zu einer Auswahl von Individuen, die in weiteren Schritten nicht mehr berücksichtigt werden. Kriterium für die Auswahl der Individuen ist dabei ihr Fitnesswert. Bei diesem Schritt werden keine neuen Lösungen generiert. Statt dessen wird die Qualität von Lösungen bewertet, und es werden Konsequenzen aus dieser Bewertung gezogen. Der Selektionsoperator bildet damit Aspekte der Prozesse Selektion und Fortpflanzung in natürlichen Systemen ab (Deb, 2000b).

Zur Realisierung dieses Schrittes sind verschiedene Modelle denkbar, und diese sind in verschiedenen Anwendungen bereits umgesetzt worden. Nacheinander sollen im Folgenden die gängigen Formen von Selektionsoperatoren dargestellt werden. Die ersten drei Mechanismen sind stochastische Prozesse, bei denen die Anzahl der Nachkommen eines Individuums der Population nur über die Gesamtpopulationsgröße festgelegt ist, die beiden folgenden sind deterministische Prozesse, bei denen jedes Individuum maximal einen Nachkommen in der folgenden Generation haben kann. Zur Veranschaulichung soll jeweils die Beispielpopulation herangezogen werden, die in Abbildung 3.2 dargestellt ist. Die Population besteht

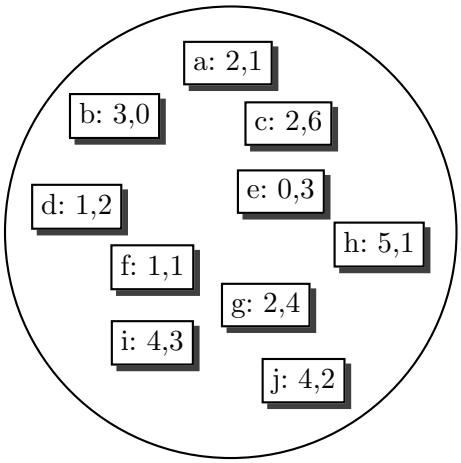


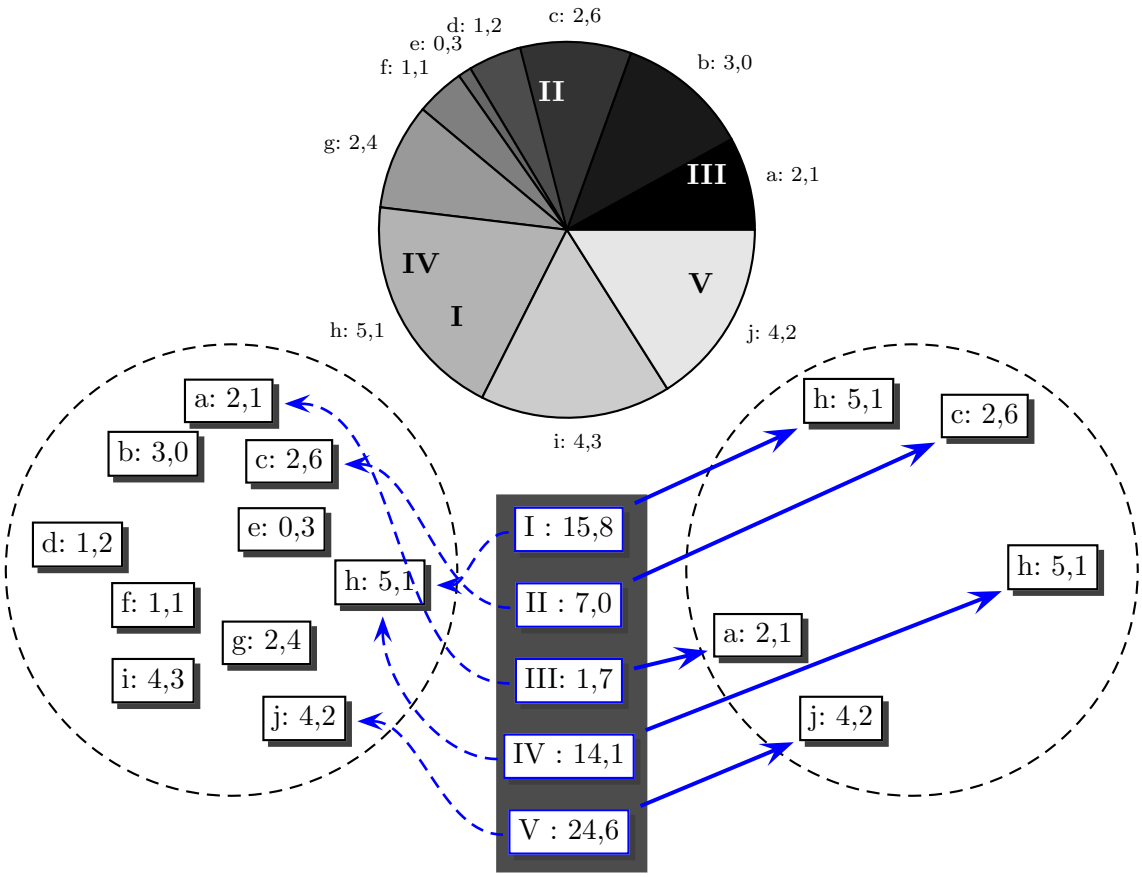
Abbildung 3.2: Beispielpopulation mit angegebenen Fitnesswerten

aus den zehn Individuen a bis j, die jeweils durch einen Buchstaben und ihren Fitnesswert repräsentiert sind.

3.2.4.1 Roulette-Wheel Selection

Das Prinzip der *Roulette-Wheel Selection* oder auch *Proportional Selection* wurde von Holland (1975) in das Feld der genetischen Algorithmen eingebracht und verwendet wiederum eine innere Metapher. Ein

imaginäres Rouletterad wird zusammengesetzt aus Taschen, die mit den einzelnen Individuen korrespondieren. Dabei bestimmt sich die Größe des Kreissektors, den eine Tasche ausfüllt, aus dem Fitnesswert des Individuums dergestalt, dass die Größe der Tasche proportional bemessen wird zum Anteil des Fitnesswertes des Individuums an der Summe aller Fitnesswerte in der Population. Für eine Population von n Individuen werden n imaginäre Kugeln in das virtuell gedrehte Rouletterad eingeworfen. Dabei wird die Endposition einer Kugel als eine zufällig ermittelte Position auf dem Kreisumfang imaginiert. Der Einfachheit halber können die Einheiten so gewählt werden, dass die Länge des Kreisumfangs der Summe der Fitnesswerte der Individuen entspricht. Zudem wird vom Roulettespiel soweit abstrahiert, dass die Struktur und Breite der Taschen keinen Einfluss auf die Zielposition der Kugel haben sollen. Damit wird der spielerische Charakter der Benennung dieser Methode noch einmal im doppelten Sinne unterstrichen. Das mit den Zieltaschen der Kugeln korrespondierende Individuum erhält dabei jeweils einen Nachkommen in der nachfolgenden Generation, wobei Einzelindividuen mehr als einen Nachkommen haben können und sogar müssen, wenn das Verfahren nicht zu einer reinen Neuanordnung der Ausgangspopulation führen soll, siehe dazu Abbildung 3.3.



$$\sum_{i \in pop} Fitness_i = 26,2$$

Abbildung 3.3: Auswahl der ersten fünf Nachkommen durch Roulette-Wheel Selection: Individuum h hat bereits 2 Nachkommen für den nächsten Schritt, da die Zufallszahlen 14,1 und 15,8 auf das h zugeordnete Kreissegment verweisen.

Die Wahrscheinlichkeit, bei jeder simulierten Raddrehung für ein Individuum a einen Nachkommen zu

erhalten, ergibt sich dementsprechend als:

$$p_a = \frac{fit_a}{\sum_{i=1}^{n} fit_i} \tag{3.7}$$

Voraussetzung für die Anwendung der Methode sind damit Fitnesswerte größer oder gleich Null (wären alle Fitnesswerte kleiner oder gleich Null, würde dies zu einer kontraintuitiven Auswahl führen). Für zwei Individuen a und b gilt dementsprechend:

$$\frac{p_a}{p_b} = \frac{fit_a}{fit_b} \tag{3.8}$$

Für ein Individuum a ergibt sich damit der Erwartungswert

$$E(n_a) = n \cdot p_a, \quad bzw. \tag{3.9}$$

$$E(n_a) = n \cdot \frac{fit_a}{\sum_{i=1}^{n} fit_i} \tag{3.10}$$

für die Anzahl seiner Nachkommen, die für die Folgegeneration berücksichtigt werden. Ist der Fitnesswert eines Individuums doppelt so groß wie der eines anderen Individuums, ist der Erwartungswert für die Anzahl der Nachkommen ebenfalls doppelt so groß. Allerdings ist es wichtig, hier festzuhalten, dass die Bezeichnung „Anzahl der Nachkommen“ im Sinne von Replikationen an dieser Stelle voreilig ist, da in den folgenden Schritten genetische Operatoren den Code der Replikationen verändern können.

Einen alternativen Algorithmus, das *Stochastic Universal Sampling* (SUS), bei dem die Population genau einmal durchlaufen wird und jeweils die Anzahl der Kopien festgelegt wird, beschreibt Grefenstette (2000a).

3.2.4.2 Tournament Selection

Die Methode der *Tournament Selection* ist ebenfalls eine stochastische Methode, deren Auswahlprozess sich aber in einer Anzahl von Schrittpaaren vollzieht, bei denen der zweite Schritt jeweils einer deterministischen Regel folgt (s. Blickle, 2000). Zudem wird nicht der Bereich des Glücksspiels, sondern der Bereich des Turnierwettkampfs für die namensgebende Metapher herangezogen.

Dieser (Teil-)Algorithmus erfordert die Bestimmung zweier Parameter, m und k. Dabei bestimmt m die Turniergröße, k bestimmt die Anzahl der Sieger pro Turnier. Üblicherweise wird k auf 1 gesetzt. In einem ersten Schritt werden zufällig m Individuen der Ausgangspopulation ausgewählt. Dabei ist die Wahrscheinlichkeit der Auswahl für jedes Individuum unabhängig von seiner Fitness, die Wahlen sind damit gleichwahrscheinlich. Übertragen auf die Roulette-Metapher der *Roulette-Wheel Selection* entspricht dies der Auswahl einer Population der Größe m mit gleichmäßig aufgeteilten Roulette-Taschen. Ebenso wie bei dieser Methode können auch in diesem Beispiel mehrere Instanzen eines Individuums im Turnier berücksichtigt werden. In einem zweiten Schritt werden nun die Individuen ihrer Fitness nach geordnet, und die k Individuen mit der höchsten Fitness werden in die Folgegeneration übernommen (siehe Abb. 3.4).

Es müssen damit $\lceil \frac{n}{k} \rceil$ Turniere durchgeführt werden, um die Folgegeneration vollständig auszuwählen. Die Turniere werden unabhängig voneinander ausgetragen, damit stehen für jedes Turnier alle Individuen zur Verfügung.

3.2.4.3 Probabilistic Tournament Selection

Die *Probabilistic Tournament Selection* entspricht weitgehend der *Tournament Selection*, allerdings ist bei dieser Variante der zweite Schritt ebenfalls mit einer stochastischen Komponente versehen. Wie zuvor wird eine zuvor bestimmte Anzahl von Individuen zufällig (bei gleicher Wahrscheinlichkeit für jedes Individuum bei jeder Ziehung) aus der Population ausgewählt. Im zweiten Schritt werden k Individuen aus der Menge der ausgewählten Individuen in die Folgegeneration aufgenommen. Dazu muss vor dem ersten Turnier eine Wahrscheinlichkeitsverteilung festgelegt werden, die jedem Rangplatz in Bezug auf die Fitnesswerte eine Wahrscheinlichkeit zuordnet, mit der dieser Rangplatz berücksichtigt werden soll. Nahe liegend ist eine abnehmende Wahrscheinlichkeit für die Rangplätze. Damit wird also eine zweite virtuelle Roulettescheibe

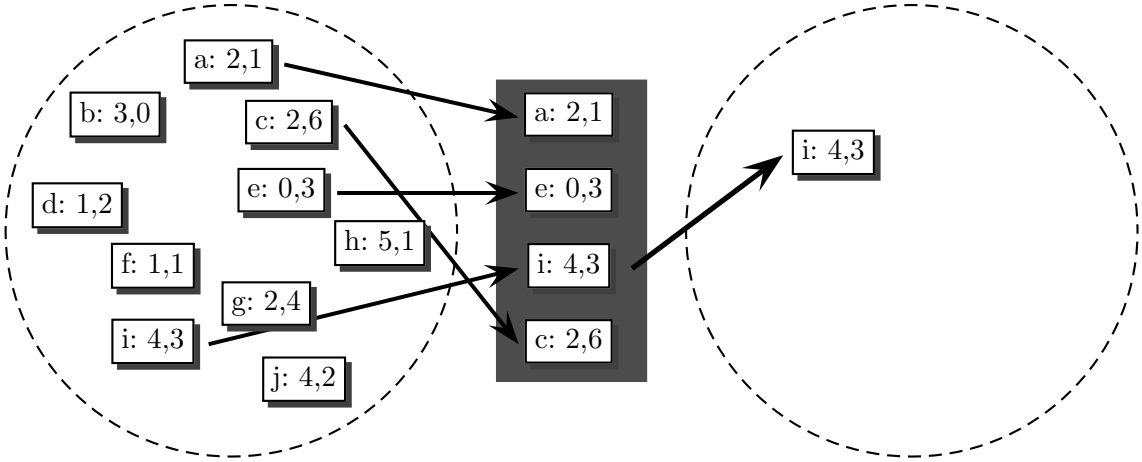


Abbildung 3.4: Tournament Selection mit einer Turniergröße von 4 und einer Siegeranzahl von 1: Im ersten Turnier wurde der Sieger ermittelt, neun weitere Turniere folgen

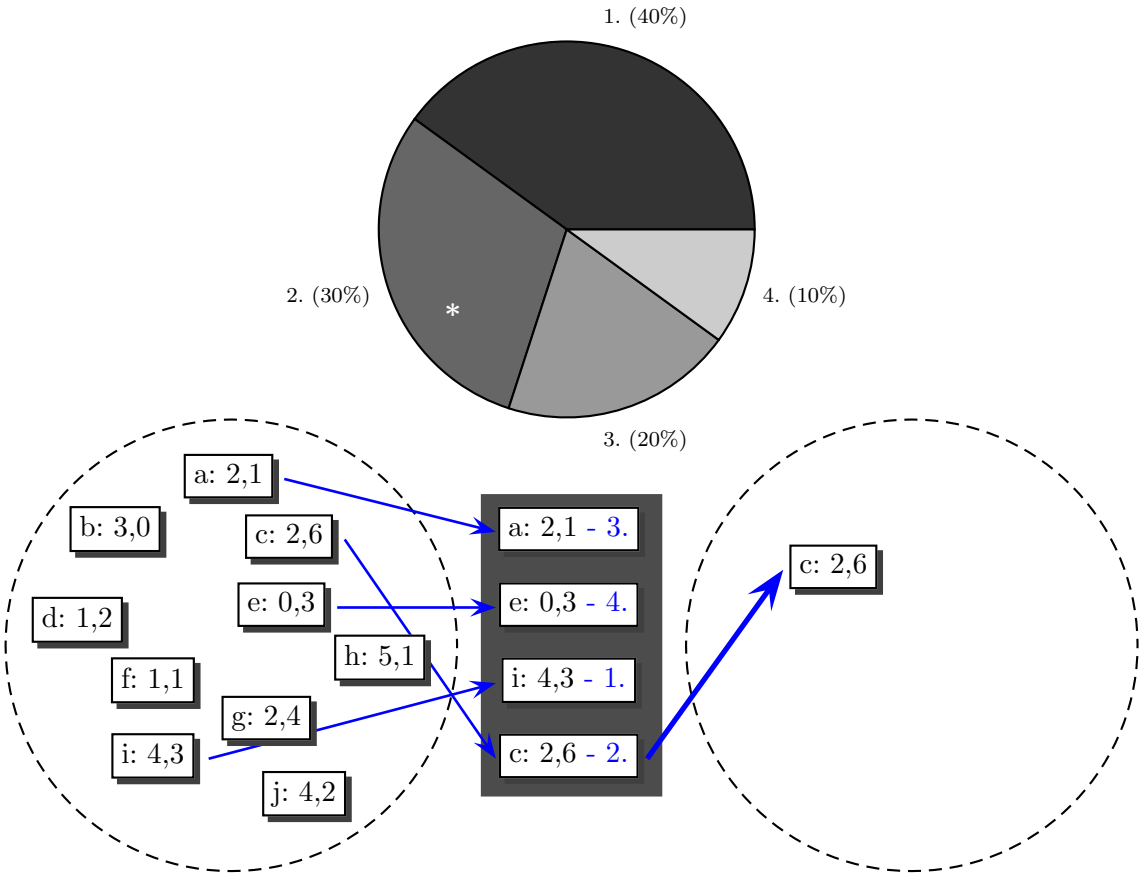


Abbildung 3.5: Probabilistic Tournament Selection: Im Beispiel wurde im ersten Turnier mit vier Teilnehmern der zweite Platz ausgewählt. Die übrigen neun Plätze werden in neuen Turnieren ermittelt

ins Spiel gebracht, deren k Taschengrößen durch die vorgegebene Wahrscheinlichkeitsverteilung bestimmt sind (s. dazu Abb. 3.5).

Dabei ist es erneut theoretisch möglich, dass ein Individuum alle Nachkommen stellt, die aus einem Turnier hervorgehen, indem entweder k Kopien dieses Individuums in jedes Turnier aufgenommen werden, oder indem in n Ziehungen jeweils immer einer der Rangplätze gezogen wird, die einer der Kopien dieses Individuums zugeordnet sind.

3.2.4.4 (μ, λ)- Selection

Diese Form der Selektion wurde, wie auch die folgende Form, im Bereich der Evolutionsstrategien entwickelt, sie lässt sich aber modifiziert auch in genetischen Algorithmen verwenden. Wie bereits angekündigt, gehören diese beiden Methoden zu den deterministischen Verfahren. Das Selektionsprinzip ist rangbasiert und arbeitet mit zwei unterschiedlichen Populationsgrößen. Dabei bestimmt μ die Größe der Ausgangspopulation ($\mu > 0$), λ bestimmt die Größe einer für diese Methode spezifischen Zwischenpopulation ($\lambda \geq \mu$). Zunächst werden λ Nachfahren aus der bestehenden Generation generiert. Im einfachsten Fall ($\mu = \lambda$) würde die gesamte Population einfach übernommen. Für den Fall $\lambda = k \cdot \mu$ mit $k \in \mathbb{N}$ könnten k Kopien jedes Mitglieds der Ausgangspopulation generiert werden. Bei anderen Werten von λ könnte eines der oben genannten Selektionsverfahren zum Einsatz kommen, um die Zwischenpopulation zu generieren, solange Fitnesswerte für die Population vorliegen. Ebenso wäre eine Zufallsauswahl geeignet, um die restlichen Plätze zu füllen. Dieser erste Schritt schafft also eine zweite Population, deren Mitglieder mit Mitgliedern der ersten Population zunächst identisch sind. Dann werden die Operatoren der folgenden Schritte angewendet und neue Fitnesswerte berechnet. Diese Fitnesswerte werden in eine Rangreihe gebracht, und nur die ersten μ Werte werden ausgewählt, um erneut die Basis zu bilden für die Erstellung von λ Individuen (s.a. Abb. 3.6).

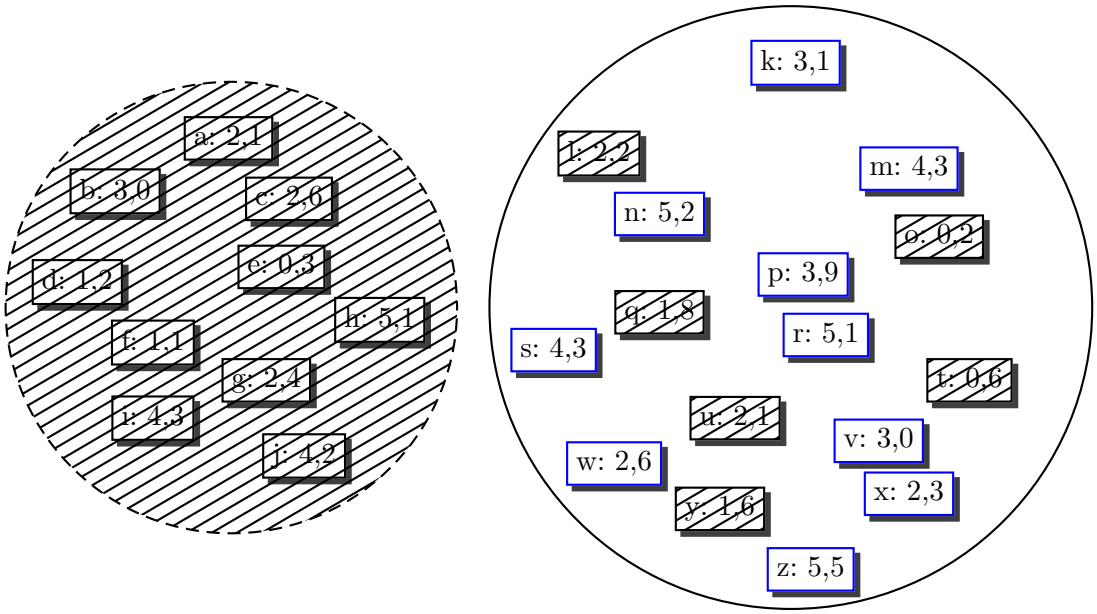


Abbildung 3.6: (μ, λ)-Selektion mit $\mu = 10$ und $\lambda = 16$: Alle schattierten Individuen fallen in diesem Schritt heraus

Beginnend mit der zweiten Generation besteht der Algorithmus deshalb immer aus zwei Schritten:

1. Die Zwischenpopulation mit λ Individuen wird deterministisch entsprechend der Rangfolge der Werte auf μ Individuen reduziert.

2. Aus den μ Individuen werden durch Kopieren oder durch Selektionsprozesse erneut λ Individuen erstellt.

In der ersten Generation entfällt der erste Schritt. Die Sprache von zwei Populationen ist bei dieser Methode ein wenig irreführend, da die kleinere der beiden Populationen streng genommen nur zwischen den beiden Schritten existiert und ihre Existenz damit als doppelt virtuell betrachtet werden kann.

3.2.4.5 $(\mu + \lambda)$- Selection

Die Sprache von zwei Populationen ist bei dieser Variante des Selektionsverfahrens nahe liegender. Erneut besteht die Basispopulation aus $\mu > 0$ Individuen. Es wird erneut wie zuvor eine Zwischenpopulation aus λ (hier: $\lambda > 0$) Individuen erstellt. Diese durchlaufen den Zyklus der Operatoranwendungen, und ihre Fitness wird anschließend aktualisiert. An dieser Stelle zeigt sich die Differenz zwischen den beiden Selektionsvarianten: Für die Reduktion der Anzahl der Individuen werden nicht nur die Mitglieder der Zwischenpopulation berücksichtigt, sondern die Mitglieder beider Populationen, die in eine gemeinsame Rangfolge gebracht werden. Die Individuen mit den μ höchsten Fitnesswerten in Bezug auf beide Populationen bilden die Folgegeneration, aus der eine neue Zwischengeneration geformt wird (s. Abb. 3.7).

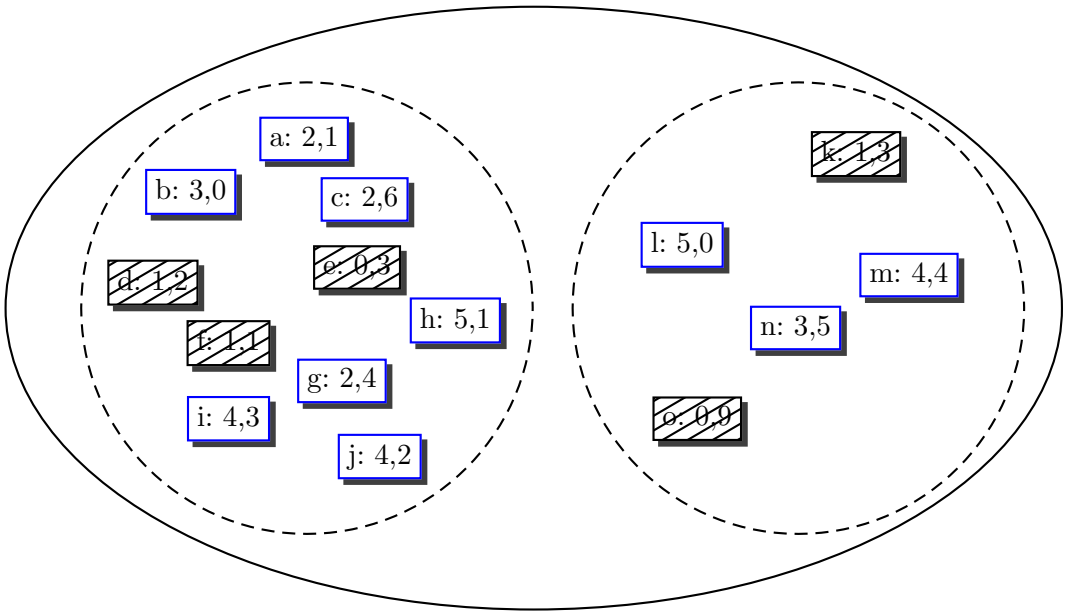


Abbildung 3.7: $(\mu + \lambda)$-Selektion mit $\mu = 10$ und $\lambda = 5$: Die schattierten Individuen fallen in diesem Schritt heraus

Die beiden Schritte werden also modifiziert zu:

1. Die beiden Populationen mit $\mu + \lambda$ Individuen werden deterministisch entsprechend der Rangfolge der Werte auf μ Individuen reduziert.

2. Aus den μ Individuen werden durch Kopieren oder durch Selektionsprozesse λ Individuen erstellt. Was hier aus dem Bereich der Biologie simuliert wird, ist die simultane Relevanz zweier Generationen für die Erzeugung von Nachkommen, in gewisser Weise wird den Individuen also eine Lebensdauer gegeben, die über eine einzelnen Generation hinausgeht. Sarma und De Jong (2000) sprechen hier auch von einem *Overlapping Population System* im Gegensatz zu einem *Nonoverlapping Population System*, bei dem wie bei den zuvor vorgestellten Methoden die alte Population vollständig durch Nachkommen ersetzt wird.

3.2.4.6 Vergleich der Einzelverfahren

Die Wahl für ein bestimmtes Einzelverfahren muss problem- bzw. anwendungsspezifisch getroffen werden, da sich Stärken und Schwächen der Verfahren nicht unabhängig vom Einsatzgebiet bestimmen lassen. *Roulette-Wheel Selection* stellt bspw. sicher, dass Genome, die relativ zu anderen eher schlecht abschneiden, nicht vollständig aus der Population verschwinden, so dass ihre Bestandteile für Rekombinationen noch zur Verfügung stehen. Der gleiche Umstand macht sie aber auch langsamer, da Zeit dafür aufgewendet wird, weniger angepasste Lösungen zu evaluieren.

Umgekehrt führt ein Verfahren wie *Tournament Selection* zu einer Eliminierung von Individuen mit geringer Fitness durch direkte Konkurrenz. Durch Wahl größerer Turniergrößen kann dieser Effekt verstärkt werden, durch Wahl einer größeren Anzahl von Siegern hingegen werden explizit Lösungen aufgenommen, die anderen hinsichtlich ihrer Fitness unterlegen sind. Der Vergleich zwischen *Roulette-Wheel Selection* und *Tournament Selection* hängt also deutlich von der Wahl der Parameter ab. Zudem kann eine vorhergehende Skalierung der Fitnesswerte die Eigenschaften der *Roulette-Wheel Selection* entscheidend verändern (s. dazu Abschnitt 3.2.5).

Die Geschwindigkeit der Konvergenz einer Population und die Anfälligkeit für lokale Optima stehen in einem Trade-Off-Verhältnis zueinander, eine gründlichere Suche im Lösungsraum wird mit einer Erhöhung der Laufzeit erkauft, die auch bei der Geschwindigkeit heutiger Prozessoren ein entscheidendes Argument gegen den Einsatz bieten kann. Die Probabilistic *Tournament Selection* stellt eine Art Mittelweg zwischen den beiden ersten Verfahren dar, dessen Angemessenheit aber ebenfalls nicht problem- und lösungsraumsunabhängig postuliert werden kann.

Die beiden deterministischen Verfahren bieten die Möglichkeit einer tieferen Analyse bei jedem Generationsschritt ohne den Nachteil, durch Vergrößerung der Populationen einen hohen Anteil von Individuen mit niedrigen Fitnesswerten zu erzeugen, da ein konstanter Prozentsatz dem Filter zum Opfer fällt. Hierbei liefert die $(\mu + \lambda)$-Variante die Möglichkeit, sehr explorative oder disruptive Operatoren zum Einsatz zu bringen, da die Ausgangsindividuen weiterhin zur Auswahl stehen und Misserfolge nicht zu negativen Sprüngen in der Entwicklung führen müssen. Erkauft wird dieser Vorteil durch einen Bias gegen potenzielle Zwischenstufen auf dem Weg aus lokalen Optima heraus. Verglichen mit den stochastischen Methoden sind diese Varianten weniger tolerant gegenüber Individuen mit niedrigen Fitnesswerten, da diese keine Chance haben, in die Folgegeneration übernommen zu werden. Durch die unter Umständen sehr große Anzahl generierter Lösungen wird dieser Effekt verstärkt, was die oben beschriebenen Chancen und Probleme aufwirft.

3.2.4.7 Varianten und Ergänzungen

Zahlreiche Modifikationen und Neuentwicklungen von Selektionsmethoden werden weltweit angewandt. Hier sei nur auf ein paar konzeptuell interessante Entwicklungen verwiesen. Was sich bei dem Vergleich von $(\mu+\lambda)$- und (μ, λ)-Verfahren bereits angeboten hat, ist das Konzept der Lebensdauer von Individuen. Es ist ebenso möglich, die Lebensdauer über eine Generation zu erhöhen, dabei unter Umständen auch Sonderregeln wie ein Inzestverbot zwischen Parental- und Filialgenerationen einzuführen.

Eine Variante der beschriebenen Methoden stellen Methoden mit einem *Generation Gap* dar (Sarma & De Jong, 2000). Dieser Wert gibt den Anteil der Population an, der bei einem Generationswechsel ersetzt wird. Für $G = 1$ ergibt sich die Ersetzung der gesamten Population, für $G = \frac{1}{\mu}$ wird nur ein einziges Individuum ersetzt. Dieses Prinzip wurde zuerst von De Jong (1975) untersucht. Der (μ, λ)-Selektion verwandt ist die so genannte *Threshold Selection* oder auch *Truncation Selection* (s. Grefenstette, 2000b), bei der ebenso wie bei der erstgenannten Form bei einem Teil der Population die Fortpflanzungswahrscheinlichkeit auf 0 gesetzt wird. Der übrige Teil wird allerdings nicht deterministisch übernommen, sondern durch eine Zufallsauswahl mit Gleichwahrscheinlichkeit der Alternativen. Formal ergibt sich:

$$P_{sel}(i) = \begin{cases} 0 & \text{if } rank(i) < (1-T)\mu \\ \frac{1}{T\mu} & \text{if } rank(i) \geq (1-T)\mu \end{cases} \tag{3.11}$$

mit $0 < T \leq 1$. Der Parameter T wird als Schwelle (*threshold*) bezeichnet.

Es ist möglich, dass von einer Generation zur nächsten das Individuum mit dem höchsten Fitnesswert in der früheren Generation entweder aufgrund eines probabilistischen Verfahrens nicht berücksichtigt wird für die nächste Generation oder durch die Applikation von Operatoren zerstört wird. Um diese Art der Verschlechterung der besten Lösung zwischen zwei Generationen zu vermeiden, bieten sich *elitistische* Strategien an. „Elitistisch“ bedeutet, dass das beste oder eine bestimmte Anzahl von besten Individuen einer Generation ohne Modifikation in die Folgegeneration übernommen werden, entweder deterministisch oder unter der Bedingung, dass sie einen höheren Fitnesswert aufweisen als die besten Individuen der neuen Generation(Sarma & De Jong, 2000). Elitismus ist im $(\mu+\lambda)$-Verfahren automatisch implementiert, Elitismus wird aber auch als zusätzliche Option bei anderen Verfahren angewendet. Vor- und Nachteil sind wiederum Zeitersparnis auf der einen, unter Umständen fehlende Gründlichkeit auf der anderen Seite.

Eine weitere Klasse von Verfahren benutzt eine Metapher aus der Thermodynamik. Bei der so genannten *Boltzmann Selection* wird ein *Simulated-Annealing*-Mechanismus zum Einsatz gebracht (Mahfoud, 2000). Dabei wird ausgehend von der Vorstellung einer progressiv abkühlenden Temperatur das Selektionsprizip graduell über die Generationen hinweg abgewandelt, so dass am Anfang die Qualität einer Lösung für die Auswahl kaum eine Rolle spielt und gegen Ende des Prozesses schlechtere Lösungen nahezu keine Chance mehr haben.

Kennzeichnend für die Boltzmann Selection ist das so genannte *Boltzmann Trial*: Sind i und j mögliche Lösungen bzw. Individuen und sind ihre Fitnesswerte mit f_i bzw. f_j gekennzeichnet, dann ergibt sich die Wahrscheinlichkeit dafür, dass sich i gegen j durchsetzt als:

$$p_i = \frac{1}{1 + e^{\frac{f_i - f_j}{T}}} \tag{3.12}$$

Dass mit dieser Definition das Prinzip des Simulated Annealing implementiert werden kann zeigen die Abbildungen 3.8 und 3.9.

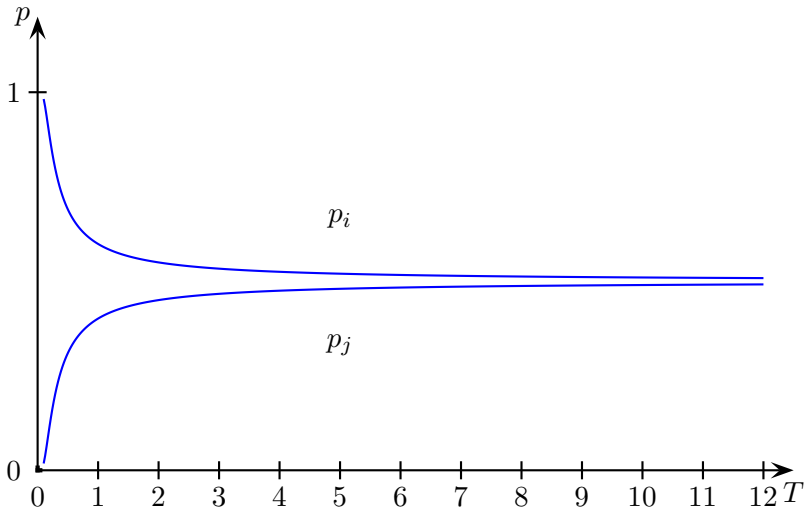


Abbildung 3.8: Boltzmann Trial 1: Wahrscheinlichkeiten bei $f_i - f_j = 0.4$

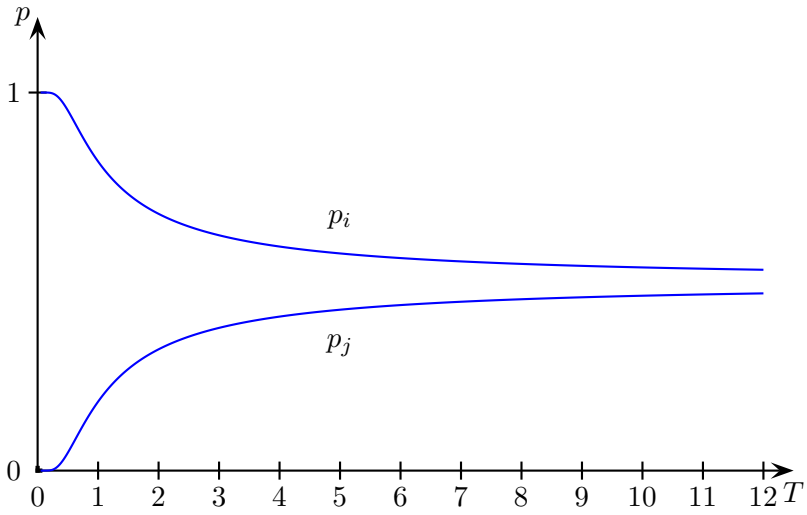


Abbildung 3.9: Boltzmann Trial 2: Wahrscheinlichkeiten bei $f_i - f_j = 3$

Für große Werte von T liegen die Wahrscheinlichkeiten für beide Alternativen um $p = 0,5$. Sinkt T, gehen die Wahrscheinlichkeiten gegen 1 bzw. gegen 0, was den oben beschriebenen Anforderungen entspricht.

Um eine frühzeitige Konvergenz zu verhindern und den Suchraum weiträumiger zu untersuchen, wird ebenfalls die Methode der *Disruptive Selection* zum Einsatz gebracht (Fogel, 2000a; Kuo & Hwang, 1993). Dabei wird die Fitnessfunktion so umgeformt, dass sich der Fitnesswert $u(x)$ eines Individuums berechnet

als

$$u(x) = |f(x) - f(t)| \,. \tag{3.13}$$

Dabei stellt $f(t)$ den Mittelwert aller Fitnesswerte zu einem Zeitpunkt t dar. Man könnte diese Methode als eine Skalierungsmethode bezeichnen, allerdings liegt der inhaltliche Fokus auf einer gravierenden Modifikation des Selektionsprinzips. Es werden durch diese Methode gleichzeitig die besten als auch die schlechtesten Individuen bevorzugt für die nächste Generation berücksichtigt, es entsteht so unter Umständen eine größere Variabilität im künstlichen Genpool. Die Nützlichkeit dieser Methode ist sicherlich abhängig von der Problemstellung (Fogel, 2000a).

Goh, Lim und Rodrigues (2003) schlagen die Variante der *Sexual Selection* vor. Die Population wird bei dieser Variante willkürlich in Männchen und Weibchen aufgeteilt. Dabei wird jedes Weibchen deterministisch für eine Paarung mit einem probabilistisch ermittelten Männchen gewählt. Die Autoren nutzen dabei die *Tournament Selection* mit einer Turniergröße von zwei. Der Vorteil dieser Methode wird zusammengefasst als:

> This design ensures that all the females get to reproduce regardless of their fitness and aims to facilitate *exploration* of the search space. On the other hand, bias towards the better performing males *exploits* the knowledge that the current chromosome-like structure of the individual is good. (Goh et al., 2003, S.127)

Ohne das Turnier der Männchen gäbe es keinerlei Selektionsdruck, der Mechanismus der Weibchen stellt eine breitere Suche sicher, da auch Weibchen mit relativ geringer Fitness berücksichtigt werden.

Dies soll als Überblick über alternative Konzeptionen und Variationen des Selektionsschrittes genügen. Abschließend lässt sich unterstreichen, dass die Auswahl des Selektionsverfahren vor dem Hintergrund der Auswahlentscheidungen für den gesamten Algorithmus getroffen werden muss. Führen Rekombinations- oder Mutationsschritte zu einer radikalen Umstrukturierung der Individuen, sind „konservative“ Selektionsverfahren angebrachter als bei vorsichtigeren Veränderungen (s.a. Hancock, 2000). Eine grundlegend andere Herangehensweise wird abschließend im folgenden Abschnitt diskutiert.

3.2.4.8 Interaktive Selektion

Eine ganz andere Zugangsweise zur Selektion bietet die *interaktive Selektion* (Banzhaf, 2000): Während bei den bisher beschriebenen Verfahren die Entscheidung für die positiv selegierten Individuen durch einen Algorithmus getroffen wurde, greift bei dieser Form der Selektionsmethoden der Anwender aktiv in den Prozess ein. Eine der ersten implementierten Anwendungen findet sich bei Dawkins (1987) zur Entwicklung von künstlichen *Biomorphs*. Ein Anwendungsgebiet, bei dem die Notwendigkeit zur Interaktion einsichtig wird, sind bspw. ästhetische Entscheidungen bei künstlich generierten Entwürfen oder künstlichen „Kunstwerken“. Damit die Entscheidungen handhabbar bleiben, muss die Auswahl allerdings auf eine überschaubare Anzahl von Möglichkeiten beschränkt bleiben, was eine große Einschränkung für die Anwendung evolutionärer Methoden bedeutet. Zum Teil können diese Probleme überwunden werden durch eine Integration des Anwenders in weitere Schritte des Algorithmus. Weitere Anwendungsgebiete für interaktive Selektion sind die Bereiche Animation, Bildverarbeitung, dynamische Systeme, Flugzeugdesign und Texturen (Banzhaf, 2000).

3.2.5 Fitnessskalierung

Um die Überlegungen zu verstehen, die zu der Praxis der Fitnessskalierung geführt haben, ist die Vertrautheit mit Selektionsverfahren - insbesondere mit der *Roulette-Wheel Selection* - notwendig, auch wenn die Skalierungsverfahren zeitlich genau vor diesem Schritt zum Einsatz kommen.

3.2.5.1 Additive Skalierung

Das vielleicht nahe liegende Problem, dass sich bei der Verwendung proportionaler Selektionsverfahren einstellt, ist folgendes: Nehmen die Fitnesswerte alle hoch positive Werte an, dann wird die Streuung der Werte in ihrer Auswirkung auf die Selektionswahrscheinlichkeiten durch den Abstand von Null reduziert, da für die Zuteilung der Selektionswahrscheinlichkeiten die Verhältnisse der Fitnesswerte von Bedeutung sind. So würden der Menge der Fitnesswerte $F_1 = \{1, 2, 3, 4, 5\}$ die Wahrscheinlichkeiten

$P_1 = \left\{\frac{1}{15}, \frac{2}{15}, \frac{1}{5}, \frac{4}{15}, \frac{1}{3}\right\}$ zugeordnet, der Menge $F_2 = \{11, 12, 13, 14, 15\}$ die Wahrscheinlichkeiten $P_2 = \left\{\frac{11}{65}, \frac{12}{65}, \frac{1}{5}, \frac{14}{65}, \frac{3}{13}\right\}$. Obwohl die Intervalle zwischen den Fitnesswerten für beide Populationen identisch sind, unterscheiden sich die Intervalle zwischen den zugeordneten Fitnesswerten stark voneinander. Sollte dieser Umstand nicht gewünscht sein, ist die einfachste Methode der Skalierung die Subtrakion des kleinsten in der Population auftretenden Fitnesswertes von allen Fitnesswerten der Population (Eshelman, 2000). Für die Population $A_x = \{a_1, a_2, \ldots, a_n\}$ zum Zeitpunkt t mit der Fitnessfunktion f ergibt sich die Skalierungsfunktion Φ als:

$$\Phi(f(a_i(t))) = f(a_i(t)) - \min_{a_i \in A_x} f(a_i(t)) \tag{3.14}$$

Nachteil dieser Methode ist ihre Anfälligkeit für Ausreißer nach unten.

3.2.5.2 Lineare Skalierung

Bei einer Anwendung proportionaler Verfahren ergibt sich im Verlauf der Generationen zusätzlich folgende Problematik: Zu Beginn des Durchlaufs weisen Populationsmitglieder sehr stark streuende Fitnesswerte auf, durch die Selektion wird aber gegen Ende des Laufes die Population auf sehr ähnliche Lösungen konvergieren, deren Fitnesswerte sich kaum noch unterscheiden. Dementsprechend ist die proportionale Selektionsmethode am Anfang eines Laufs wenig nachsichtig gegenüber schlechteren Lösungen, gegen Ende eines Laufs werden alle Lösungen mit fast gleicher Wahrscheinlichkeit selegiert. Um diesem Umstand zu begegnen, führt Goldberg (1989) zunächst die Methode der linearen Skalierung ein. Lineare Skalierung verwendet Skalierungsfunktionen der Form:

$$\Phi(f(a_i(t))) = af(a_i(t)) + b. \tag{3.15}$$

Ziel der vorgeschlagenen Skalierung ist die Möglichkeit der Bestimmung eines Erwartungswerts für die Anzahl der Nachkommen des Individuums mit dem höchsten Fitnesswert. Festgelegt wird, dass ein Individuum mit exakt durchschnittlicher Fitness einen Erwartungswert von 1 behalten soll. Im folgenden soll gelten:

$$\bar{f}(A_x(t)) \equiv \frac{1}{|A_x(t)|} \sum_{a_i(t) \in A_x(t)} f(a_i(t)) \tag{3.16}$$

$$\tag{3.17}$$

dementsprechend soll gelten:

$$\Phi(\bar{f}(A_x(t))) = \bar{f}(A_x(t)), \tag{3.18}$$

$$\Phi(f(a_{max}(t))) = C_{mult} \cdot \bar{f}(A_x(t)). \tag{3.19}$$

Die Abbildungen 3.10 und 3.11 zeigen die Bestimmung der Skalierungsfunktionen Φ_2 und $\Phi_{1,5}$ für $C_{mult} = 2$ bzw. $C_{mult} = 1.5$. Dabei seien $a_{min}(t)$ und $a_{max}(t)$ die Populationsmitglieder mit kleinstem, bzw. größtem Fitnesswert in der Population zum Zeitpunkt t.
Die Methode wirft allerdings das Problem auf, dass die Nonnegativitätsbedingung für die Fitnesswerte verletzt werden kann. Da lineare Funktionen streng monoton sind und die Größenreihenfolge der Messwerte als transformationsinvariant gesetzt wird, genügt die Betrachtung des Funktionswertes für das Element $a_{min}(t)$. Da für die Funktionen die beiden Punkte $(f(a_{av}(t))), \Phi(f(a_{avg}(t)))$ und $(f(a_{max}(t)), C_{mult} \cdot f(a_{avg}(t))$ festgelegt sind, ergibt sich der Funktionswert für das kleinste Element als:

$$\Phi(f(a_{min}(t))) = \bar{f}(A_x(t)) - (C_{mult} - 1) \cdot \bar{f}(A_x(t)) \cdot \frac{\bar{f}(A_x(t)) - f(a_{min}(t))}{f(a_{max}(t)) - \bar{f}(A_x(t))} \tag{3.20a}$$

$$= \bar{f}(A_x(t)) \left(1 - (C_{mult} - 1) \cdot \frac{\bar{f}(A_x(t)) - f(a_{min}(t))}{f(a_{max}(t)) - \bar{f}(A_x(t))}\right). \tag{3.20b}$$

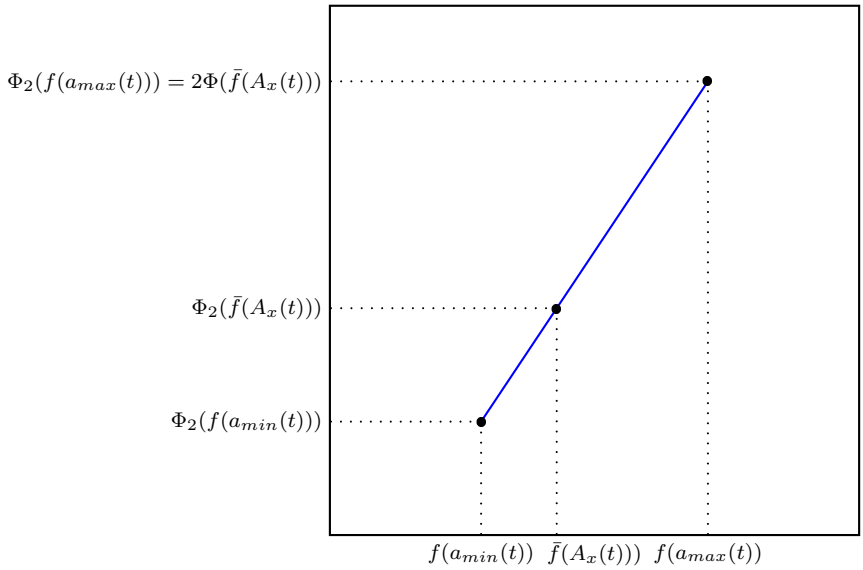


Abbildung 3.10: Lineare Skalierung 1: Transformation bei $C_{mult} = 2$

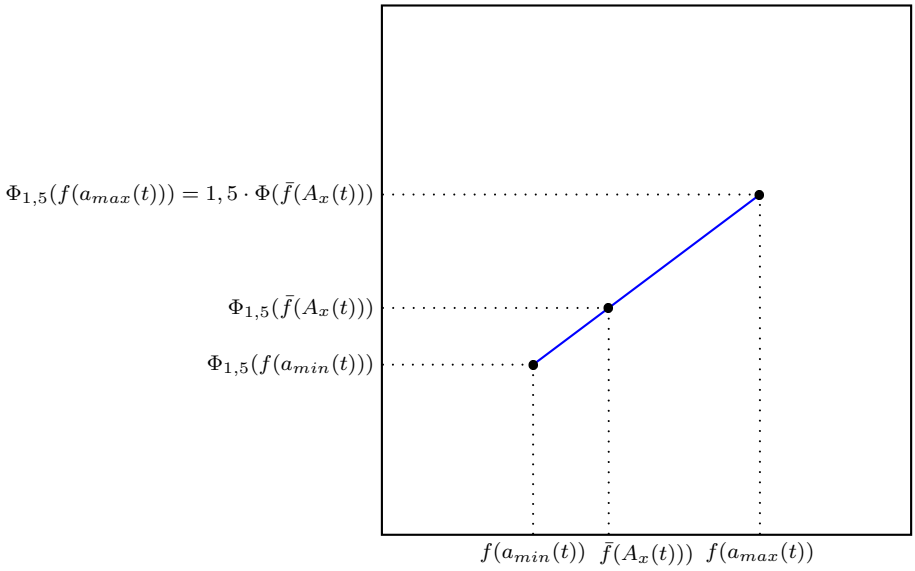


Abbildung 3.11: Lineare Skalierung 2: Transformation bei $C_{mult} = 1.5$

Die Forderung $\Phi(f(a_{min})) \geq 0$ (Nonnegativität) führt unter $C_{mult} > 1$ (Bevorzugung von Individuen mit höheren Fitnesswerten) und $f(a_{max}(t)) - \bar{f}(A_x(t)) > 0$ (Positivität der Streuung) zu:

$$\bar{f}(A_x(t)) \left(1 - (C_{mult} - 1) \cdot \frac{\bar{f}(A_x(t)) - f(a_{min}(t))}{f(a_{max}(t)) - \bar{f}(A_x(t))}\right) \geq 0 \tag{3.21a}$$

$$\equiv \quad \frac{f(a_{max}(t)) - C_{mult}\bar{f}(A_x(t)) + C_{mult}f(a_{min}(t)) - f(a_{min}(t))}{f(a_{max}(t)) - \bar{f}(A_x(t))} \geq 0 \tag{3.21b}$$

$$\equiv \quad f(a_{max}(t)) - C_{mult}\bar{f}(A_x(t)) + (C_{mult} - 1)f(a_{min}(t)) \geq 0 \tag{3.21c}$$

$$\equiv \quad f(a_{min}(t)) \geq \frac{C_{mult}\bar{f}(A_x(t)) - f(a_{max}(t))}{C_{mult} - 1} \tag{3.21d}$$

Bei einer Streuung der Fitnesswerte von 0 wäre jede Funktion mit nichtnegativem Wertebereich als Transformationsfunktion zulässig, wenn eine proportionale Selektionsmethode zum Einsatz kommt. Für den allgemeinen Fall markiert Ungleichung 3.21d die Bedingungen für eine Anwendung der oben beschriebenen Vorgehensweise für die lineare Skalierung. Da $f(a_{min}(t))$ aber ein empirischer Wert ist, kann der Fall eintreten, dass zu einem Zeitpunkt t für ein bestimmtes C_{mult} die Ungleichung nicht erfüllt ist. In diesem Fall würde eine lineare Skalierung zu negativen Fitnesswerten führen. Abbildung 3.12 zeigt ein Beispiel

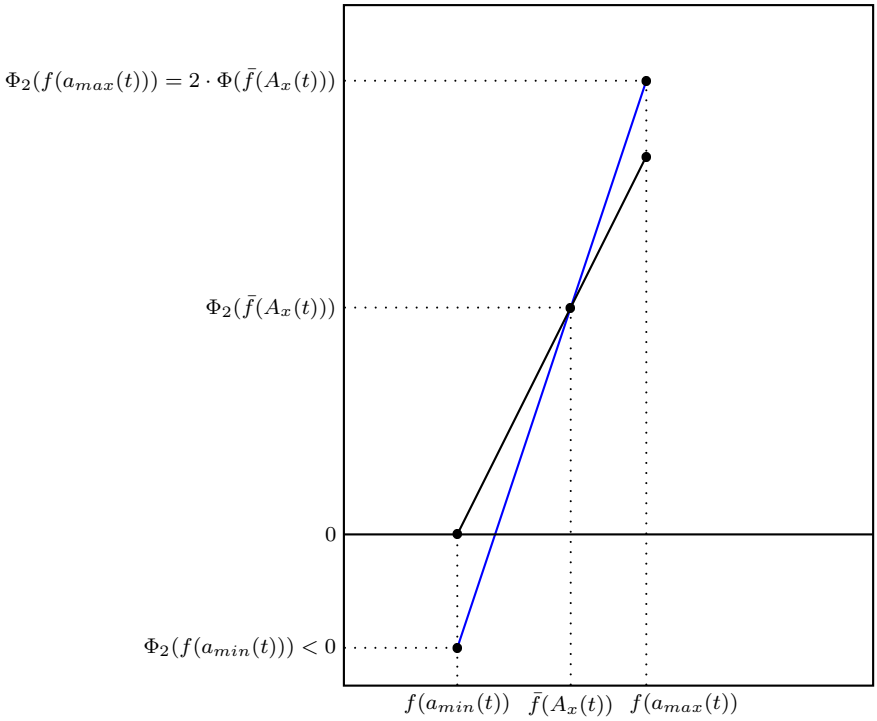


Abbildung 3.12: Lineare Skalierung 3: Transformation bei $C_{mult} = 2$ mit Verletzung der Nonnegativitätsbedingung; zusätzlich Funktion für $C_{mult} = \frac{5}{3}$

für das beschriebene Problem.

Aufgelöst für C_{mult} ergibt sich aus Ungleichung 3.21d:

$$C_{mult} \leq \frac{f(a_{max}(t)) - f(a_{min}(t))}{\bar{f}(A_x(t)) - f(a_{min}(t))} = C^*_{mult}(t) \tag{3.22}$$

Diese Ungleichung gibt die obere Grenze für C_{mult} an, wenn negative Fitnesswerte vermieden werden sollen. Eine übliche Strategie in der Anwendung von Skalierungen ist das Festsetzen eines Wertes C_{mult}, dessen Anwendbarkeit für jede Generation überprüft wird, um gegebenenfalls auf den höchsten zulässigen Wert korrigiert zu werden. Will man die lineare Skalierungsfunktion in Abhängigkeit von C_{mult} berechnen

über die beiden bekannten Funktionswerte, so erhält man:

$$\begin{aligned}
&\Phi_{C_{mult}}(f(a_i(t))) \\
=&\frac{C_{mult}\bar{f}(A_x(t)) - \bar{f}(A_x(t))}{f(a_{max}(t)) - \bar{f}(A_x(t))} f(a_i(t)) + \bar{f}(A_x(t)) - \bar{f}(A_x(t)) \frac{C_{mult}\bar{f}(A_x(t)) - \bar{f}(A_x(t))}{f(a_{max}(t)) - \bar{f}(A_x(t))} \\
=&\frac{(C_{mult}-1)\bar{f}(A_x(t))}{f(a_{max}(t)) - \bar{f}(A_x(t))} f(a_i(t)) + \bar{f}(A_x(t)) \frac{f(a_{max}(t)) - C_{mult}\bar{f}(A_x(t))}{f(a_{max}(t)) - \bar{f}(A_x(t))}
\end{aligned} \tag{3.23}$$

Setzt man für C_{mult} die Höchstgrenze $C^*_{mult}(t)$ aus Ungleichung 3.22 ein, so resultiert:

$$\begin{aligned}
\Phi_{C^*_{mult}(t)}(f(a_i(t))) = &\frac{(\frac{f(a_{max}(t)) - f(a_{min}(t))}{\bar{f}(A_x(t)) - f(a_{min}(t))} - 1)\bar{f}(A_x(t))}{f(a_{max}(t)) - \bar{f}(A_x(t))} f(a_i(t)) + \\
&\bar{f}(A_x(t)) \frac{f(a_{max}(t)) - \frac{f(a_{max}(t)) - f(a_{min}(t))}{\bar{f}(A_x(t)) - f(a_{min}(t))}\bar{f}(A_x(t))}{f(a_{max}(t)) - \bar{f}(A_x(t))} \\
= &\frac{\bar{f}(A_x(t))}{\bar{f}(A_x(t)) - f(a_{min}(t))} f(a_i(t)) - \frac{\bar{f}(A_x(t))}{\bar{f}(A_x(t)) - f(a_{min}(t))}
\end{aligned} \tag{3.24}$$

Aus 3.21d, 3.23 und 3.24 ergibt sich damit die allgemeine Gleichung:

$$\begin{aligned}
&\Phi_{C_{mult}}(f(a_i(t))) = \\
&\begin{cases}
\frac{(C_{mult}-1)\bar{f}(A_x(t))}{f(a_{max}(t)) - \bar{f}(A_x(t))} f(a_i(t)) + \bar{f}(A_x(t)) \frac{f(a_{max}(t)) - C_{mult}\bar{f}(A_x(t))}{f(a_{max}(t)) - \bar{f}(A_x(t))} & , C_{mult} \leq C^*_{mult}(t) \\
\frac{\bar{f}(A_x(t))}{\bar{f}(A_x(t)) - f(a_{min}(t))} f(a_i(t)) - \frac{\bar{f}(A_x(t))}{\bar{f}(A_x(t)) - f(a_{min}(t))} & , C_{mult} > C^*_{mult}(t)
\end{cases}
\end{aligned} \tag{3.25}$$

mit

$$C^*_{mult}(t) = \frac{f(a_{max}(t)) - f(a_{min}(t))}{\bar{f}(A_x(t)) - f(a_{min}(t))} \tag{3.26}$$

Eine Anwendung auf die Population $A_1(0)$ mit den Fitnesswerten $\{f_{a_1}(0) = 0,8; f_{a_2}(0) = 1,2; f_{a_3}(0) = 1,3; f_{a_4}(0) = 1,4; f_{a_5}(0) = 1,4; f_{a_6}(0) = 1,8; f_{a_7}(0) = 2,2; f_{a_8}(0) = 2,5; f_{a_9}(0) = 2,8; f_{a_{10}}(0) = 3,1; f_{a_{11}}(0) = 3,2; f_{a_{12}}(0) = 3,3; f_{a_{13}}(0) = 3,5; f_{a_{14}}(0) = 3,6; f_{a_{15}}(0) = 3,9\}$ und für drei verschiedene Parameterwerte zeigt Abbildung 3.13.

3.2.5.3 Sigma Truncation

Eine Variante der linearen Skalierung, die insbesondere das Problem einzelner Ausreißer angeht, ist das Prinzip der vorgeschalteten *Sigma Truncation*. Diese folgt dem Kalkül, Populationsmitglieder mit relativ niedriger Fitness aus dem Selektionsprinzip auszuschließen, indem ihre Fitness auf 0 gesetzt wird. Alle anderen Werte können dementsprechend näher an den Nullpunkt gesetzt werden. Um diesen Effekt zu formalisieren, wird auch diese Methode mit einem Parameter c versehen. Das namensgebende σ referiert zudem auf die Verwendung der Standardabweichung der Fitnesswerte in der Population. Ein wenig gegen den strikten Gebrauch des Symbols wird hier die empirische Standardabweichung betrachtet. Der Parameter c legt den Bereich von Fitnesswerten fest, der im Selektionsverfahren ignoriert werden soll. Dabei werden alle Werte auf 0 gesetzt, deren Fitnesswert eine untere Grenze nicht übertrifft, die durch die Differenz von Mittelwert und c-facher Standardabweichung definiert wird, von den übrigen Werten wird das Produkt aus Standardabweichung und c subtrahiert. Je höher c gesetzt wird, desto weniger selektiv ist das Verfahren. Übliche Werte liegen zwischen 1 und 3 (Goldberg, 1989). Nach Grefenstette (2000a) lässt sich dementsprechend die Skalierungsfunktion angeben als:

$$\Phi_c(f(a_i(t))) = \begin{cases}
f(a_i(t)) - (\bar{f}(A_x(t))) - c \cdot \sigma(A_x(t)) & , f(a_i(t)) > \bar{f}(A_x(t))) - c \cdot \sigma(A_x(t)) \\
0 & , f(a_i(t)) \leq \bar{f}(A_x(t))) - c \cdot \sigma(A_x(t)) \\
1 & , \sigma(A_x(t)) = 0
\end{cases} \tag{3.27}$$

mit

$$\sigma(A_x(t)) = \sqrt{\frac{1}{|A_x(t)| - 1} \sum_{a_i \in A_x(t)} \left(f(a_i(t)) - \bar{f}(A_x(t))\right)^2}. \tag{3.28}$$

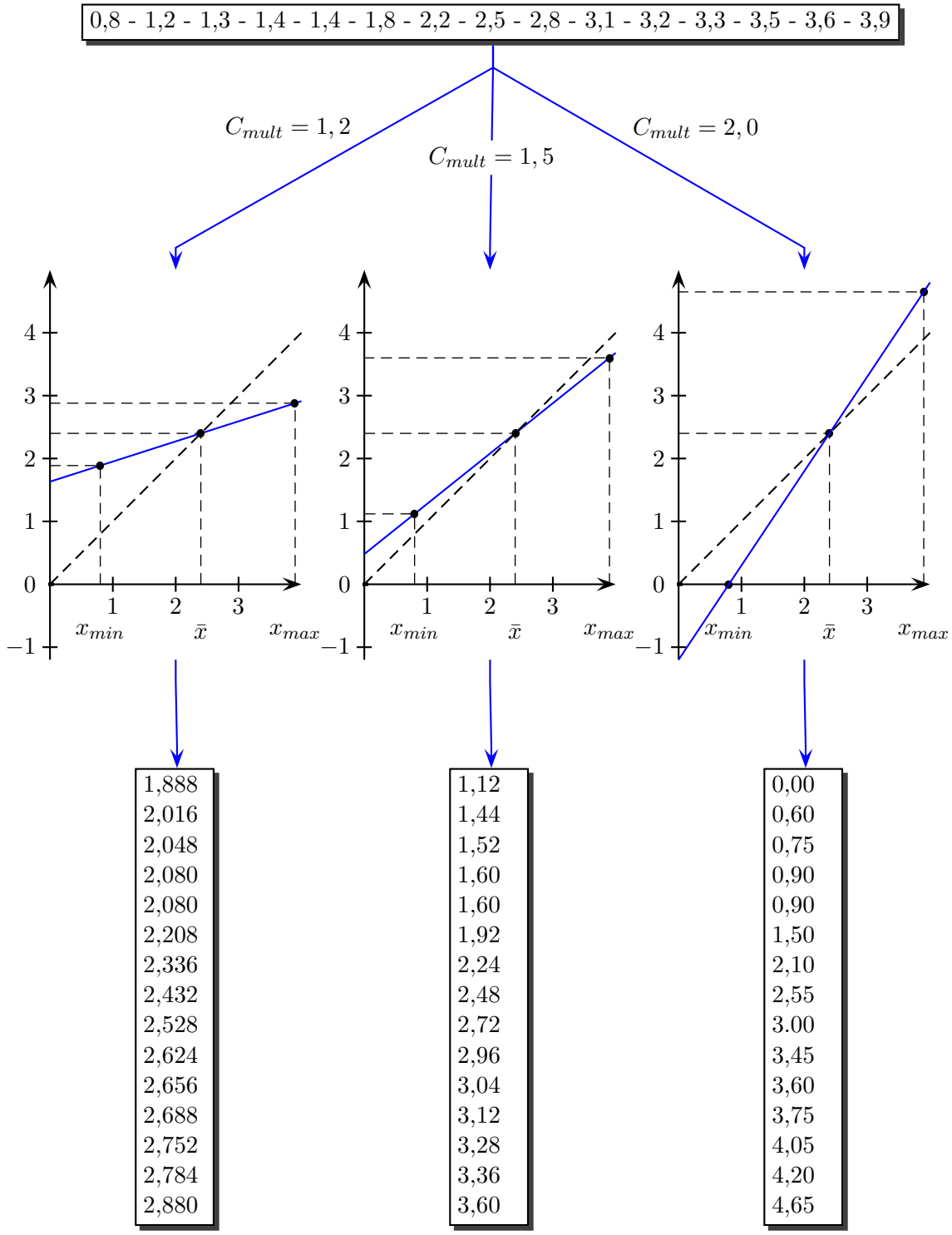


Abbildung 3.13: Lineare Skalierung für $A_1(0)$: Transformationen für drei verschiedene Parameterwerte

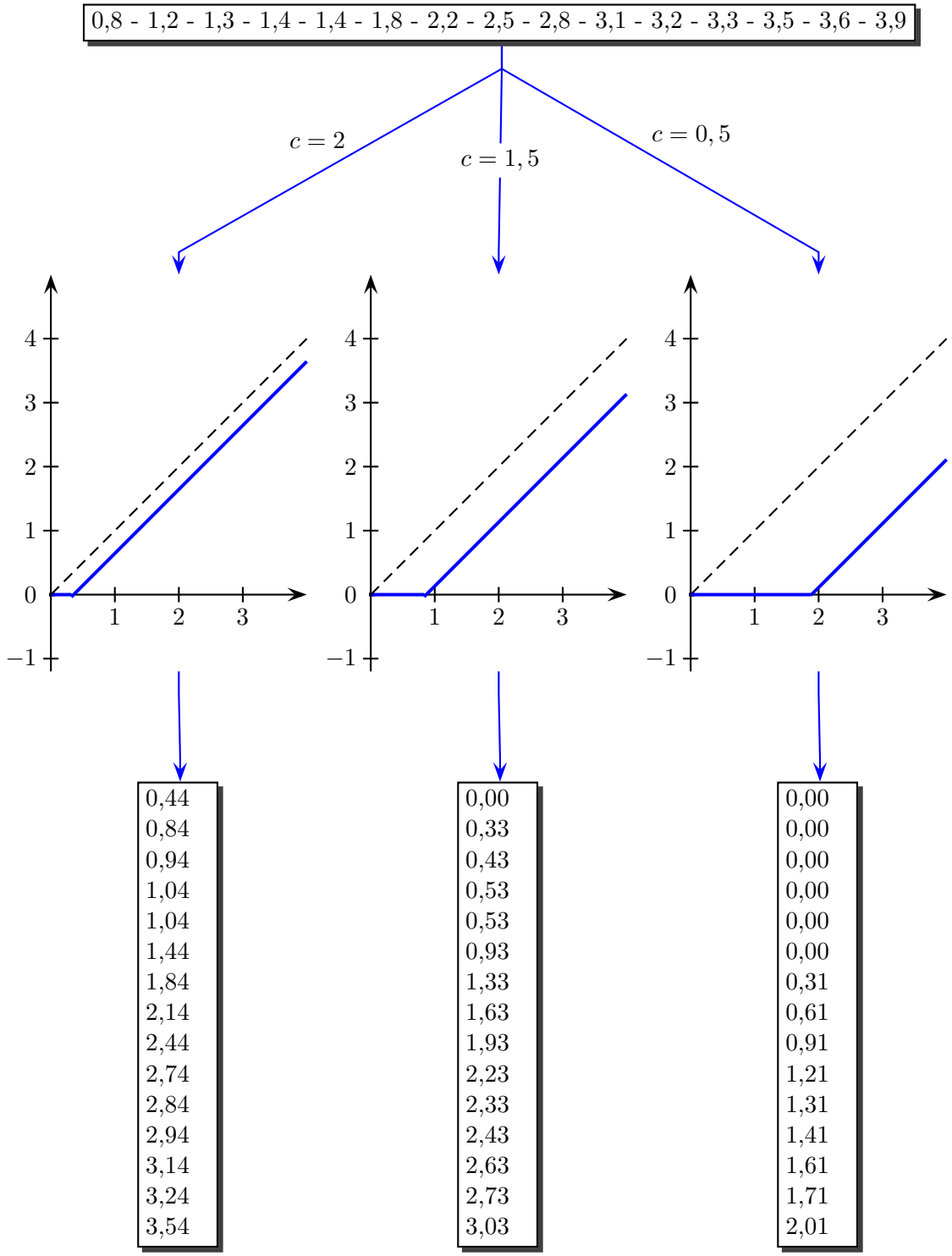


Abbildung 3.14: Sigma Truncation für $A_1(0)$: Transformationen für drei verschiedene Parameterwerte

Mitchell (1996) weist darauf hin, dass für den Fall einer Standardabweichung von 0 alle Individuen der Population den gleichen Erwartungswert für die Anzahl der Nachkommen erhalten sollten, deshalb wurde eine dritte Bedingung in die Formel eingeführt. Abbildung 3.14 zeigt die Anwendung der Transformation mit drei verschiedenen Parameterwerten auf die Population $A_1(0)$. Es ist durchaus üblich, das Verfahren der Sigma Truncation der linearen Skalierung vorzuschalten (Goldberg, 1989).

3.2.5.4 Power Scaling

Neben der linearen Skalierung sind zahlreiche nichtlineare Varianten der Skalierung denkbar. Exemplarisch soll hier die Variante des *Power Scalings* vorgestellt werden (s. Goldberg, 1989). Hierbei werden die Fitnesswerte mit Hilfe einer Potenzfunktion skaliert in der Form:

$$\Phi_k(f(a_i(t))) = f(a_i(t))^k \tag{3.29}$$

Der Parameter k bestimmt die Form der Funktion, wobei in praktischen Anwendungen oft Werte im Bereich knapp über 1 eingesetzt werden (Michalewicz, 1996). Je höher k gewählt wird, desto stärker werden höhere Fitnesswerte gegenüber niedrigen Fitnesswerten im Selektionsprozess bevorzugt, bei Werten unter 1 hingegen rücken die Werte stärker zusammen. Die Wahl von k ist damit stark problemabhängig und kann nicht ohne Kenntnis über Kodierung und Fitnessfunktion getroffen werden. Abbildung 3.15 zeigt eine Anwendung des *Power Scaling* mit drei verschiedenen Parametersetzungen für $A_1(0)$.

3.2.5.5 Transformation in Rangwerte

Eine letzte vorgestellte Skalierungstransformation beruht nicht auf der Verwendung einer stetigen Funktion, sondern auf der Umformung der Fitnesswerte in Rangwerte. Als Konsequenz werden alle Differenzen zwischen Fitnesswerten zu einer Standarddifferenz zwischen Vorgänger und Nachfolger in der Rangreihe reduziert. Dabei können die Fitnesswerte über die Transformation

$$\Phi_k(f(a_i(t))) = s_u + \frac{rg(a_i(t), A_x(t)) - 1}{|A_x(t)| - 1} \cdot (s_o - s_u) \tag{3.30}$$

mit

$$rg(a_i(t), A_x(t)) \equiv \text{Rang von Individuum } a_i \text{ in der Population } A_x \text{ zum Zeitpunkt t} \tag{3.31}$$

auf jedes beliebige Intervall $[s_o, s_u]$ mit $s_o, s_u \in \mathbb{R}$ abgebildet werden. Die untere und die obere Grenze dieses Intervalls (s_u und s_o) stellen die Parameter für die Methode dar. Abbildung 3.16 demonstriert die Anwendung der Methode für drei Parametersetzungen auf $A_1(0)$.

Für die Varianten der Tournament Selection oder die (μ, λ)- bzw. $(\mu + \lambda)$-Verfahren wäre die Anwendung der beschriebenen Methode sinnlos, da diese Verfahren nur Ranginformationen ausnutzen und diese bei dieser Transformation invariant bleiben. Die Angemessenheit der Methode bei einem proportionalen Selektionsalgorithmus muss problemabhängig bestimmt werden.

3.2.6 Abbruchkriterien

Ein evolutionärer Prozess hat kein natürliches Ende: So wenig wie die natürliche Evolution ein Ziel kennt, so stark ist der Anwender allerdings an einem Ergebnis interessiert, wenn der Bereich des maschinellen *Online-Learning* einmal ausgeklammert wird. Da das Kriterium zum Abbruch des Algorithmus nicht in einem der anderen Schritte enthalten ist, muss es explizit in den Algorithmus eingeführt werden. Auch an dieser Stelle gibt es eine Reihe von Auswahlmöglichkeiten.

3.2.6.1 Generationenzahl

Das einfachste Kriterium stellt sich als Angabe einer bestimmten Zahl von Generationen G_{gen} dar, nach der der Algorithmus angehalten wird. Diese Setzung ist nicht ohne Risiko. Ein zu niedriger Wert von G_{gen} unterbricht die Suche zu früh, bevor überhaupt der relevante Bereich im Suchraum gefunden worden ist, oder bevor dieser gründlich genug untersucht werden konnte. Ein zu hoher Wert von G_{gen} hingegen ist zunächst einmal unnötig zeitintensiv. Dieser Faktor kann durch eine entsprechende Prozessorausstattung sicherlich drastisch, wenn auch nicht beliebig reduziert werden. Zum anderen kann es aber bei einer zu

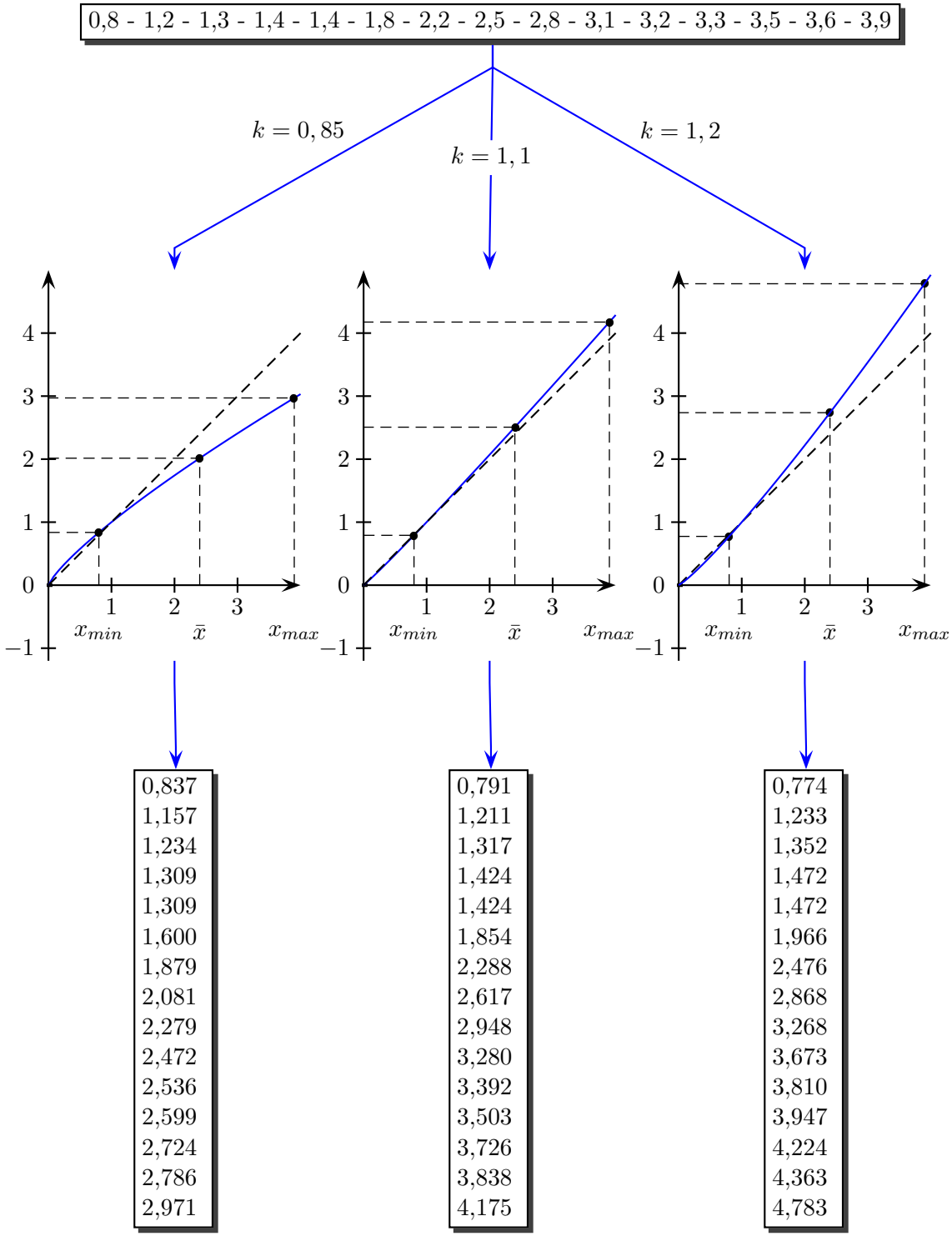


Abbildung 3.15: Power Scaling für $A_1(0)$: Transformationen für drei verschiedene Parameterwerte

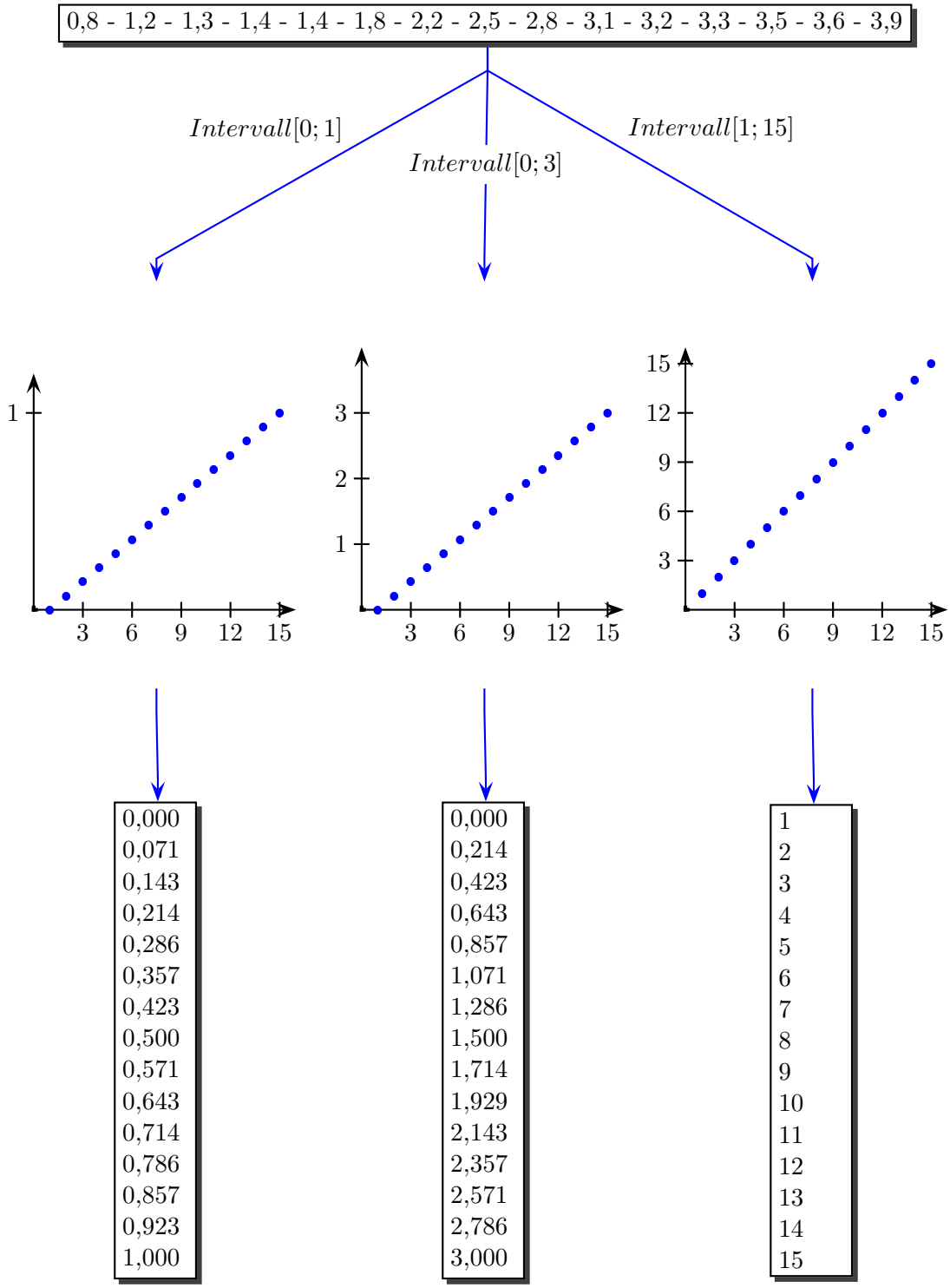


Abbildung 3.16: Rangtransformation für $A_1(0)$: Transformationen für drei verschiedene Parameterwerte

langen Laufzeit zu einer Überanpassung an die Daten kommen. Ist das Ziel der Applikation ein reines Fitting, ist eine Überanpassung nicht möglich (eine vollständige Anpassung wäre optimal), hingegen wird dieses Phänomen bei prädiktiven Modellen oder Generalisierungen über den Trainingsdatensatz hinaus zu einem substanziellen Problem. Aus dem Grund muss dieses Kriterium mit Vorsicht eingesetzt werden, unter Umständen sind mehrere Läufe mit verschiedenen Werten für G_{gen} empfehlenswert.

3.2.6.2 Zielkriterium für die Modellgüte

Eine zweite Möglichkeit, die Generationenfolge zu beenden, ist die Setzung einer Grenze für den Fitnesswert des besten Individuums $G_{f_{max}(A_x(t)}$ oder für die maximale Klassifikationsleistung G_{class}. Die Evolution wird so lange fortgesetzt, bis diese Grenze vom besten Individuum der Population überschritten wird.

Die Setzung dieses Wertes erfordert Kenntnisse über den Lösungsraum. Sind die Forderungen zu hoch, d.h. existiert im gesamten Lösungsraum keine Lösung, die dem geforderten Qualitätsmaßstab genügt, läuft der Algorithmus unendlich lange. Ist die Grenze zu niedrig gesetzt, wird die Möglichkeit verschenkt, eine bessere Lösung zu finden. Eine Alternative stellt dementsprechend eine Kombination der beiden vorgestellten Möglichkeiten dar: Um ein *Overfitting* zu vermeiden, wird ein Kriterium $G_{f_{max}(A_x(t)}$ gesetzt, gleichzeitig aber wird auch eine Grenze G_gen eingeführt, bei der der Algorithmus abgebrochen wird, wenn bis zu diesem Zeitpunkt die Grenze noch nicht überschritten worden ist. Die beiden Abbruchkriterien werden also disjunktiv verknüpft.

Wahlweise kann auch eine Minimalzahl von Generationen M_{gen} gesetzt werden, damit eine Suche nicht zu schnell abgebrochen wird. Dementsprechend sähe die Schleife in Pseudocode folgendermaßen aus:

```
GENERATIONENSCHLEIFE(population, anzgen, M_gen, G_gen, G_fmax)
1  while (anzgen < M_gen) OR ((BESTEFITNESS(population) < G_fmax) AND (anzgen<G_gen))
2  do NEUE_GENERATION(population)
3      anzgen ← anzgen + 1
4  ABBRUCH(population, anzgen)
```

Listing 3.1: Pseudocode für die Generationenschleife mit drei Grenzwerten

3.2.6.3 Stagnation des Fortschritts

Ist nicht genug über den Suchraum der Anwendung bekannt, um einen der oben genannten Werte bestimmen zu können, lässt sich eine Lösung finden, die lediglich die beobachtbare Dynamik des Suchprozesses untersucht. Hierzu werden die beiden Parameter U_{rel} und n_{gen} bestimmt. Ist die Verbesserung im besten Modell für n_{gen} Generationen maximal U_{rel}, dann wird der Algorithmus abgebrochen. Aufgrund der stochastischen Natur des Algorithmus sollten sehr niedrige Werte für n_{gen} vermieden werden, außerdem sollte eine Mindestlaufzeit eingedacht sein. Es ist ebenfalls zu bedenken, dass die Feinoptimierung durch einen genetischen Algorithmus unter Umständen sehr lange dauern kann, andererseits droht gerade in diesem Bereich eine Überanpassung an die Daten.

Die Wahl von U_{rel} ist ebenfalls problemabhängig und hängt insbesondere von der verwendeten Fitnessfunktion ab, wenn Fitnesswerte zur Kontrolle untersucht werden. Auch hier können mehrere Läufe notwendig sein, um geeignete Werte abschätzen zu können. Neben dem Fitnesskriterium kann auch die Klassifikationsleistung als Kontrollwert betrachtet werden. Dies ist unter Umständen dann relevant, wenn in den Fitnesswert Faktoren eingerechnet sind, die eher der Steuerung der Suche dienen als einer Beschreibung der Lösungsqualität.

3.2.6.4 Kreuzvalidierung

Eine Methode, die bei Neuronalen Netzwerken in vielen Fällen Verwendung findet (Woike, 2000), ist die Methode der Kreuzvalidierung zur Bestimmung des Abbruchzeitpunkts. Dazu wird das Set der Daten in ein Trainingsset und ein Testset aufgeteilt. Sichtbar für den genetischen Algorithmus ist für alle anderen Operatoren ausschließlich das Trainingsset, lediglich zur Überprüfung des Abbruchkriteriums wird auf

das Testset zurückgegriffen. Ein Indikator für eine Überanpassung an die Daten unter Verlust von Generalisierbarkeit ist eine Verschlechterung der Ergebnisse für das Testset bei einer gleichzeitigen Verbesserung der besten Lösung im Trainingsset, das im Idealfall die über generalisierbare Merkmale hinausgehenden Besonderheiten des Trainingssets nicht teilt. Das Verfahren könnte ähnlich eingesetzt werden wie das Verfahren zur Evaluation der Stagnation des Fortschritts. In regelmäßigen Generationsabständen wird anhand des Testsets die Fitness überprüft, und das Verfahren wird abgebrochen, wenn die Verbesserung über eine Anzahl von Vergleichen hinweg eine bestimmte Verbesserung nicht überstiegen hat, was eine Verschlechterung mit einschließt (s.a. Abbildung 3.17.

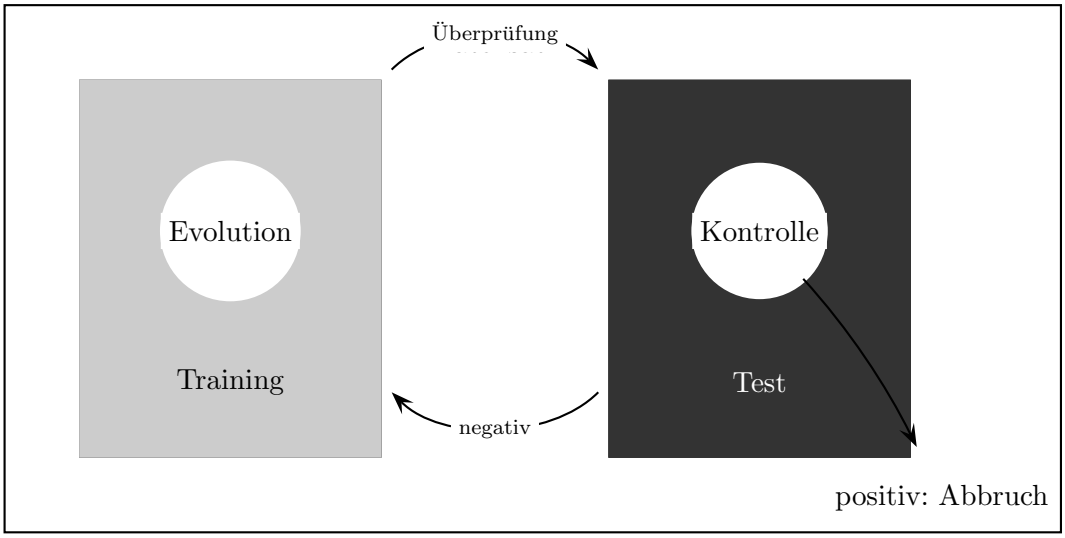


Abbildung 3.17: Abbruch über Kreuzvalidierung: Die Evolution wird solange fortgesetzt, bis sich die Ergebnisse im Testset verschlechtern.

Prinzipiell sind auch andere Prinzipien denkbar. Beispielsweise könnten Charakteristika der Population herangezogen werden, um im Falle einer Konvergenz der Population auf einen einzigen Genotyp und einer Anzahl erfolgloser Mutationen das Verfahren abzubrechen. Ein genetischer Algorithms wehrt sich dabei keineswegs gegen die Kreativität eines Anwenders, der problemspezifisch eigene Lösungen integrieren möchte.

3.3 Genetische Operatoren

In diesem Abschnitt sollen zwei charakteristische Operatorklassen von genetischen Algorithmen diskutiert werden, deren Eigenschaften zusammen mit denen des Selektionsoperators den Suchprozess in wesentlicher Weise gestalten. Zunächst sollen dabei Mutationsoperatoren und im Anschluss Rekombinationsoperatoren vorgestellt werden.

3.3.1 Mutation

Ohne irgendeine Form der Mutation oder Variation blieben in einem Bereich die Versatzstücke, die zum Handeln zur Verfügung stehen immer die gleichen, eine Erschaffung von etwas Neuem bliebe auf das Kopieren des Bestehenden beschränkt, vielleicht in neuen Kombinationen, aber ohne neuartige Veränderung eines Bestandteils. Es gäbe in der Sprache keine Neologismen, keine Neuentstehung von Untersuchungsgebieten oder Konzepten, die sich nicht vollständig als Rearrangement von bestehenden erklären ließen, die Entwicklungsmöglichkeiten wären stark eingeschränkt. Natürlich wäre im Gegenzug die Vorhersagbarkeit dieser Bereiche und die damit verbundene Sicherheit deutlich größer, es würden mit Sicherheit auch weniger Fehler begangen, aber bezeichnenderweise sind es gerade Fehler, die Neues kreieren und Lösungsideen erzeugen können. Hätte es im Prozess des biologischen Kopierens keine Fehler gegeben, wäre Evolution undenkbar gewesen, Kopierfehler erzeugen erst die Vielfalt, die zu einer Auswahl führen kann, eine Auswahl zwischen identischen Objekten stellt sich hingegen als reichlich sinnlos dar.

3.3.1.1 Funktion der Mutation in genetischen Algorithmen

Wäre die im Genpool der Population enthaltene Information invariant, so wäre die Suche beschränkt auf Lösungen, die im Kern bereits in der ersten Population vorliegen. Sicherlich könnten Teile der Lösung rearrangiert, schwächere Elemente aussortiert werden, aber alles, was in der ersten Population nicht angelegt wäre, könnte im Suchverfahren nicht gefunden werden.

Um dieser Einschränkung zu entkommen, machen die meisten genetischen Algorithmen Gebrauch von einem sehr einfachen genetischen Operator. Die *Mutation* lässt sich am besten beschreiben als zufallsbedingte Veränderung der genetischen Information. Dabei entsteht ein natürliches Dilemma: Ist die Wahrscheinlichkeit für eine relevante Veränderung der Struktur zu hoch, ist die Population sehr unbeständig, und die Suche im Lösungsraum kann nicht konzentriert erfolgen, sie bleibt sprunghaft und sporadisch. Ist die Wahrscheinlichkeit aber zu niedrig, so wird die Suche zwar gründlich und beharrlich sein, unter Umständen aber über lange Zeit in Bereichen des Lösungsraums erfolgen, in denen die gewünschte Lösung nicht zu finden ist. Die so genannte *Mutationsrate* hat also einen entscheidenden Anteil an der Form des Suchprozesses.

3.3.1.2 Implementierung der einfachen Mutation

Die einfachste Form der Mutation für ein binär kodiertes Genom lässt sich in folgendem Pseudo-Code beschreiben:

```
MUTATIONSOPERATOR(population)
1  for a ← 1 to population.size
2  do for i ← 1 to LENGTH(population.individuum[a].genomstring)
3      do r ← ZUFALLSZAHL(0, 1)
4          if r < mutationsrate
5             then if population.individuum[a].genomstring[i] = 0
6                     then population.individuum[a].genomstring[i] ← 1
7                     else  population.individuum[a].genomstring[i] ← 0
```

Listing 3.2: Pseudocode für den Mutationsoperator

Für jede Stelle im Genom eines jeden Individuum wird durch eine (pseudo-)zufällig generierte Zahl ermittelt, ob eine Mutation stattfindet oder nicht. Im positiven Fall wird eine 0 durch eine 1, bzw. eine 1 durch eine 0 ausgetauscht. Abbildung 3.18 demonstriert die Funktionsweise des Operators in visueller

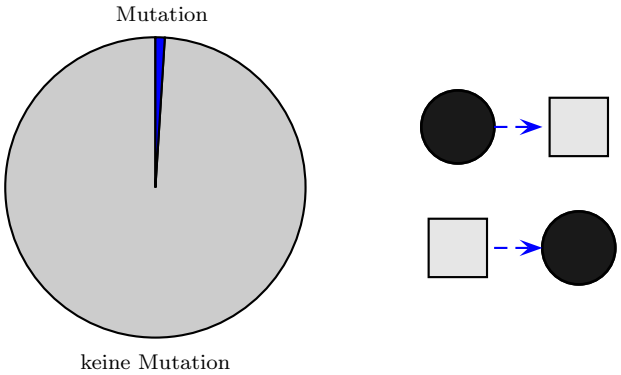


Abbildung 3.18: Funktionsweise des Mutationsoperators: die zwei Formen symbolisieren die Zahlen 0 und 1 (die Zuordnung ist für das Beispiel irrelevant)

Sprache. Quadrate und Kugeln symbolisieren in diesem Fall die Zahlen 0 und 1. Für ein Beispielgenom ist in Abbildung 3.19 die Mutationsschleife durchlaufen worden. Üblicherweise wird die Mutationsrate auf sehr kleine Werte festgelegt, ein Richtwert wäre etwa bei einem Genom der Länge n ein Wert von $p_{mut} = \frac{1}{n}$.

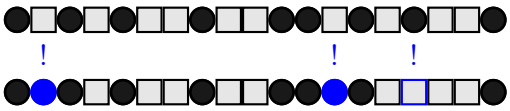


Abbildung 3.19: Beispiel für die Anwendung des Mutationsoperators: An den mit einem Ausrufezeichen markierten Positionen ist eine Ersetzung vorgenommen worden

3.3.2 Rekombination

In diesem Abschnitt soll *Rekombination* als eine Zusammenstellung von Elementen verstanden werden, die ohne eine Veränderung oder Umpositionierung der Elemente auskommt. Das Austauschen zweier Knöpfe an einem Kleidungsstück würde in diesem Sinne als Rekombination betrachtet. Sogar der Austausch von Hosenbeinen zwischen zwei Hosen wäre eine Rekombination, das Aufnähen eines Flickens oder das Kürzen eines Hosenbeins wären hingegen Mutationen. Das Potenzial der Rekombination liegt in der geschickten Kombination dessen, was vorliegt, was in gewisser Weise schon besessen wird, aber nicht unbedingt vorteilhaft gebündelt. Dass aus Bestehendem Neuartiges durch reine Rekombination entstehen kann, zeigen eindrucksvoll die Techniken der Parallelmontage im Film, das Zitieren in wissenschaftlichen Texten, das Phänomen von Coverversionen in der Musikindustrie (die allerdings oft eher als Mutationen zu klassifizieren sind), intertextuelle Bezüge in der Literatur oder auch Unterschiede zwischen Köchen, die auf die gleichen Zutaten zurückgreifen können. In Firmennetzwerken wird das Wissensmanagement zudem ebenfalls zu einer Frage der Rekombination des Wissens Einzelner.

Damit wird deutlich, dass der Erfolg von Rekombinationsprozessen auf den Einsatz der Mutation angewiesen ist, damit eine Vielzahl an Elementen entstehen kann, die rekombiniert werden können. Im Rahmen von genetischen Algorithmen wird die Rekombination allerdings nicht wie in den Beispielen explizit als kreative Technik eingesetzt, sondern als stochastisch determinierter, technischer Operator. Unterschiede hinsichtlich der Qualität von Lösungen werden ausschließlich von Selektionsoperatoren erfasst, die Rekombination ist wie die Mutation blind für die Konsequenzen ihrer Anwendung - sie ist ein syntaktischer Operator - was die Notwendigkeit der Kombination von Selektions-, Mutations- und Rekombinationsoperatoren unterstreicht.

3.3.2.1 Funktion der Rekombination in genetischen Algorithmen

In der Sprechweise der Forschungsgemeinschaft wird die Rekombination in genetischen Algorithmen oft als Cross-Over-Prozess in Analogie zum biologischen Vorbild beschrieben. Rekombination bedeutet im Gegensatz zur Mutation, dass bezogen auf die Veränderung von Genomstrukturen keine Elemente isoliert modifiziert werden, sondern ausschließlich Kopierprozesse zum Einsatz kommen. Kopieren beinhaltet in diesem Fall nicht die Verschiebung von Elementen zwischen zwei Stellen des gleichen Genoms, sondern die Auswahl aus verschiedenen Varianten einer Genomstelle für jede Stelle einzeln. Kombiniert werden also Varianten für Stellen in einem Genom in Analogie zu Allelen, *re*-kombiniert werden sie, weil sie in der Population bereits vorhanden sein müssen. Im Folgenden sollen die gebräuchlichen Formen des Operators diskutiert werden.

3.3.2.2 Single Point Cross-Over

Die einfachste Variante der Rekombination stellt das *Single Point Cross-Over* dar. Alle Cross-Over-Operatoren beziehen sich im Unterschied zu Mutationsoperatoren auf Genompaare oder auf Gruppen von Genomen, in Parallele zu Prozessen der geschlechtlichen Fortpflanzung in biologischen Systemen. Es kommt dabei zu einer Rekombination der genetischen Information aller Beteiligten. Im klassischen Fall wird der Operator auf Individuenpaare angewandt. Hierzu wird die Population in Paare unterteilt, wobei für jedes Paar mit einer Wahrscheinlichkeit p_cross der Cross-Over-Operator zum Einsatz kommt. Die Vorgehensweise kann aus folgendem Pseudo-Code rekonstruiert werden:

```
SINGLE_POINT_CROSS_OVER(pop)
 1  ZUFALLSREIHENFOLGE(pop)
 2  for a ← 1 to pop.sizeDIVDIV2
 3  do r ← ZUFALLSZAHL(0, 1)
 4      if r < p_cross
 5         then s ← ZUFALLSZAHL_GANZ(1, LENGTH(pop.indiv[2 ∗ a − 1].genom) − 1)
 6              for i ← 1 to s
 7              do child[1].genom[i] ← pop.indivi[2 ∗ a − 1].genom[i]
 8                 child[2].genom[i] ← pop.indivi[2 ∗ a].genom[i]
 9                 for i ← s + 1 to LENGTH(pop.indiv[2 ∗ a − 1].genom)
10              do child[1].genom[i] ← pop.indiv[2 ∗ a].genom[i]
11                 child[2].genom[i] ← pop.indiv[2 ∗ a − 1].genom[i]
12                 pop.indiv[2 ∗ a − 1].genom ← child[1].genom
13              pop.indiv[2 ∗ a].genom ← child[2].genom
```

Listing 3.3: Pseudocode für Single-Point Cross-Over

Es wird eine Genomposition zufällig ausgewählt, um die genetische Information der Eltern zu rekombinieren. Dabei entstehen zwei Nachkommen, die die Stellen der Eltern einnehmen. Dazu wird das Genom der Eltern an der Trennungsstelle in zwei Stücke geteilt und auf die beiden Nachkommen verteilt. Abbildung 3.20

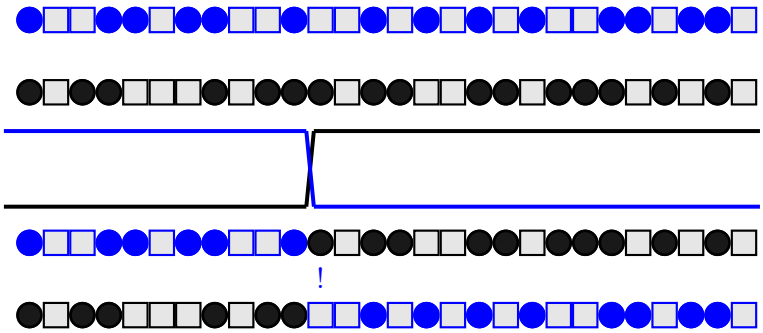


Abbildung 3.20: Beispiel für die Anwendung des Single Point Cross-Overs: Das Ausrufungszeichen markiert die Trennungsstelle

demonstriert, wie das Genom der Nachkommen aus den Genomen der Eltern zusammengesetzt werden. Der Einsatz dieses Operators ist stark abhängig von der Wahl einer Cross-Over-Wahrscheinlichkeit. Übliche Werte für diesen Parameter liegen zwischen 0,45 und 0,95 (Booker, 2000).

3.3.2.3 Two Point Cross-Over

Das Single Point Cross-Over führt in der Anwendung zu Schwierigkeiten: Die Wahrscheinlichkeit dafür, dass nah beieinander liegende Stellen im Genom getrennt werden, ist kleiner als die Wahrscheinlichkeit für weiter entfernt liegende Stellen, da jede Trennungsstelle im Genom gleichwahrscheinlich ist und die Anzahl der möglichen Trennungsstellen bei einem wachsenden Abstand ansteigt. Unter Umständen kann dieser Effekt gewünscht sein, wenn etwa Blöcke von Information gezielt geschützt werden sollen, im allgemeinen führt dies aber zu dem Problem, dass bestimmte gewünschte Weiterentwicklungen sehr lange dauern, weil die Wahrscheinlichkeit für das notwendige Cross-Over-Ereignis zu klein ist. Um dieses Ungleichgewicht zu korrigieren, wurde die Methode des Two-Point-Cross-Overs eingeführt. Hierzu wird das Genom der Eltern an zwei Stellen getrennt und auf die Nachkommen so aufgeteilt, dass jedes neue Genom die beiden Endstücke des Genoms eines Elternteils und den Mittelteil des Genoms des zweiten Elternteils erhält. Hierzu wird eine zweite Trennungsstelle zufällig ermittelt, so dass eine Trennung vollzogen werden kann wie sie in Abbildung 3.21 veranschaulicht ist.

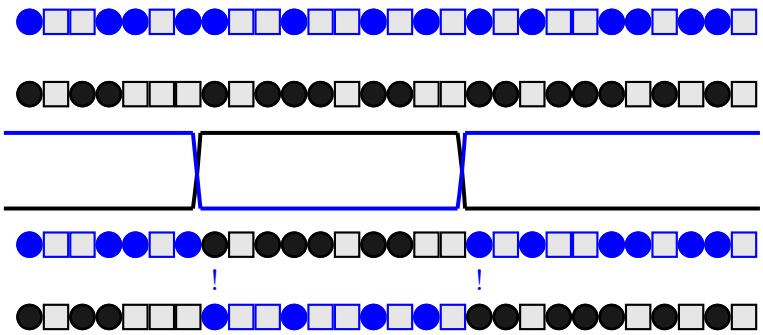


Abbildung 3.21: Beispiel für die Anwendung des Two Point Cross-Overs: Die Ausrufungszeichen markieren die Trennungsstellen

3.3.2.4 Uniform Cross-Over

Sind die so entstehenden Blöcke immer noch zu stark strukturiert für die Lösung eines bestehenden Problems, kann die Anzahl der Trennungsstellen weiter erhöht werden. Radikal wird die Rekombination aber erst dann, wenn die Zuordnung jeder einzelnen Stelle im Genom der Eltern zu den beiden Nachkommen zufällig bestimmt wird, mit der Möglichkeit, einen Parameter p_1 zu setzen, dessen Differenz zu 0,5 den Grad angibt, zu der ein Individuum einem einzelnen Elternteil gleicht. Dabei bestimmt p_1 die Wahrscheinlichkeit, mit der die Information an einer Genomstelle vom ersten Elternteil auf den ersten Nachkommen übertragen wird (vgl. Abbildung 3.22). Die Setzungen 0 oder 1 führen zu einer Kopie der beiden Eltern, ansonsten ergibt sich die Ähnlichkeit aus der Bernoulli-Verteilung.

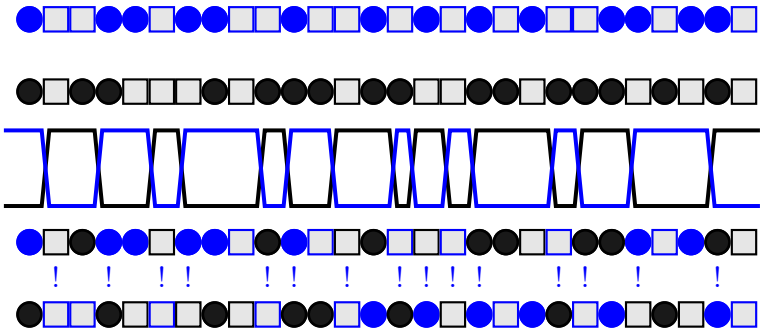


Abbildung 3.22: Beispiel für die Anwendung des Uniform Cross-Overs: Die Ausrufungszeichen markieren die Trennungsstellen

Soll erreicht werden, dass zusammenhängende Blöcke von beiden Elternteilen entstehen,aber nicht in der Form des Point-Cross-Overs kann eine Wechselwahrscheinlichkeit p_k für das *Uniform P-Cross-Over* definiert werden, durch die die Verteilung der Informationen in folgender Weise gesteuert wird: Es hilft die Vorstellung eines Lesegeräts, dass genau zwei Spuren auslesen kann, die den Genomen der Eltern entsprechen. An jeder Stelle wechselt das Lesegerät mit der Wahrscheinlichkeit p_k die Spur. Ist p_k klein gewählt, so entstehen durchschnittlich größere zusammenhängende Stücke als bei einem großen Wert. Abbildung 3.23 zeigt das Ergebnis eines Durchlaufs mit dem Operator Uniform P-Cross-Over.

3.3.2.5 Mehr als zwei Elternteile

Theoretisch ist das Cross-Over nicht auf den Fall von genau zwei Elternteilen beschränkt. So ist es kein Problem, einen Operator zu konstruieren, der die Information von mehr Elternteilen verwaltet. Bei k Elternteilen ist es die einfachste Methode, die Genome an jeweils identischen, aber zufällig ermittelten $k-1$ Stellen in k Teile aufzuteilen. Hier hilft die Anwendung der Struktur von lateinischen Quadraten, siehe Tabelle 3.4 für ein lateinisches Quadrat mit vier Elementen (s. Tabelle 3.4). Die Nachkommen würden aus den neun Stücken so zusammengesetzt, dass jeder Nachkomme Informationen aller k Elternteile erhält. Ein Beispiel für vier Elternteile ist Abbildung 3.24 zu entnehmen.

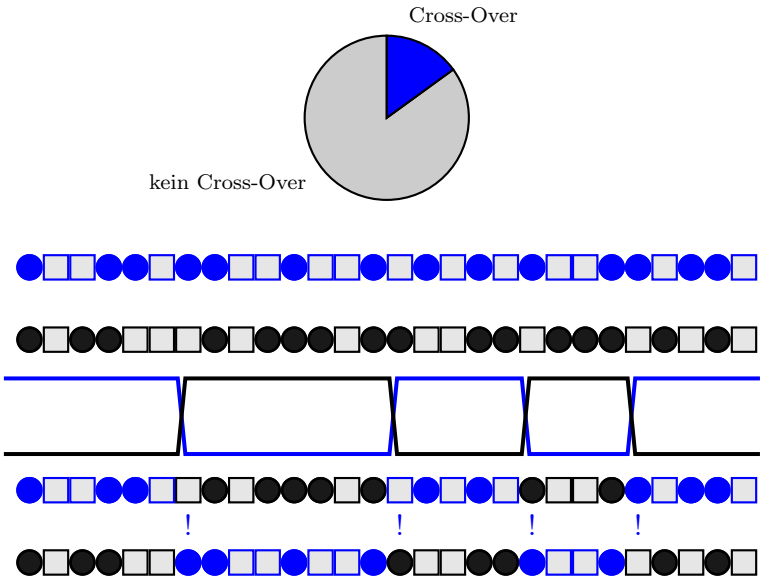


Abbildung 3.23: Beispiel für die Anwendung des Uniform P-Cross-Overs: Die Ausrufungszeichen markieren die Trennungsstellen

Tabelle 3.4: Lateinisches Quadrat für vier Elemente

1:	A	B	C	D
2:	D	A	B	C
3:	C	D	A	B
4:	B	C	D	A

Naheliegend ist auch die Erweiterung auf $2k - 1$ Teile und dementsprechend $2k$ Bestandteile als Generalisierung des Two-Point-Cross-Overs. Das Uniform-Cross-Over ist ebenfalls anwendbar unter Berücksichtigung der Tatsache, dass ab dem zweiten Nachkommen nicht mehr alle Informationen für eine Zufallsauswahl zur Verfügung stehen und erst der letzte Nachkomme durch die Verteilung der ersten $k-1$ determiniert ist. Ebenso kann das Lesegerät mit k Spuren und dem Blockieren von Spuren mit gelesener Information, die einen Wechsel erzwingen, konzipiert werden.

Es ist auch möglich, die Informationen der gesamten Population als Pool zu verstehen, aus dem jeder Nachkomme zusammengestellt wird (Eiben, 2000). Dabei kann eine große Zahl von Elternteilen bspw. über einen Mehrheitsentscheid die Stellen im Genom bestimmen oder auch die Information aller Eltern zu einer Wahrscheinlichkeitsverteilung für die Nachkommen transformiert werden.

3.4 Erstellung von hierarchischen Regelmengen mit genetischen Algorithmen

Genetische Algorithmen werden in dieser Arbeit als Lösungsansatz für das Problem der Regelinduktion vorgestellt. Der Vorteil gegenüber gierigen Algorithmen liegt in der Möglichkeit, Interaktionen von Prädiktoren leichter aufzudecken. Dieser Vorteil konnte für künstlich konstruierte Datensätze demonstriert werden (vgl. Freitas, 2002). Der Vorteil schwindet bei Datensätzen, die keine komplexen Interaktionsmuster aufweisen. Im Folgenden sollen einige Möglichkeiten und Herausforderungen der Implementierung genetischer Algorithmen für die Regelinduktion diskutiert werden: zunächst die Frage der grundsätzlichen Vorgehensweise für die Konstruktion von Regelmengen, dann die Kodierung von Antezedenzbedingungen und Konsequenzen, zuletzt Möglichkeiten der Variablenselektion.

3.4.1 Pittsburgh vs. Michigan: Ein Repräsentationsproblem

Bei der Entwicklung von Regelmengen für die Klassifikation durch genetische Algorithmen wird zwischen dem Pittsburgh- und dem Michigan-Ansatz unterschieden (Freitas, 2002). Im Michigan-Ansatz

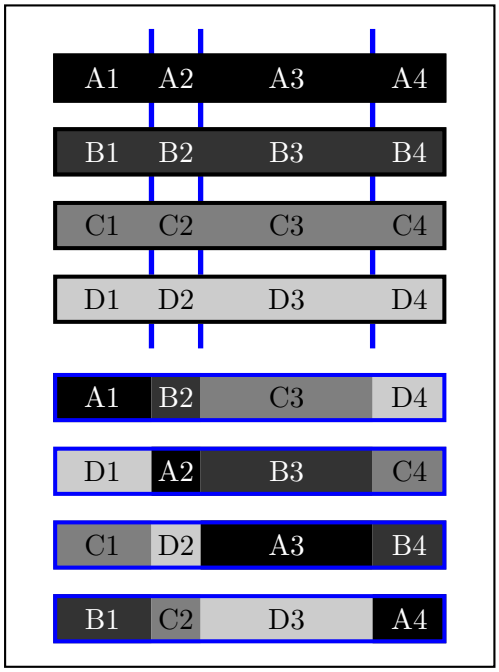


Abbildung 3.24: Mögliches Schema für das Cross-Over mit vier Eltern

repräsentieren Individuen im genetischen Algorithmus einzelne Regeln, im Pittsburgh-Ansatz entspricht jedem Genom eine vollständige Regelmenge. Die an dieser Stelle resultierenden Konsequenzen für die Entwicklung von Regelmengen sind die Folgenden:

Im Pittsburgh-Ansatz ist die Beziehung zwischen Individuen und Lösungen geradlinig. Jedes Individuum stellt eine vollständige Lösung dar. Probleme resultieren allerdings aus der Größe des Suchraums, die sich aus dem Produkt der Suchräume für jede einzelne Regel ergibt. Das verlangsamt die Entwicklung von geeigneten Lösungen und vermehrt unter Umständen die Anzahl lokaler Optima. Weiterhin ist die Bestimmung der Regelanzahl ein fundamentales Problem bei Wahl dieser Vorgehensweise. Im Michigan-Ansatz werden Individuen als Repräsentation einzelner Regeln betrachtet. Für die Aufgabe, von Einzelregeln zu Regelmengen zu gelangen, stehen einige Alternativen zur Verfügung. Wird die Gesamtpopulation als Regelmenge betrachtet, muss sichergestellt werden, dass nicht alle Individuen auf eine Lösung konvergieren. Hier stehen Methoden wie *Fitness Sharing* oder *Niching* zur Verfügung, in denen Individuen für Nähe zu anderen Individuen bestraft werden, mit der Leitidee, dass Individuen an der gleichen Position sich den Fitnessvorteil untereinander teilen müssen.

Andererseits kann die Erstellung von Regeln sequenziell erfolgen, was sich vor allem bei hierarchischen Regelsystemen anbietet, da die Anwendung früherer Regeln nicht durch die Anwendung späterer Regeln beeinflusst wird. Hierzu wird im Anschluss an jede Erstellung einer Regel der Teil der Stichprobe aus der Analyse entfernt, der durch diese Regel klassifiziert werden kann. Die Strategie kann als konditionale Strategie bezeichnet werden, die durch Pfadabhängigkeit gekennzeichnet ist. Abhängig von der Wahl der ersten Regeln verändern sich die Bewertungen späterer Regeln. Während die wiederholte Suche in den meisten Fällen immer noch einen Zeitvorteil gegenüber dem Pittsburgh-Ansatz aufweisen kann, muss die Menge optimaler Regeln nicht die optimale Regelmenge sein, da unter Umständen eine schwächere erste Regel eine bessere Gesamtleistung ermöglichen kann.

Weijters und Paredis (2002) sprechen auch von einem Fokus auf die lokale vs. die globale Ebene, um zwischen den Ansätzen zu differenzieren. Im Rahmen der mit dieser Arbeit einhergehenden Softwareerstellung sind beide Ansätze umgesetzt worden, der Michigan-Ansatz dabei in seiner konditionalen, sequenziellen Form.

3.4.2 Kodierung von Antezedenzbedingungen für Regeln

Die Kodierung von Antezedenzbedingungen kann zunächst einmal in zwei Schritten konzipiert werden. Zunächst wird für jeden Prädiktor das Selektionsintervall durch obere und untere Grenze kodiert, für die jeweils ein Genomabschnitt reserviert wird, der in seiner fixen Länge die erforderliche Präzision zulässt.

Dabei wird definiert, dass Werte in dem Intervall, das durch die Grenzen eingeschlossen wird, zu einer Erfüllung der Antezedenzbedingung führen. Soll das Intervall invertiert werden, wäre ein weiteres Bit zur Kodierung der Interpretation der Grenzen erforderlich (innen vs. außen). Weiterhin entsteht aus einer fixen Positionierung der Intervallgrenzen eine genetische Nähe zwischen angrenzenden Prädiktoren in dem Sinne, dass bei einem Single-Point-Crossover die Wahrscheinlichkeit größer wird, dass die Werte für zwei angrenzende Prädiktoren mit höherer Wahrscheinlichkeit an das gleiche Individuum vererbt werden als die Werte zweier entfernter Prädiktoren. Um generell das Problem der positionalen Abhängigkeit der Wirkweise von genetischen Operatoren anzugehen, ist unter anderem das so genannte *Messy Coding* eingeführt worden, bei dem die Kodierung von Werten und die Zuordnung von Werten zu Prädiktoren getrennt voneinander kodiert würden. In der vorliegenden Arbeit ist das Problem durch die Bereitstellung von genetischen Operatoren gelöst worden, deren Ergebnisse durch die beschriebene genetische Nähe nicht beeinflusst werden. Zuletzt lässt sich auch ein Teil des Genoms als Aktivierungsgenom verstehen, in dem für jeden Prädiktor binär kodiert wird, ob er für die Prüfung der Antezedenzbedingung berücksichtigt werden soll oder nicht. Mit einem solchen Genomteil kann so das Problem der Variablenselektion gelöst werden.

Die vorliegende Software stellt die Möglichkeit der Verwendung eines Aktivierungsgenom zur Verfügung, arbeitet aber mit einer festen Länge für Genome, Irrelevanz muss also bei Verzicht auf das Aktivierungsgenom durch die Grenzen des Intervalls erzeugt werden.

3.4.3 Kodierung von Prädiktionswerten für Regeln

Durch die Antezedenzbedingungen wird die Menge von Objekten definiert, die von einer Regel klassifiziert werden. Für die Entscheidung, wie diese Objekte klassifiziert werden, gibt es prinzipiell mehrere Vorgehensweisen.

Der naheliegende Weg besteht aus einer direkten Kodierung in das Genom als zusätzliches Bit. Damit würde das Regelergebnis ebenfalls den Kräften der künstlichen Selektion unterworfen. Freitas (2002) argumentiert gegen diese Vorgehensweise, da die Beziehung zwischen Konsequenz und Antezedenzbedingung sehr stark ist: Die Mutation der Konsequenz nach vielen Generationen würde die Qualität einer Lösung invertieren. Der Nutzen wäre in frühen Generationen vorhanden, würde aber in späteren Generationen durch die Gefahr der Zerstörung aufgewogen werden. Ein alternatives Vorgehen legt die Regelkonsequenz nicht genetisch fest, sondern lediglich das Entscheidungsprinzip: Eine Regel wird immer so entschieden, dass ein wählbares Qualitätskriterium maximiert wird. Das kann bedeuten, dass immer so entschieden wird, dass die Mehrheit der Fälle richtig klassifiziert wird, oder das ein Kostenkriterium durch die Entscheidung minimiert bzw. ein Nutzenkriterium maximiert wird. Im Allgemeinen wird die Konsequenz nach Beobachten der Zusammensetzung der durch die Regel zu klassifizierenden Objektmenge fixiert.

Weiterhin wäre es möglich, die gleiche Konsequenz für die gesamte Population zu verwenden, damit würden durch das Regelsystem nicht klassifizierte Fälle automatisch der verbleibenden Kategorie zugeordnet. Verwendet man eine Speziesregel, so werden die Cross-Over-Paarungen auf Individuen mit gleicher Regelkonsequenz eingeschränkt, so dass sich effektiv zwei Subpopulationen bilden, die getrennt voneinander evolvieren (Freitas, 2002). In der vorliegenden Software werden mehrere der vorgeschlagenen Vorgehensweisen mit Ausnahme der Evolution der Konsequenz als Alternative angeboten.

3.4.4 Spezielle Operatoren für die Regelinduktion

Für die Evolution von Regeln bieten sich weiterhin einige spezielle genetische Operatoren an. Freitas (2002) diskutiert das Prinzip der Generalisierung und Spezifizierung von Regeln durch Cross-Over und Mutation. Bei der Mutation besteht eine Spezialisierung aus der Erhöhung der Restriktivität der Antezedenzbedingung, d.h. aus einer Verkleinerung von einem oder mehreren Intervallen. Die Generalisierung besteht dementsprechend aus einer Relaxierung der Antezedenzbedingung und demnach aus einer Vergrößerung von einem oder mehreren Intervallen. Praktisch ist fast jede Wertveränderung entweder eine Generalisierung oder eine Spezialisierung, ein gezielter Mutationsoperator könnte aber ein Feintuning von Lösungen ermöglichen, indem gezielte Manipulationen an den Werten einzelner Prädiktoren vorgenommen werden.

Bei einem Generalisierungs-Cross-Over würden die neuen Individuen nicht zufällig aus den Genomteilen der Eltern gebildet: Statt dessen werden bei jeder Wahl zwischen zwei Genomteilen jene Teile gewählt, die

zu einer höheren Generalisierung führen. Wird beispielsweise eine untere Schranke für einen Prädiktorwert vererbt, so würden beide Kinder den kleineren der beiden möglichen Werte erben, bei einer oberen Schranke den größeren. Bei einem Spezialisierungs-Cross-Over wäre das Vorgehen dementsprechend umgekehrt, hier werden die Nachkommen maximal spezialisiert. Es wird deutlich, dass diese Operatoren nicht als einzige Operatoren Verwendung finden sollten, sondern lediglich als Ergänzung dienen können. Sonst würde entweder eine Population von Regeln erzeugt werden, die jedes Objekt klassifizieren, oder eine Population, die kein Objekt mehr klassifizieren kann. Verschiedene Formen regelspezifischer Operatoren sind auch in die Software integriert worden.

3.4.5 Auswahl der Prädiktoren

Zentral für die Generalisierbarkeit von Lösungen ist neben der Verfügbarkeit einer repräsentativen und genügend großen Stichprobe von zu klassifizierenden Objekten auch eine vernünftige Entscheidung für die Menge der eingesetzten Prädiktoren. Freitas (2002) unterscheidet zwei Auswahlprinzipien voneinander, namentlich *Filter*- und *Wrapper*-Ansätze:

Bei einem Filter-Ansatz wird die gesamte Menge aller Attribute in das Data Mining-Verfahren aufgenommen. Das Verfahren selbst filtert die Teilmenge benötigter Attribute heraus. Die Verwendung von Aktivierungsgenen entspricht zum Beispiel dem Filter-Ansatz: Alle Attribute stehen potenziell zur Verfügung, durch Setzen der Bits wird bestimmt, welche Teilmenge für die Klassifikation relevant wird. Andere Filter-Ansätze minimieren Kriterien wie Inkonsistenz durch Auswahl von Attributteilmengen oder spezifizieren Einschluss- oder Ausschlusskriterien für Prädiktoren wie Verfahren der schrittweisen Regression oder Diskriminanzanalyse. Bei einem Wrapper-Ansatz wird der zur Verfügung stehende Trainingsdatensatz in zwei Teile getrennt, einen Trainings-Trainings-Teil und einen Trainings-Test-Teil[2]. Jede mögliche Teilmenge von Attributen wird jetzt in dem Sinne überprüft, dass anhand des Trainings-Trainings-Datensatzes ein Modell konstruiert wird, dessen Qualität am Trainings-Test-Datensatz bestimmt werden kann. Die in dieser Hinsicht stärkste Lösung wird als endgültige Attributsmenge dann in das Klassifikationsverfahren eingebettet, das am Test-Datensatz geprüft werden kann.

Der Wrapper-Ansatz ist deutlich langsamer als der Filter-Ansatz. Das Ungleichgewicht wird gravierender, je größer die Menge der Attribute ist, da die Anzahl der möglichen Teilmengen mit der Zahl der Attribute exponentiell wächst. Gleichzeitig wird der Ansatz oft als robuster gesehen als der Filter-Ansatz. Bei begrenzten Ressourcen wird es oft zu einer empirischen Frage, welchem Ansatz der Vorzug gegeben werden soll.

In der vorliegenden Software ist mit dem Aktivierungsgenom die Möglichkeit für einen Filter-Ansatz eingerichtet worden. Für die Erstellung von Downhill-Modellen wird hingegen auch ein Wrapper-Ansatz angeboten, der nach Angabe von Mindest- und Maximalvariablenzahl alle möglichen Teilmengen der Attribute mit diesen Spezifikationen auf ihre Leistungsfähigkeit untersucht und die beste Auswahl bestimmt.

3.5 Genetische Algorithmen und Einschränkungen im Lösungsraum

Im idealen Anwendungsfall entspricht jede mögliche Ausprägung im virtuellen Genom von Individuen einer möglichen Lösung für das Zielproblem, allerdings gibt es mehrere Gründe, warum künstliche Genome auftreten können, die keiner akzeptablen Lösung entsprechen. Zu nennen wären hier unzulässige Werte, Restriktionen und Besonderheiten des Problemraums. Es soll zunächst ein kurzer Überblick über Probleme der beschriebenen Art gegeben werden, um im Anschluss einige übliche Lösungsmöglichkeiten zu skizzieren.

3.5.1 Das Problem unzulässiger Werte und Restriktionen

Ein Genom kodiert im Prinzip beliebige numerische oder symbolische Werte, deren Beziehung zu empirischen Handlungen oder Gegebenheiten entscheidend von der Repräsentation des Problems und der gewählten Kodierung abhängt. Die genetischen Operatoren sind zunächst blind gegenüber Werteberei-

[2] Der Testdatensatz wird für die Bestimmung der Qualität des konstruierten Klassifikationsalgorithmus verwendet, deshalb darf er nicht für die Entscheidung über Teilkomponenten des Algorithmus herangezogen werden. Freitas (2002) bezeichnet ein derartiges Vorgehen als „Todsünde“.

chen oder Voraussetzungsverletzungen, da Genome syntaktisch und nicht semantisch bearbeitet werden. Das Genom eines Individuums kodiert eine unzulässige Lösung, wenn

- eine grundsätzlich fehlerhafte Repräsentation des Problems oder eine grundsätzlich fehlerhafte Kodierung gewählt wurde;
- Variablen Werte außerhalb ihres unzulässigen Wertebereichs zugewiesen werden;
- Abhängigkeiten zwischen Variablen verletzt werden;
- die rekonstruierte Lösung keine Lösung für das Ausgangsproblem darstellt.

In vielen Problemen lässt sich ein Suchraum als Körper in einem mehrdimensionalen Raum konzipieren, der durch die zulässigen Ausprägungen des Genoms aufgespannt wird. Dabei entspricht jedem Parameterwert eine Dimension in diesem Suchraum, und die möglichen Parameterwerte entsprechen der Ausdehnung des Körpers in dieser Dimension. Enthält ein Suchraum Wertbelegungen für Parameter, die den zulässigen Wertebereich verlassen, so gibt es einen Unterraum im Lösungsraum, der unzulässige Lösungen enthält.

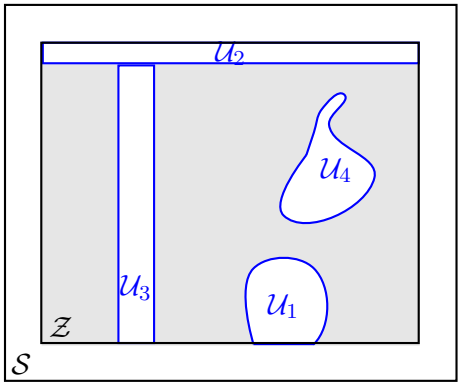


Abbildung 3.25: Beispiel für einen Suchraum ($\mathcal{S}$), der sich aus einem Unterraum mit zulässigen Lösungen($\mathcal{Z}$) und einem Unterraum mit unzulässigen Lösungen ($\mathcal{U} = \mathcal{U}_1 \cup \mathcal{U}_2 \cup \mathcal{U}_3 \cup \mathcal{U}_4$) zusammensetzt

In ähnlicher Weise tritt das Problem unzulässiger Lösungen auf, wenn Abhängigkeiten zwischen Variablen verletzt werden, wenn also zum Beispiel ein höherer Wert für eine obere Grenze kodiert wird als für eine untere Grenze oder Restriktionen bezüglich von Produktionsmitteln verletzt werden. Da die möglichen Symbole eines Genoms diskrete Werte annehmen, gibt es abzählbar viele Lösungen, die ein Genom kodieren kann, die Menge dieser Lösungen soll trotzdem als Suchraum bezeichnet werden, der als Menge $\mathcal{S}$ beschrieben werden soll. Der Suchraum setzt sich nun zusammen aus der Menge der zulässigen Lösungen $\mathcal{Z}$ und der Menge der unzulässigen Lösungen $\mathcal{U}$ (vgl. Michalewicz & Fogel, 2000). Das Problem wird verschärft, wenn die beiden Teilmengen nicht aus benachbarten Lösungen bestehen, sondern der Suchraum mehrere Partitionen aufweist, die aus zulässigen und unzulässigen Lösungen bestehen, wie im Beispiel in Abbildung 3.25. Zum Umgang mit Genomen, die eine Lösung l kodieren mit $l \in \mathcal{U}$, werden verschiedene Möglichkeiten diskutiert.

3.5.2 Bestrafung und Belohnung als Lösungsansatz

Ein Bestrafungsansatz läuft auf die differenzielle Bewertung von Lösungen hinaus, die sich in der Menge zulässiger Lösungen und in der Menge unzulässiger Lösungen befinden, so dass mit

$$o_z : \mathcal{Z} \rightarrow \mathbb{R} \tag{3.32}$$

und

$$o_u : \mathcal{U} \rightarrow \mathbb{R} \tag{3.33}$$

zwei unterschiedliche Bewertungsfunktionen definiert werden (Michalewicz, 2000a). Dabei definiert ein Bestrafungsansatz die Beziehung zwischen den Funktionswerten für eine Lösung l in der Form

$$o_u(l) = o_z(l) - p(l). \tag{3.34}$$

Die einfachste Lösungsmöglichkeit zum Umgang mit unzulässigen Lösungen ist die maximale Bestrafung von Individuen, die unzulässige Lösungen anbieten. Dies kann erreicht werden durch die Vergabe von Fitnesswerten, die sicher stellen, dass keine Nachkommen an die Folgegeneration abgegeben werden können. Michalewicz und Fogel (2000) bezeichnen dieses Vorgehen als „death penalty"-Heuristik. Diese Heuristik ist sinnvoll, wenn der unzulässige Bereich zusammenhängend und am Rand des zulässigen Bereichs liegt, wie etwa $\mathcal{U}_2$ im Beispiel.

Eine flexiblere und in anderen Situationen mehr versprechende Alternative ist die Verwendung von Belohnungs- oder Penaltyfunktionen. Penaltyfunktionen bestrafen die Übertretung erlaubter Parameterwertgrenzen oder die Verletzung von Restriktionen. Ist das temporäre Tolerieren unzulässiger Lösungen notwendig, um Brücken zwischen unzusammenhängenden Bereichen des Raums zulässiger Lösungen zu ermöglichen, stellt die Subtraktion einer Konstante die einfachste Art der Bestrafung dar. Wird die Konstante sehr hoch gewählt, entspricht die Vorgehensweise der oben genannten Heuristik. Ist das Ausmaß der Überschreitung quantifizierbar, kann ein alternativer Bestrafungsmechanismus zum Einsatz kommen. Quantifizierbar ist eine Überschreitung beispielsweise durch Abzählen der verletzten Wertebereiche oder Restriktionen (Eiben & Ruttkay, 2000), aber auch durch Angabe der Distanz der Lösung von der nächstgelegenen zulässigen Lösung bei Verwendung eines angemessenen Distanzmaßes. Liegt ein Überschreitungsmaß s vor, so kann bspw. mit einfachen Cut-Offs gearbeitet werden, so dass eine Überschreitung geringeren Ausmaßes toleriert wird, oder es kann eine Funktion für den Zusammenhang zwischen Überschreitung und Bestrafung angegeben werden, beispielsweise in der linearen Form:

$$p = a_1 \cdot s + a_0 \tag{3.35}$$

In diesem Fall wird mit a_0 ein Grundbestrafungswert bei Überschreitung festgelegt, der mit der Höhe der Überschreitung linear anwächst,gesteuert durch den Koeffizienten a_1. Diese Lösung könnte erneut mit einem oder zwei Cut-Off-Werten kombiniert werden, so dass eine leichte Überschreitung immer toleriert wird und eine schwere Überschreitung immer zum Ausschluss führt. Auch andere funktionale Zusammenhänge sind denkbar. Die Bestrafung kann zudem adaptiv erfolgen in dem Sinne, dass die Funktion mit einem Koeffizienten gewichtet wird, dessen Wert in einer Generation von dem Auftreten unzulässiger Lösungen in Vorgängergenerationen abhängt. Damit werden Überschreitungen schwächer gewichtet, wenn zulässige Lösungen beobachtet wurden, und stärker, wenn die besten Lösungen unzulässig waren. So wird eine Exploration von Randbereichen unterstützt, eine Exploration außerhalb der zulässigen Bereiche gehemmt (Smith & Coit, 2000).

Logischerweise kann eine Bestrafung nur dann erfolgen, wenn Lösungen außerhalb der zulässigen Grenzen überhaupt bewertbar sind. Führt eine Überschreitung der Grenzen zu unzulässigen Lösungen in dem Sinne, dass die Lösung unsinnig und damit semantisch leer ist, ist eine Bestrafung keine geeignete Methode der Behandlung dieser Lösung. Die Wahl der Methode muss die Problemstruktur adäquat berücksichtigen. Michalewicz und Fogel (2000) sehen eine Abhängigkeit von dem Verhältnis zwischen zulässigem Lösungsraum und Suchraum, den topologischen Eigenschaften des zulässigen Suchraums, der Art der Bewertungsfunktion, der Anzahl der Variablen, der Anzahl von Einschränkungen, der Art der Einschränkungen und der Anzahl aktiver Einschränkungen an der Stelle des Optimums. Damit ist der Einsatz der Bestrafungsfunktion als nichttrivial zu bewerten.

In Ergänzung oder an Stelle einer Bestrafung kann - nicht ungleich den Empfehlungen der Lerntheorie - auch mit Belohnungen gearbeitet werden. Gibt es bspw. eine Quantifizierbarkeit des Abstands von der nächstgelegenen unzulässigen Lösung, so kann mit diesem Abstand eine Belohnung gewichtet werden, um Individuen in Abhängigkeit von der Gefahr zu belohnen, dass Nachkommen unzulässige Lösungen kodieren. Ist der zulässige Lösungsraum sehr klein, kann die Suche so auf relevante Stellen des Suchraums konzentriert werden. Auch hier sind lineare oder andere funktionale Zusammenhänge und Verbindungen mit Cut-Off-Werten denkbar. Belohnung und Bestrafung könnten zudem in Kombination verwendet werden, wobei die genannten Abhängigkeiten auch bei der Konstruktion von Belohnungsfunktionen berücksichtigt werden müssen.

Smith und Coit (2000) unterscheiden zwischen „exterior penalty functions" und „interior penalty functions", je nachdem ob sich die Bestrafung auf Lösungen innerhalb oder außerhalb des zulässigen Raums erstreckt. Diese Unterscheidung deckt sich mit der hier getroffenen Aufteilung in Belohnung und Bestrafung, wenn man die Belohnungsfunktion invertiert.

3.5.3 Reparaturmechanismen als Lösungsansatz

Eine Alternative zur Bestrafung stellt die *Reparatur* von Lösungen an unzulässigen Stellen im Suchraum dar. Mit Reparatur ist eine Transformation des Genoms gemeint, die als Ergebnis eine zulässige Lösung produziert. Ein Reparaturmechanismus kann verschiedene Formen annehmen:

Im einfachsten Fall wird das Genom vollständig gelöscht und durch ein neu erstelltes Genom ersetzt, das eine zulässige Lösung darstellt, auch wenn hier der Begriff „Reparatur“ etwas gedehnt wird. Ist die Topologie des zulässigen Raums bekannt und ist im Suchraum eine Distanzmetrik sinnvoll einsetzbar, so kann eine unzulässige Lösung durch die ihr am nächsten liegende zulässige Lösung ersetzt und in dieser Art und Weise „repariert“ werden. Ist die Art oder die Richtung der notwendigen Reparaturoperation hingegen unbekannt, so können Werkzeuge bereit gestellt werden, die in zufälliger Reihenfolge oder mit zufällig gewählten Parametern appliziert werden, um gegebenenfalls mit einer zeitlichen Begrenzung eine Reparatur zu versuchen. Ist zumindest die Art der Abweichung von zulässigen Lösungen bekannt, so kann im Einzelfall die Reparatur sogar sichergestellt werden, zum Beispiel dann, wenn es eine triviale zulässige Lösung gibt, zu der eine unzulässige Lösung schrittweise transformiert werden kann.

Michalewicz (1996) setzt in GENOCOP III eine Reparaturmethode um, die mit zwei Populationen operiert. In einer ersten Population sind ausschließlich zulässige Lösungen als Referenzpunkte vertreten, die somit durch die Bewertungsfunktion evaluiert werden können, in der zweiten Population können unzulässige Lösungen auftreten. Diese werden in der Weise repariert, dass zunächst ein zulässiger Referenzpunkt aus der ersten Population zufällig ausgewählt wird. Im Anschluss werden Lösungen auf Zulässigkeit geprüft, die zufällig aus der Menge der Lösungen ausgewählt werden, die auf der Verbindungslinie im Suchraum zwischen unzulässiger Lösung und Referenzpunkt liegen. Die erste zulässige Lösung, die bei diesem Prozess gefunden wird, ersetzt die unzulässige Lösung. Andere Formen der Reparatur leiden unter der Abhängigkeit von der Problemstellung, so dass eine Vielzahl von Reparaturmechanismen für unterschiedliche Probleme erstellt wurden (Michalewicz & Fogel, 2000). Da ähnliche Probleme bei der Wahl der Parameterwerte entstehen, stellt Michalewicz (2000b) unter anderem eine Methode vor, bei der durch Segregierung von Populationen mit unterschiedlichen Parametersetzungen eine Ausbalancierung erreicht wird.

Eine interessante Alternative stellt die Verwendung von Dekodern da. Dabei wird der ursprüngliche Suchraum $\tilde{\mathcal{S}}$, der durch die virtuellen Gene aufgespannt wird, durch einen Dekoder auf den zulässigen Raum der Lösungen $\mathcal{Z}$ abgebildet, also $\tilde{\mathcal{S}} \rightarrow \mathcal{Z}$. Dabei müssen Dimensionalität und Struktur des genetischen Suchraums und des Suchraums für Lösungen $\mathcal{S}$ nicht identisch sein. Durch dimensionsreduzierende Verfahren könnte so etwa ein genetischer Suchraum geringerer Dimensionalität konstruiert werden. So könnten hier selbstorganisierende topologische Karten (Kohonen, 1982a, 1982b) oder Vektorquantisierungsverfahren, aber auch Varianten der Hauptkomponentenanalyse (Jackson, 1991) oder der Clusteranalyse (vgl. Everitt, 2001) zum Einsatz kommen.

Es gibt bei einigen Reparaturmechanismen eine Nähe zum so genannten *Baldwin-Effekt*, der auf eine Kombination von individuellem Lernen und genetisch gestützter Evolution verweist (Michalewicz & Fogel, 2000; Mayley, 1996): Baldwin-Effekte treten dann auf, wenn ein Organismus einen Fitnessvorteil durch ein differenziell verteiltes Lernvermögen gewinnt. Proximal sind deshalb die Lernvorgänge verantwortlich für seine höhere Fitness, die aber in ihrer Struktur in der genetischen Ausstattung angelegt sind. Eine klare Trennung von ontogenetischen und phylogenetischen Anteilen ist hier nicht zu leisten. Die Einführung von Reparaturmechanismen begünstigt in ähnlicher Art und Weise eine Menge von Lösungen, die in ihrer genetischen Grundform eine Zieleigenschaft nicht aufweisen, diese aber in Interaktion mit einer Reparaturregel in einem virtuellen Sinne „entwickeln“ können:

> The introduction of individual learning can radically alter fitness landscapes. This is especially true if the learning algorithm operates on phenotypes according to a fundamentally different process. (Anderson, 2000, S. 317)

Dabei muss beachtet werden, dass Lernprozesse auch einen Schutz für Genome mit genetisch bedingt niedriger Fitness bedeuten können, was zu Vorteilen in den frühen Phasen eines Algorithmusdurchlaufs führt, aber die genetische Konvergenz in späteren Phasen auch beeinträchtigen kann. Baldwin-Effekte sind auch in natürlichen Umgebungen beobachtet wurden und reichen bereits aus, um einen naiven Determinismus-Glauben an die vollständige Determination des Organismus durch seinen genetischen Code auszuhebeln, da die Berücksichtigung von Umwelt und Interaktionen unablässig wird. Man muss

diese Art von Interaktion zwischen Evolution und Lernen abgrenzen von einem anderen Prozess, der so genannten Lamarckschen Vererbung von Gelerntem an Nachkommen:

> Lamarckian inheritance is rejected as a biological mechanism under the modern synthesis, since it is difficult to envision a process by which acquired information can be transferred into the gametes. Nevertheless, the practical utility of Lamarckian algorithms has been demonstrated in some evolutionary optimization application. (Anderson, 2000, S. 308)

Es kann erneut festgehalten werden, dass die biologische Plausibilität der organismischen Entsprechung von Bestandteilen virtueller Algorithmen im Bereich des Data Mining eine untergeordnete Rolle spielt. Umgekehrt ermöglicht die Einführung Lamarckscher Operatoren die Exploration einer virtuellen Evolution, die nicht an die möglicherweise kontingenten oder suboptimalen Eigenschaften beobachtbarer Evolution gebunden ist.

3.6 Der Partikelschwarm und Optimierungsaufgaben

In diesem Abschnitt sollen Partikelschwärme nach dem Vorbild der Schwärme von Kennedy und Eberhart (1995, 2001) für die Lösung von Problemen besprochen werden, die mit reellen Zahlen kodiert werden können. Schwärme können als Agentenmodell beschrieben werden und gehören damit zum Bereich des Künstlichen Lebens. Grundprinzip ist ein Suchverhalten, das über randomisierte Bewegungen und Kommunikation zwischen parallelen Suchprozessen funktioniert. Die Mischung aus Unabhängigkeit von Schwarmteilen und Parallelität mit Koordination und Konvergenz ermöglicht dabei eine effiziente Suche auf der Basis des Gesamtschwarms.

In der Darstellung sollen zunächst die Bestandteile der Schwarmmetapher erläutert werden, also Suchraum, Partikelstruktur und die Kommunikation zwischen Partikeln. Im Anschluss werden der Algorithmusablauf und seine Einzelschritte präsentiert.

3.6.1 Komponenten des Partikelschwarmalgorithmus

3.6.1.1 Der Suchraum

Der Suchraum wird aufgespannt durch orthogonale Achsen, die jeweils einem der unabhängig voneinander zu schätzenden Parameterwerte entsprechen. So wird ein Problem mit sechs Parametern zu einer Suche im sechsdimensionalen Raum. Jeder Punkt in diesem Raum entspricht deshalb einer möglichen Lösung. Die Kodierung ist deshalb einfach, lediglich die Eindeutigkeit der Zuordnung von Parametern zu Achsen muss sichergestellt werden. Sind Parameter durch einfache Beziehungen verbunden, die bspw. den Wert eines der Parameter aus den Werten der anderen Parameter eindeutig bestimmen lassen, dann muss dieser letzte Parameter nicht im Suchraum berücksichtigt werden. Im allgemeinen wird der Suchraum kontinuierlich konzipiert. Kennedy und Eberhart (2001) diskutieren Modelle für binäre Modelle, in dieser Arbeit wird später ein Modell für Variablen mit diskreten Werten vorgestellt. Die Dimensionen sind zudem unbegrenzt, weshalb für Anwendungen, die Restriktionen für Parameterwerte aufweisen, Modifikationen des Modells erforderlich werden.

3.6.1.2 Die Partikel

Ein Partikel stellt einen sehr einfachen Agenten da. Der Agent wird mit lediglich vier Eigenschaften versehen. Die erste Eigenschaft ist seine Position im Suchraum, die durch Angabe eines Vektors in $\mathcal{R}^n$ definiert wird, die zweite Variable ist seine Geschwindigkeit, die durch einen Vektor gleicher Dimensionalität repräsentiert wird. Die Geschwindigkeit ist die Bewegung, die ein Partikel bei Ausbleiben von Beschleunigung in einem Zeitschritt ausführen würde. Die anderen beiden Eigenschaften sind Speicherstellen für Positionen. Zum einen wird für jeden Partikel die Stelle im Raum festgehalten, die im Lauf der Suche des Partikels bisher das beste Ergebnis erbrachte, zum anderen wird jene Stelle gespeichert, die für alle Nachbarn des Partikels (s. den folgenden Abschnitt) das bisher beste Ergebnis erbrachte. Bildlich gesprochen erhält jeder Agent ein Gedächtnis für die vorläufig beste eigene Lösung und ein Gedächtnis für die vorläufig beste Lösung, die von seinen Nachbarn kommuniziert wurde. Damit sind die vier Eigenschaften definiert, deren Wert alle durch Vektoren in $\mathcal{R}^n$ angegeben werden (siehe auch Abbildung 3.26).

Bei einer bildlichen Betrachtung sind Position und Geschwindigkeit externale Variablen, da sie von außen wahrgenommen werden können, die beiden Gedächtnisstellen hingegen sind internale Variablen, die lediglich über Kommunikationsprozesse vermittelt werden können.

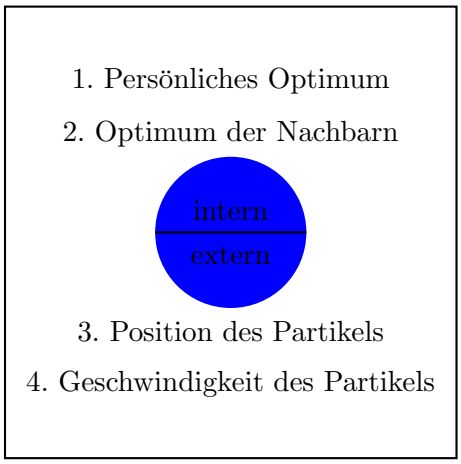


Abbildung 3.26: Interne und externe Variablen der Schwarmpartikel

Diese einfache Struktur genügt bereits, um ein komplexes Suchverhalten zu generieren, das zu einer effizienten Abdeckung des Suchraums führt.

3.6.1.3 Nachbarschaft und Kommunikation

Der Kommunikationsprozess zwischen Nachbarn ist sehr eingeschränkt, es kann lediglich eine Raumposition einseitig und fehlerfrei übermittelt werden. Nachbarschaft ist eine zweistellige Relation: Zwei

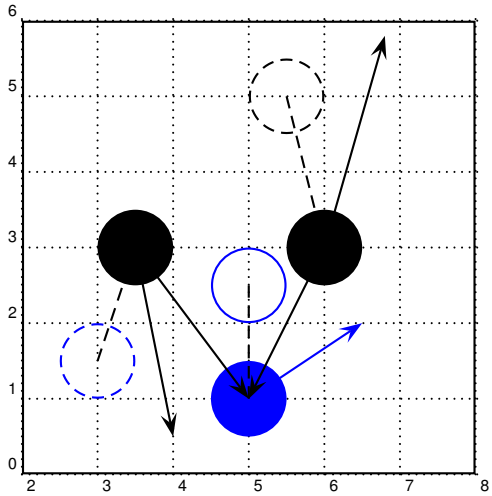


Abbildung 3.27: Repräsentation der Variablenwerte eines Partikels im zweidimensionalen Raum: Zu dem Partikel auf Position $(5; 1)$ gehören die Variablen Position mit dem Wert $(5; 1)$, Geschwindigkeit mit dem Wert $(1,5; 1)$, persönliches Optimum mit dem Wert $(5; 2,5)$ und bestes persönliches Optimum der Nachbarn mit dem Wert $(3; 1,5)$.

Partikel heißen genau dann benachbart, wenn sie in einer Kommunikationsbeziehung stehen, also wenn der erste Partikel mit dem zweiten Partikel kommuniziert. Die Relation ist nicht-reflexiv und nicht notwendigerweise symmetrisch. Nachbarschaft ist zentral für die Funktionsweise des Partikelschwarms. Jeder Partikel ist in der Lage, sein bisher bestes Suchergebnis den Nachbarn mitzuteilen, und sowohl sein eigenes Gedächtnis als auch das vermittelte Gedächtnis beeinflussen die Beschleunigungskräfte, die auf ihn ausgeübt werden. Dementsprechend bestimmt die Nachbarschaftsstruktur mit, welche Agenten in ähnlichen Regionen des Raums suchen werden. Der genaue Algorithmus wird weiter unten beschrieben. Die vier Variablenwerte eines Partikels mit zwei Nachbarn in einem Schwarmausschnitt ist in Abbildung 3.27 visualisiert. Jeder der drei Partikel ist mit seiner Position, seinem Geschwindigkeitsvektor und seinem persönlichen Optimum dargestellt. Es wird hier davon ausgegangen, dass das persönliche Optimum des Nachbarn auf Position $(3,5; 3)$ eine bessere Lösung repräsentiert als das Optimum des zweiten Nachbarn.

3.6.1.4 Schematischer Ablauf der Partikelschwarmoptimierung

Der Verlauf der Partikelschwarmoptimierung ist in Abbildung 3.28 zusammengefasst. Nach einer Initia-

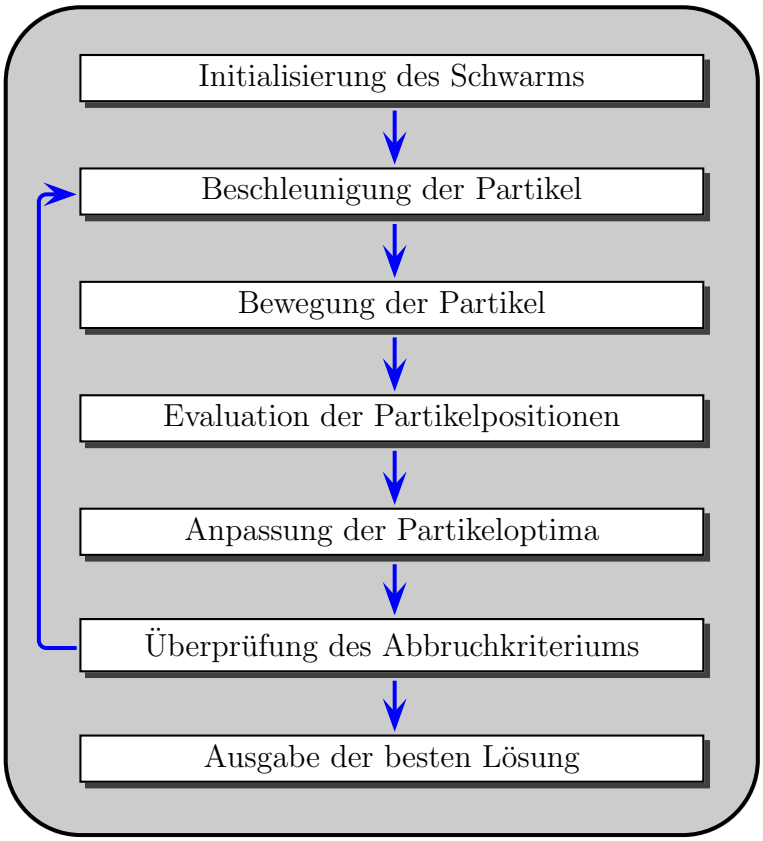


Abbildung 3.28: Abfolge der Partikelschwarmoptimierung

lisierung des Schwarms wird eine Schleife aus Beschleunigung, Bewegung, Bewertung und Anpassung durchlaufen, bis das Abbruchkriterium erfüllt ist. Dabei ist auch die Möglichkeit der differenziellen Festlegung von Topologie und weiteren Parametern der Teilschritte erneut eine Vielzahl von Algorithmen denkbar, die diesem Schema genügen. „Partikelschwarmoptimierung“ bezeichnet dementsprechend - wie der Begriff „Genetischer Algorithmus“ - eine Klasse von Algorithmen und nicht einen einzelnen, vollständig definierten Algorithmus. Im folgenden sollen die Einzelschritte mit ihren Wahlmöglichkeiten vorgestellt werden.

Zur Initialisierung des Partikelschwarms ist zunächst die Festlegung der Partikelanzahl erforderlich. Hierfür gibt es kaum Ansatzpunkte außer Feststellungen, dass weniger Partikel benötigt werden als Individuen in Genetischen Algorithmen. In einem zweiten Schritt müssen die Anfangspositionen und Anfangsgeschwindigkeiten der Partikel festgelegt werden, in einem dritten Schritt die Kommunikationstopologie. Schließlich wird ein Evaluationsschritt eingefügt, um die Initialisierung des Schwarms abzuschließen.

3.6.2 Initialisierung des Partikelschwarms

Für die Bestimmung der Anfangspositionen von Schwarmpartikeln gelten ähnliche Regeln wie für die Bestimmung der Anfangsindividuen in Genetischen Algorithmen. Folgende Möglichkeiten können zum Einsatz kommen:

- Randomisierte Erzeugung: Für jede Dimension jedes Partikels wird ein zufälliger Wert bestimmt. Die Verteilung für die Bestimmung kann von den spezifischen Eigenschaften der Dimensionen abhängig gemacht werden.

- Systematische Auswahl der ersten Positionen: Die Partikel werden systematisch über den Suchraum verteilt. Auch hier kann die Systematik die spezifischen Eigenschaften der Dimensionen oder die spezifische Kommunikationstopologie (s.u.) berücksichtigen.

- Empiriegestützte Auswahl der ersten Positionen: Die Partikel werden aufgrund empirischer Ergebnisse und Vorstudien im Suchraum platziert. Zu diesem Zweck ist Expertenwissen notwendig.
- Bestimmung der ersten Positionen durch einen Hilfsalgorithmus: Die ersten Positionen werden als Ergebnis eines oder mehrerer Hilfsalgorithmen festgelegt. Diese Alternative kann relevant sein, wenn Hilfsalgorithmen anfällig sind für lokale Minima und Räume zwischen lokalen Minima untersucht werden sollen oder auch in dem Fall, wenn Hilfsalgorithmen wenig präzise, aber schnell arbeiten.

Die Auswahl der Vorgehensweise liegt im Ermessen des Anwenders. Wird Vorinformation genutzt, kann das die Laufzeit des Algorithmus reduzieren, die Eingabe suboptimaler Positionen kann den Algorithmus unter Umständen frühzeitig in Regionen des Suchraums binden und die Suche nach besseren Positionen verhindern. Durch die Verwendung anderer Algorithmen geht die von diesen ermittelte beste Lösung im Laufe der Anwendung der Partikelschwarmoptimierung nur dann verloren, wenn tatsächlich eine bessere Lösung gefunden wird. Ein Partikel speichert seine Anfangsposition solange als persönliche Bestposition, bis er sich an eine Stelle des Suchraums mit besserer Lösungsqualität bewegt.

3.6.3 Wahl der Kommunikationstopologie

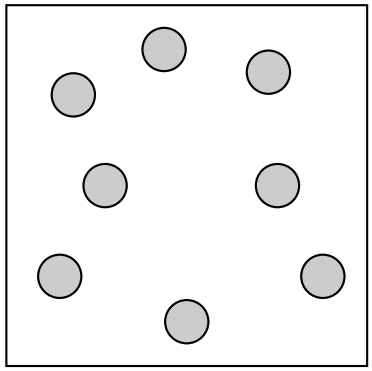


Abbildung 3.29: Darstellung einer Gruppe von Partikeln ohne Kommunikationsrelationen

Je nach Aufgabenstellung bieten sich verschiedene Topologien zur Abstimmung zwischen Gruppenmitgliedern an. Einige wichtige Klassen sollen an dieser Stelle vorgestellt werden. Die Topologien werden als Graph dargestellt, die Ecken des Graphen sind die Partikel, die Kanten stellen Kommunikationsrelationen dar, gerichtete Kanten entsprechen einseitigen, ungerichtete Kanten oder beidseitig gerichtete Kanten zweiseitigen Kommunikationsrelationen. Abbildung 3.29 zeigt eine Gruppe von Agenten ohne Kommunikationstopologie. Für die Darstellung der Kommunikationsverbindungen ist die räumliche Position eines Agenten unerheblich. Die vorgestellten Topologien zeigen verbundene Gruppen von Partikeln, die Verwendung mehrerer unverbundener Gruppen ist denkbar, aber unüblich. Es sind mehr Topologien als die vorgeschlagenen konstruierbar, trotzdem können wesentliche Unterschiede an diesen Strukturen verdeutlicht werden.

3.6.3.1 Totalvernetzung

In dieser Struktur ist jedes Element mit jedem anderen Element verbunden. Information, die einem Agenten zur Verfügung steht, steht allen Agenten durch einfache Kommunikation zur Verfügung. Dementsprechend bilden alle Agenten zusammen eine große Einheit. Diese Bündelung wird erkauft durch den Verzicht auf Unabhängigkeit. Mit Ausnahme der persönlichen Gedächtnisstellen sind viele Stellen redundant belegt. Soll aber eine bestimmte Region des Suchraums gründlich durchsucht werden, kann diese Topologie von Nutzen sein. Ebenso droht aber die Gefahr, dass die Festlegung auf einen Bereich zu früh erfolgt, und die Gesamtgruppe in lokalen Minima haften bleibt. Abbildung 3.30 (links) zeigt eine Topologie mit Totalvernetzung.

3.6.3.2 Räder

In einer Radstruktur (*wheel structure*) oder auch Sternstruktur ist ein einziger Agent mit allen anderen Agenten verbunden, die jeweils ausschließlich mit ihm verbunden sind. Dieser Agent bündelt die gesamte

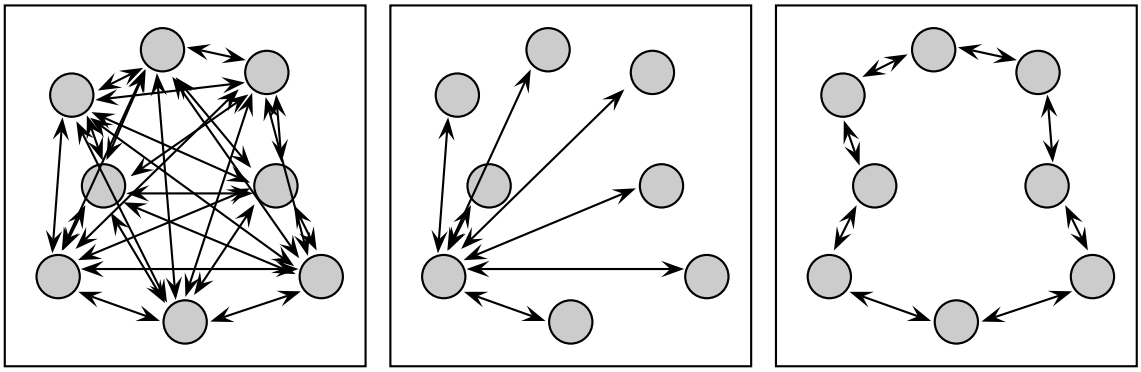


Abbildung 3.30: Darstellung einer Kommunikationstopologie mit Totalvernetzung (links), einer Rad-Kommunikationstopologie (Mitte) und einer Ring-Kommunikationstopologie (rechts)

Information der Topologie und verfügt als einziger in einem Schritt über den kollektiven Wissensschatz. Gleichzeitig verteilt er dieses Wissen an alle übrigen Agenten, wenn er sich in die entsprechenden Teile des Suchraums bewegen sollte. Diese Struktur unterscheidet sich von der Totalvernetzung mit Ausnahme der Tatsache, dass die Kommunikation zwischen zwei nichtzentralen Partikeln nur über das Explorationsverhalten des zentralen Agenten erfolgen kann und damit auf jeden Fall verzögert wird. In Abbildung 3.30 (Mitte) ist eine solche Struktur dargestellt.

3.6.3.3 Ringe

In Ring- oder Kreis-Strukturen bilden alle Agenten zusammen eine geschlossene Kette. Jeder Agent ist mit genau zwei Nachbarn verbunden, so dass Information relativ lokal verteilt ist. In dieser Struktur ist individuelle Exploration möglich, deren Ergebnisse erst langsamer an die übrigen Agenten verteilt werden. Abbildung 3.30 (rechts) demonstriert ein derartiges Verknüpfungsschema. Da Information Schritt für Schritt den Ring durchlaufen muss, ist eine individuelle Exploration für einige Zeitschritte möglich, so dass unter Umständen bessere Positionen außerhalb der Regionen mit lokalen Minima gefunden werden können, bevor der Schwarm auf eine Region konvergiert.

3.6.3.4 Randomisierte Topologien

Durch Angabe einer Verbindungswahrscheinlichkeit p_v kann bestimmt werden, welcher Anteil der möglichen Verbindungen durchschnittlich vorhanden sein wird, bzw. mit welcher Wahrscheinlichkeit eine einzelne Verbindung besteht. So kann bei unbekannten Eigenschaften des Suchraums mit verschiedenen Anfangszuständen experimentiert werden. Der Wahrscheinlichkeitswert für das Vorliegen einer Verbindung zwischen zwei Partikeln kann so spezifiziert werden, dass zwischen $p_v = 0$ und einer damit verbundenen Totalisolierung der Elemente und $p_v = 1$ und der damit verbundenen Totalvernetzung alle Schattierungen durchlaufen werden können. Diese beiden Extremwerte sind allerdings keine sinnvollen Wahlen, da keine Randomisierung für die Erstellung der Topologie erforderlich wäre. Die Totalisolation führt zudem zu einer radikalen Einbuße in der Effizienz des Suchalgorithmus, da keine Information zwischen Agenten ausgetauscht werden kann und statt dessen lediglich parallele unverbundene Suchprozesse ablaufen würden. Abbildung 3.31 zeigt eine randomisierte Topologie mit $p_v = 0.25$ und eine randomisierte Topologie bei anderer Parameterwahl ($p_v = 0.5$).

Die Anzahl aller möglichen Relationen ergibt sich bei k Partikeln als $k(k-1)$, da für jeden Agenten $k-1$ einseitige Kommunikationsrichtungen existieren. Damit ergibt sich die erwartete Anzahl von Verbindungen als $p_v k(k-1)$. Soll die Anzahl der Verbindungen exakt festgelegt werden, könnte die Erstellung auch über eine randomisierte Auswahl von genau $p_v k(k-1)$ paarweise verschiedenen Verbindungen erfolgen.

3.6.4 Beschleunigung und Bewegung

3.6.4.1 Bewegungsprinzip

Die Regeln für die Bewegung der Partikel im Suchraum sind schlicht. Die Zeit ist für den Suchprozess in diskrete Schritte unterteilt und für jeden Partikel i sind Position und Geschwindigkeit für einen Zeitschritt

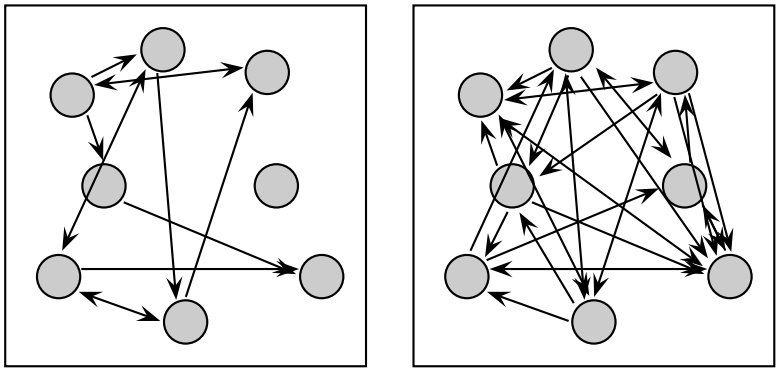


Abbildung 3.31: Darstellung von randomisierten Kommunikationstopologien mit $p_v = 0,25$ (links) und $p_v = 0,5$ (rechts)

t gespeichert. Diese beiden Informationen lassen sich als Positionsvektor $\vec{x}_i$ und als Geschwindigkeitsvektor $\vec{v}_i$ festhalten. Wie bereits erwähnt würde der Positionsvektor im folgenden Zeitschritt $t+1$ bei fehlender Beschleunigung dem Ergebnis der Vektoraddition von Positions- und Geschwindigkeitsvektor entsprechen. Geht man davon aus, dass $\vec{v}_i(t)$ die Bewegung beschreibt, die in einem Zeitschritt t im Vergleich zum letzten Zeitschritt $t-1$ zurückgelegt wurde (also nach Einbeziehung aller möglichen Beschleunigungen), dann entspricht dies Gleichung 3.36:

$$\vec{x}_i(t) = \vec{x}_i(t-1) + \vec{v}_i(t) \tag{3.36}$$

Würde keine Beschleunigung auf den Partikel einwirken, dann wäre $\vec{v}_i(t+1) = \vec{v}_i(t)$. Der Suchprozess wäre in diesem Fall eine wenig interessante eindimensionale Durchquerung des Suchraums in fixen Intervallen. Aus diesem Grund sollen in die Berechnung der neuen Position zusätzlich zu Position und Geschwindigkeit auch die gespeicherte individuelle beste Position $\vec{p}_{b_i}(t)$ und die gespeicherte beste Position der Nachbarn $\vec{p}_{n_i}(t)$ einbezogen werden, dementsprechend ist

$$\vec{x}_i(t) = f\big(\vec{x}_i(t-1), \vec{v}_i(t-1), \vec{p}_{b_i}(t-1), \vec{p}_{n_i}(t-1)\big). \tag{3.37}$$

3.6.4.2 Ermittlung von Bewegung und Geschwindigkeit

Die Einbeziehung der gespeicherten Positionen erfolgt über einen Zufallsprozess. Dabei kommen zwei Zufallsvariable zum Einsatz, die aus einer gleichförmigen Verteilung zwischen 0 und einem Maximalwert $\varphi_{b_{max}}$ bzw. $\varphi_{n_{max}}$ annehmen können. Kennedy und Eberhart (2001) verwenden gleiche Bezeichnungen für die Zufallsvariablen, die Werte der Zufallsvariablen und das Maximum der Zufallsverteilung, hier sollen φ_b und φ_n die in einem Zeitschritt unabhängig voneinander bestimmten Werte bezeichnen, $\varphi_{b_{max}}$ und $\varphi_{n_{max}}$ die maximalen Werte. Dabei wird zudem φ definiert als:

$$\varphi = \varphi_{b_{max}} + \varphi_{n_{max}}, \text{ mit } 0 < \varphi_{b_{max}} < \varphi \tag{3.38}$$

bestimmt. Im allgemeinen werden die beiden Maximalgrenzen auf $\frac{\varphi}{2}$ gesetzt. Somit werden für jeden Zeitschritt zufällig zwei Zahlen zwischen 0 und $\frac{\varphi}{2}$ bestimmt, deren Summe einen Wert zwischen 0 und φ ergeben muss. Die Geschwindigkeit eines Partikels i wird zu einem Zeitpunkt t bestimmt als

$$\vec{v}_i(t) = \vec{v}_i(t-1) + \varphi_b(\vec{p}_{b_i}(t-1) - \vec{x}_i(t-1)) + \varphi_n(\vec{p}_{n_i}(t-1) - \vec{x}_i(t-1)). \tag{3.39}$$

Zusammen mit Gleichung 3.36 folgt

$$\vec{x}_i(t) = \vec{x}_i(t-1) + \vec{v}_i(t-1) + \varphi_b(\vec{p}_{b_i}(t-1) - \vec{x}_i(t-1)) + \varphi_n(\vec{p}_{n_i}(t-1) - \vec{x}_i(t-1)). \tag{3.40}$$

.

Damit kann sowohl die neue Position der Partikel als auch ihre neue Geschwindigkeit bestimmt werden. Abbildung 3.32 veranschaulicht den Bewegungsschritt eines Partikels. Die gespeicherte Geschwindigkeit würde diesen Partikel an den beiden Optimalstellen vorbei treiben, wenn keine neue Beschleunigung auf ihn einwirken würde. Der Punkt (6,5;2,8) wäre der Zielpunkt des Partikels, wenn sowohl φ_b und φ_n den

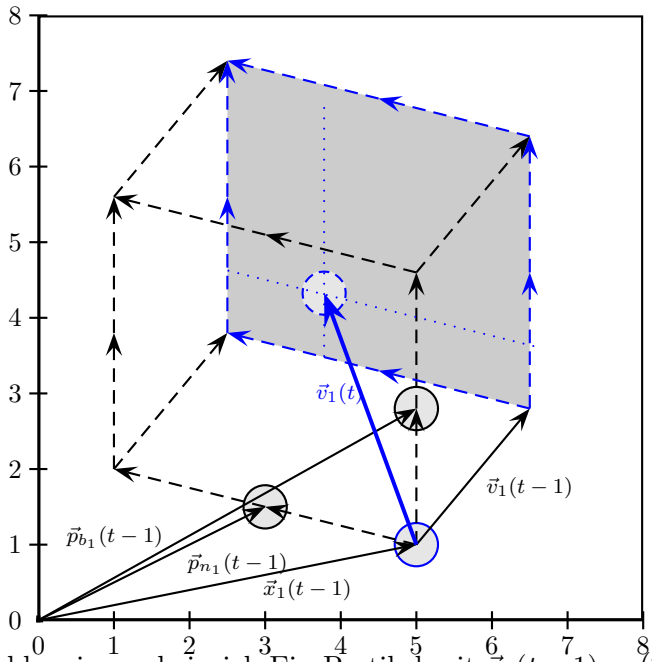


Abbildung 3.32: Beschleunigungsbeispiel: Ein Partikel mit $\vec{x}_1(t-1) = \binom{5}{1}$ wird im Zeitschritt t beschleunigt. Die beiden gespeicherten Positionen sind $\vec{p}_{n_1}(t-1) = \binom{3}{1,5}$ und $\vec{p}_{b_1}(t-1) = \binom{5}{2,8}$. Es gelten $\varphi_{n_{max}} = 2$ und $\varphi_{b_{max}} = 2$, die ermittelten Zufallszahlen sind $\varphi_n = 1,36$ und $\varphi_b = 0,47$. Es resultiert der Geschwindigkeitsvektor $\vec{v}_1(t) = \binom{-1,22}{3,326}$. Damit ist die neue Position des Partikels nach der Bewegung $\vec{x}_1(t) = \binom{3,78}{4,326}$. Der schattierte Bereich zeigt alle möglichen Zielpunkte der Bewegung bei allen möglichen Kombinationen von Werten der beiden Zufallszahlen in diesem Zeitschritt.

Wert 0 hätten. Aus der Skalarmultiplikation der Zufallswerte jeweils mit den Vektoren $\vec{p}_{n_1}(t-1)-\vec{x}_1(t-1)$ und $\vec{p}_{b_1}(t-1) - \vec{x}_1(t-1)$ ergeben sich zwei Vektoren, deren Addition die Beschleunigung des Partikels ergibt. Alle möglichen Zielpunkte der Bewegung des Partikels müssen also innerhalb des Parallelogramms liegen dessen Ecken vom Nullpunkt aus durch die Vektorensummen $\vec{x}_1(t-1) + \vec{v}_1(t-1)$, $\vec{x}_1(t-1) + \vec{v}_1(t-1) + \varphi_{b_{max}} \cdot (\vec{p}_{b_1}(t-1) - \vec{x}_1(t-1))$, $\vec{x}_1(t-1) + \vec{v}_1(t-1) + \varphi_{n_{max}} \cdot (\vec{p}_{n_1}(t-1) - \vec{x}_1(t-1))$ und $\vec{x}_1(t-1) + \vec{v}_1(t-1) + \varphi_{b_{max}} \cdot (\vec{p}_{b_1}(t-1) - \vec{x}_1(t-1)) + \varphi_{n_{max}} \cdot (\vec{p}_{n_1}(t-1) - \vec{x}_1(t-1))$ gegeben sind. In der Abbildung ist die Fläche, die aus den möglichen Zielpunkten gebildet wird, schattiert gekennzeichnet. Im Beispiel würde der Partikel den Zeitschritt $t+1$ an der Position $\vec{x}_1(t)$ mit der Geschwindigkeit $\vec{v}_1(t)$ beginnen. Es ist damit nicht ausgeschlossen, dass sich der Partikel unter Umständen weiter von den Optima entfernen wird. Bei steigendem Abstand wird die zu erwartende relativ zur bestehenden Bewegung negative Beschleunigung größer, so dass eine erneute Annäherung an die Optima zu erwarten ist.

3.6.4.3 Die Begrenzung der Geschwindigkeit

Würde man den Algorithmus in der beschriebenen Form starten, eskaliert die Geschwindigkeit der Partikel nach einigen Zeitschritten, die Oszillationen um das Optimum werden größer und größer, die Geschwindigkeit steigt kontinuierlich an. Um diesen unerwünschten Effekt zu bekämpfen wird ein Parameter V_{max} eingeführt. Jede Dimension des Geschwindigkeitsvektors wird daraufhin überprüft, ob der Betrag der Komponente den Parameterwert übersteigt und in diesem Fall auf den Parameterwert zurückgesetzt. Der Geschwindigkeitsvektor $\vec{v}^T = (v_1, v_2 ..., v_n)$ wird also so angepasst, dass für jede Komponente v_i gilt:

$$v_i = \begin{cases} V_{max}, & v_i > V_{max} \\ v_i, & |v_i| < V_{max} \\ -V_{max}, & v_i < -V_{max} \end{cases} \tag{3.41}$$

3.6.5 Die Evaluation der Lösungen

Zur Bestimmung der internen Variablen bzw. der Gedächtnisstellen der Partikel ist eine Analyse der gefundenen Lösung notwendig. Da jeder Positionsvektor eine Lösung vollständig repräsentiert, kann die

Qualität einer Lösung als Funktion eines Positionsvektors $G(\vec{x})$ definiert werden. In jedem Zeitschritt wird nun nach der Bestimmung der neuen Positionen $\vec{x}_i(t)$ für jedes Individuum der Wert $G(\vec{x}_i(t))$ bestimmt. Dann ist:

$$\vec{p}_{b_i}(t) = \begin{cases} \vec{x}_i(t) & \text{für } G(\vec{x}_i(t)) > G(\vec{p}_{b_i}(t-1)) \\ \vec{p}_{b_i}(t-1) & \text{für } G(\vec{x}_i(t)) \leq G(\vec{p}_{b_i}(t-1)) \end{cases} \tag{3.42}$$

Der Positionsvektor wird an die aktuelle Position angeglichen, wenn die Qualität der mit dieser Position verbundenen Lösung die Qualität der bisher besten Lösung übertrifft.

Nachdem alle Individuen diesen Prozess durchlaufen haben, werden nun die Nachbarschaften der Individuen untersucht. Die Relation „a kommuniziert mit b“ soll als aKb definiert werden. Dementsprechend ist:

$$\begin{aligned} \vec{p}_{n_i}(t) =& \vec{p}_{b_j}(t) \\ & \text{mit } iKj \wedge \big(\forall k\big)\big(k \neq j \wedge iKk \rightarrow (G(\vec{p}_{b_j}(t)) > G(\vec{p}_{b_k}(t))) \\ & \quad \vee (G(\vec{p}_{b_j}(t)) = G(\vec{p}_{b_k}(t)) \wedge j < k)\big) \end{aligned} \tag{3.43}$$

.

Der zweite Positionsvektor wird an die gespeicherte persönliche Optimalposition des Nachbarn angeglichen, für die die Qualitätsfunktion den höchsten Wert aufweist. Damit sind alle Variablen des Partikels für einen Zeitschritt bestimmt. Die beste Optimalposition über alle Partikel hinweg soll als $\vec{p}_o(t)$ bezeichnet werden, die Qualität dieser Lösung soll $G_{max}(t) = G(\vec{p}_o(t))$ heißen.

3.6.6 Das Abbruchkriterium

Die bisher dargestellte Abfolge von Aktionen stellt einen Zeitschritt dar. Diese Schritte werden iteriert, bis ein Abbruchkriterium erfüllt ist. Im einfachsten Fall ist dieses Kriterium das Erreichen einer bestimmten Schrittzahl, es können aber auch komplexere Kriterien zum Einsatz kommen. Prinzipiell bieten sich alle Kriterien an, die für Genetische Algorithmen definiert wurden, also im einzelnen:

- Anzahl der Schritte: Die Anzahl der Generationen wird durch die Anzahl der Zeitschritte ersetzt
- Erreichen einer vordefinierten Lösungsqualität: Der Algorithmus wird Zeitschritt nach Zeitschritt ausgeführt, bis $G_{max}(t)$ einen vorher festgelegten kritischen Wert G^* überschreitet. Alternativ kann auch eine eigene Zielfunktion $H(\vec{x})$ mit $\vec{x} \in \mathcal{R}^n$ definiert werden und ein kritischer Wert H^* und das Abbruchkriterium wird zu $H(\vec{p}_o(t)) > H^*$.
- Überschreiten einer vorher definierten Laufzeit: Der Algorithmus wird zu einem bestimmten Zeitpunkt abgebrochen, unabhängig von der bis zu diesem Zeitpunkt erreichten Lösungsqualität. Dies kann abhängig von der lokalen Realisierung zu unterschiedlichen Konsequenzen für die zu erwartende Anzahl an Zeitschritten führen.
- Kreuzvalidierung: Die Daten werden in Trainings- und Testdatensatz gespalten, und der Suchprozess wird solange fortgesetzt, wie die Verbesserung der optimalen Lösung für den Trainingsdatensatz auch zu einer Verbesserung der optimalen Lösung für den Testdatensatz führt.
- Andere Kriterien: Natürlich können auch spezifische Kriterien für Schwärme entwickelt werden. So kann zum Beispiel die Konvergenz der individuellen Partikeloptima, definiert als Streuung der Positionswerte, summiert über alle Dimensionen oder auch als durchschnittlicher Abstand zwischen zwei Partikeln herangezogen werden.

Auch für die Wahl des Abbruchkriteriums gibt es keine festen Regeln, so dass der Anwender hier erneut eine Entscheidung treffen muss. Fällt die Entscheidung für den Abbruch, so wird keine neue Beschleunigungsphase angeschlossen. Alle individuellen Optima werden untersucht und $G_{max}(t)$ und die Position $p_o(t)$ werden ein letztes Mal ausgegeben. $p_o(t)$ wird dekodiert, um die durch den Algorithmus gewonnene Lösung zu erhalten. Selbstverständlich kann ein Schwarmalgorithmus auch in eine Kette von Algorithmen integriert werden, indem alle Schwarmpositionen an einen Folgealgorithmus weitergegeben werden. Damit ist auch der abschließende Schritt des Algorithmus erläutert.

3.6.7 Der Partikelschwarm in Pseudocode

Die dargestellten Schritte für einen Zeitschritt in der Hauptschleife lassen sich in Pseudo-Code zusammenfassen:

```
SCHWARMZEITSCHRITT(pschwarm)
 1  for i ← 1 to anz_Partikel
 2  do if POS_ FITNESS(pschwarm.pkel[i].pos) >
 3        POS_ FITNESS(pschwarm.pkel[i].p_best)
 4        then for j ← 1 to anz_Dimensionen
 5               do pschwarm.pkel[i].p_best[j] ← pschwarm.pkel[i].pos[j]
 6     for k ← 1 to pschwarm.pkel[i].anz_Nachbarn
 7     do if POS_ FITNESS(pschwarm.pkel[pschwarm.pkel[i].nachbar[k]].pos) >
 8           POS_ FITNESS(pschwarm.pkel[i].n_best)
 9           then for j ← 1 to anz_Dimensionen
10                  do pschwarm.pkel[i].n_best[j]
11                     ← pschwarm.pkel[pschwarm.pkel[i].nachbar[k]].pos[j]
12     Phi_p ← RANDOM ∗ (Phi/2)
13     Phi_n ← RANDOM ∗ (Phi/2)
14     for d ← 1 to anz_Dimensionen
15     do pschwarm.pkel[i].geschw[d] ← pschwarm.pkel[i].geschw[d]
16           + Phi_p ∗ (pschwarm.pkel[i].p_best − pschwarm.pkel[i].pos[d])
17           + Phi_n ∗ (pschwarm.pkel[i].n_best − pschwarm.pkel[i].pos[d])
18        if pschwarm.pkel[i].geschw[d] > Vmax
19           then pschwarm.pkel[i].geschw[d] ← Vmax
20           else  if pschwarm.pkel[i].geschw[d] < −Vmax
21                    then pschwarm.pkel[i].geschw[d] ← −Vmax
22        pschwarm.pkel[i].pos[d] ← pschwarm.pkel[i].geschw[d]
```

Listing 3.4: Zeitschrittalgorithmus für den Partikelschwarm in Pseudocode

Die Anfangsinitialisierung, die Prüfung des Abbruchkriteriums und die Ausgabe der Lösung müssen zusätzlich spezifiziert werden.

3.6.8 Individuelle Exploration und Kommunikation

Trotz der Schlichtheit des Algorithmus bieten Schwärme einen gewissen Schutz gegen die Anziehungskraft lokaler Maxima im Suchraum. Das Prinzip, durch das diese Leistungsfähigkeit gewonnen wird, ist Abbildung 3.33 zu entnehmen. Die Farbe der Regionen im abgebildeten Suchraum markiert die Qualität der Lösungen, die dunkelste Stelle stellt das globale Optimum dar. Es ist zu sehen, dass mehr Partikel sich in dunklen als in hellen Regionen befinden, viele Partikel aber die Stellen mit lokalen Maxima besetzen. Da aber Nachbarn ihr persönliches Optimum kommunizieren, besteht die Möglichkeit, dass Partikel bildlich gesprochen aus diesen Stellen herausgezogen werden können, wenn die Zufallskombination sie in Richtung des Nachbarschaftsoptimum beschleunigt. Zudem wäre es ausreichend, wenn ein einziger Partikel das globale Optimum finden würde, in praktischen Anwendungen kann das Vorliegen dieses Falls aber zumeist nicht bestimmt werden.

Damit ist der Partikelschwarm eines der einfachsten Modelle, in denen die Unabhängigkeit einzelner Agenten und Strategien mit Kommunikationsmechanismen verbunden werden, durch die verteilte Information wieder gebündelt werden kann. Wären die Agenten vollständig isoliert, entspräche ein Durchlauf mit 20 Agenten einer zwanzigfachen Wiederholung eines Durchlaufs mit einem einzigen Agenten. Der Partikelschwarm arbeitet aber schneller, da durch die Kommunikation von Lösungen lohnenswerte Stellen im Suchraum kommuniziert werden, die die Partikel schneller in interessante Regionen ziehen. Ideal wäre also ein Verhalten, bei dem zunächst der Raum exemplarisch und großflächig nach relevanten Stellen durchsucht würde, um dann in einer späteren Phase die identifizierten Regionen näher zu durchsuchen. Dies kann im Idealfall dann der Fall sein, wenn alle Agenten in der Nähe des globalen Optimums zusam-

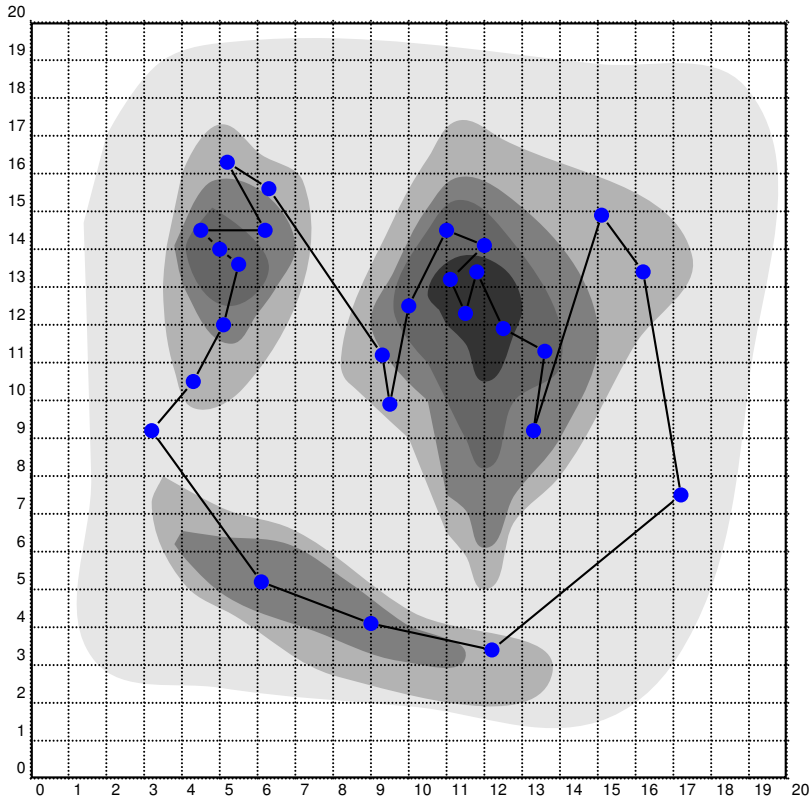


Abbildung 3.33: Partikelschwärme und lokale Maxima im Suchraum: Das Beispiel zeigt einen Partikelschwarm mit Ringtopologie während der Suche in einem zweidimensionalen Suchraum. Die Dunkelheit der Regionen soll verschiedene Niveaus der Werte der Zielfunktion für die Punkte dieser Regionen anzeigen, die dunkelste Region repräsentiert das globale Optimum.

menkommen, es kann natürlich auch der Fall sein, wenn alle Agenten um eines oder mehrere nicht-globale lokale Optima zusammenkommen. Die Zufallsabhängigkeit der Beschleunigung hingegen sorgt für eine Suche in der Nähe der identifizierten Stellen, aber auch in einigem Abstand von den Optimalstellen.

3.7 Praktische Anwendungen der vorgestellten Heuristiken

3.7.1 Anwendungen genetischer Algorithmen

Die Präsentation eines vollständigen Überblicks über Anwendungen genetischer Algorithmen in verschiedenen wissenschaftlichen Disziplinen und in der Praxis ist bereits zu diesem Zeitpunkt ein hoffnungsloses Ziel, zu breit sind die Anwendungsvariationen und zu viele Einsatzgebiete lassen sich ausmachen. Aus diesem Grund sollen exemplarisch einige Anwendungen skizziert werden, um einen Eindruck von der Breite der Problemklassen zu geben, deren Lösungsmöglichkeiten durch genetische Algorithmen in sehr unterschiedlichen Kontexten untersucht wurden. Dabei soll grob einer mäßig trennscharfen disziplinären Einordnung gefolgt werden.

Einige beispielhafte Anwendungen in der Chemie und Physik sind in Tabelle 3.5 aufgeführt, Anwendungen in der Biochemie in Tabelle 3.6. Dabei liegt ein Fokus auf der Identifizierung von Molekülstrukturen und der Auswertung komplexer Datensätze. Die Suche nach einer zu den Daten passenden räumlichen Anordnung von Atomen lässt sich durch genetische Algorithmen ebenso leisten wie die Verarbeitung hochdimensionaler Datensätze. Einen Überblick über Anwendungen in der Chemie gibt auch Leardi (2001).

Statistische Anwendungen und Anwendungen im Bereich der Visuellen Mathematik sind in Tabelle 3.7 zusammengestellt. Genetische Algorithmen wurden dabei für Klassifikationsaufgaben eingesetzt, aber auch für die Suche nach Clusterlösungen, Vektorquantisierung und die Lösung von statistischen Problemen, die als *NP-hard* eingestuft werden können. Zu dieser Klasse gehört das Rucksackproblem, bei dem ein Rucksack mit Objekten variablen Gewichts möglichst bis zur Kapazitätsgrenze gefüllt werden soll, oder das NP-vollständige TSP-Problem. Einen Überblick über statistische Anwendungsmöglichkeiten genetischer Algorithmen geben auch Chatterjee, Laudato und Lynch (1996).

Das Spektrum der Anwendungen in den Ingenieurswissenschaften reicht von Designproblemen in unter-

Tabelle 3.5: Beispiele für GA-Anwendungen in der Chemie und Physik

Quelle	Problem	Lösungsansatz
Eliseev und Nelen (1998)	Generierung und Identifizierung von Molekülstrukturen	GA
Domínguez und Arcos (2000)	Variablenreduktion in der Chromspeziesbestimmung	Partielle Kleinst-Quadrate-Regression mit GA
Chaudhury und Bhattacharyya (2000)	Stochastische Rekonstruktion von Reaktionspfaden	Eigenwertbestimmung durch GA
Oda, Nagao, Kitagawa, Shigeta und Yamaguchi (2000)	Analyse des magnetischen Verhaltens von Spin-Clustern	Monte-Carlo-Methode mit GA
Leardi (2000)	Merkmalsauswahl bei der Analyse spektraler Datensätze	Partielle Kleinst-Quadrate-Regression mit GA
Brunetti und Golosio (2001)	Kurvenanpassung in der energiedispersiven Röntgen-Fluoreszenzanalyse	GA
Monett, Méndez, Abraham, Gallardo und Román (2002)	Berechnung von Reaktivitätsverhältnissen	GA

Tabelle 3.6: Beispiele für GA-Anwendungen in der Biochemie

Quelle	**Problem**	**Lösungsansatz**
Yamashita, Wanchana und Hashida (2002)	Quantitative Struktur/Eigenschaftsanalyse von Caco2-Permeabilität	GA-basierte partielle Kleinstquadratemethode
Rockey, Laederach und Reilly (2000)	Protein-Docking	GA mit Lamarckscher Vererbung
Kamphausen et al. (2002)	Design von Molekülen mit gewünschten Eigenschaften	GA mit $n \times m$-Crossover
Gardiner, Willett und Artymiuk (2001)	Protein-Docking	GA
Pedersen und Moult (1997)	Protein Folding-Simulationen	GA-basierte Suchmethode
Jin, Leung und Weaver (1999)	Suche im biomolekularen Konformationsraum	Drei GA-Varianten
Standley, Gunn, Friesner und McDermott (1998)	Vorhersage der Tertiärstruktur von Proteinen	Simulated Annealing mit GA-Komponente

Tabelle 3.7: Beispiele für visuelle und statistische GA-Anwendungen

Quelle	Problem	Lösungsansatz
Yu und Gao (1997)	Radarbildrekonstruktion	GA
Chtioui, Bertrand und Barba (1998)	Visuelle Samenkorndiskriminierung	GA mit NN-Methode
Larrañaga, Kuijpers, Murga, Inza und Dizdarevic (1999)	Lösung des Travelling-Salesman-Problems	GA
Fränti (2000)	Vektorquantisierung	GA
Prabhu, Buckles und Petry (2000)	Visuelle Wolkentypklassifikation	GA mit Inselalgorithmus
Araki, Kaji, Yamamoto, Suzuki und Ohuch (2001)	Lösung des Problems der optimalen sequenziellen Partitionierung	GA
Chiu und Chen (2001)	Bildrekonstruktion	GA
Lipnitskii (2002)	Lösung des Rechteckpackproblems	GA
Nandy, Chaudhury und Bhattacharyya (2002)	Bestimmung der Eigenwerte und Eigenvektoren Hamiltonscher Matrizen	GA
Li, Balazs, Parks und Clarkson (2002)	Multimodale Funktionsoptimierung	GA mit Speziesregeln
Ezziane (2002)	Lösung des Rucksackproblems	GA
Ho und Lee (2003)	Bildsegmentierungsanalyse	GA mit k-means-Algorithmus
Kivijärvi, Fränti und Nevalainen (2003)	Clustering	selbstadaptiver GA
Gómez und Bielza (2004)	Knoteneliminierung in Einflussdiagrammen	GA

Tabelle 3.8: Beispiele für GA-Anwendungen in den Ingenieurswissenschaften

Quelle	Problem	Lösungsansatz
Lee (1997)	Design von Marinefahrzeugen	GA mit direct search
Steinberg und Page (1999)	Nichtlineare adaptive Flugkontrolle	GA
Yoshioka und Omatu (2000)	Trennung unabhängiger Eingangssignale	GA
Ponterosso und Fox (2000)	Optimierung verstärkter Straßenhänge	GA
Cortés, Larrañeta, Onieva, García und Caraballo (2001)	Planung von Kabelnetzwerken zur Telekommunikation	GA
Jahn (2001)	Zuweisung von Kanalfrequenzen an Satellitensysteme	GA
Li, Liu, Zhang, Tam und Yang (2001)	Strukturelle Kontrolle bei Stürmen und Erdbeben	Mehrebenen-GA
Oyama, Obayashi und Nakamura (2001)	Aerodynamisches Design von Flugzeugflügeln	GA
Koyamada, Yamada, Nishio und Kotera (2001)	Steigerung der Genauigkeit in der Thermalsimulation elektronischer Apparaturen	GA
Coello Coello und Aguirre (2002)	Design kombinatorischer Logikschaltkreise	GA
Zhong, Xing, Weicheng, Peide und Baozhi (2003)	Kontrolle der Luftversorgung bei Bergwerksbränden	GA
Whitfield, Duffy, Coates und Hills (2003)	Effiziente Prozessoptimierung durch Bestimmung von Aktivitätssequenzen	GA
Simpson und D'Souza (2004)	Bewertung von Plattformkommunalitäten in Produktfamilien	multiobjektiver GA

Tabelle 3.9: Beispiele für wirtschatsbezogene GA-Anwendungen

Quelle	Problem	Lösungsansatz
Varetto (1998)	Analyse des Insolvenzrisikos von Unternehmen	GA
Aldershof und Carducci (1999)	Suche nach stabiler Paarung von Bewerbern und Arbeitgebern	GA
Dhar, Chou und Provost (2000)	Mustererkennung für die Unterstützung von Investitionsentscheidungen	GA mit Entropiereduktion
Chang, Christensen und Zhang (2001)	Projektmanagement	GA
Aickelin und Dowsland (2002)	Entwurf des Layouts für ein Einkaufszentrum und Pächterauswahl	GA mit Decoder
Hsu, Welge, Redman und Clutter (2002)	Prädiktion und Entscheidungsunterstützung	GA
Korczak und Roger (2002)	Kauf und Verkauf von Aktien	GA
Gatarski (2002)	Design von Bannerwerbung	GA
Boll (2003)	Kooperative Produktionsplanung und -kontrolle	GA mit impliziter Kooperation
Viotti, Polettini, Pomi und Innocenti (2003)	Optimierung von Müllabfuhrrouten	GA
Lawrenz und Westerhoff (2003)	Modellierung von Wechselkursen	GA

Tabelle 3.10: Beispiele für GA-Anwendungen in der Medizin

Quelle	Problem	Lösungsansatz
Kragelund (1997)	Zeitplanung für Mediziner im Notdienst	paralleler GA
Aickelin und Dowsland (1998)	Erstellung von Einsatzplänen für Krankenschwestern	Koevolutionärer GA
Mark Bower und Rowe (1999)	Überlebensprognose bei Tumordiagnosen	Diffusions-GA
Pantoja, García, Hernández-López, Bretones und Martín (2002)	Design von Antennen zur Entdeckung von Brusttumoren	GA
Petricoin, Paweletz und Liotta (2002)	Identifikation von Proteinveränderungen in Körperflüssigkeiten	GA
Durant (2002)	Anpassung von Hörgeräten	GA
Ahmad und Bath (2004)	Mortalitätsprognose aufgrund von Risikofaktoren	GA mit Cox-Regression

Tabelle 3.11: Beispiele für GA-Anwendungen in der Robotik

Quelle	Problem	Lösungsansatz
Chen und Zalzala (1997)	Wegplanung für mobile Roboter	GA
Cheng und Lin (1997)	Steuerungsdesign für bipedale Bewegung	GA
Zhuang, Wu und Huang (1997)	Planung von Kalibrierungsexperimenten für Roboter	GA
Nakashima, Maruyama, Umeda, Kyura und Hasegawa (1998)	Vibrationskontrolle bei Robotern für Reparaturen an Stromleitungen	GA
Zhu, Qiu und Tani (2001)	Optimierung flexibler Roboterarme	GA
Tuci, Quinn und Harvey (2002)	Untersuchung von Lernverhalten mit einem roboterbasierten Modell	GA mit neuronalen Netzwerken
Gallant und Boudreau (2002)	Synthese paralleler Manipulatoren mit prismatischen Gelenken	GA

Tabelle 3.12: Beispiele für GA-Anwendungen in Simulationen

Quelle	Problem	Lösungsansatz
Birchenhall, Kastrinos und Metcalfe (1997)	Modellierung evolutionärer Dynamik	modifizierter GA
Dawid und Kope (1998)	Modellierung ökonomischer Agenten	GA
Novkovic (1998)	Simulation einer Transitionsökonomie	GA
Roy P. Pargas und Peck (1999)	Generierung von Testdaten für Softwaretests	GA
Chen und Ni (2000)	Simulation eines Oligopol-Spiels für drei Personen	GA
Özyildirim und Criss (2001)	Simulation von Aufrüstungsprozessen	GA
Hong, Huang und Lin (2001)	Suche in Spielbäumen	GA
Rees und Koehler (2002)	Modellierung von Austauschprozessen bei Gruppenentscheidungen	GA
Tiampo et al. (2002)	Simulation von Erdbeben	GA mit zwei anderen Simulatoren
Lee, Lee und Rho (2002)	Simulation der Entstehung und Dynamik strategischer Gruppen	GA
Lauret, Boyer, Riviere und Bastide (2005)	Prüfung der Thermalsimulation von Gebäuden	GA

schiedlichen Bereichen bis hin zu Kontrollaufgaben und Datenverarbeitungsproblemen (vgl. Tabelle 3.8). Dabei fällt auf, dass auch sensible Bereiche wie das Design von Flugzeugflügeln oder die Steuerung der Luftversorgung genetischen Algorithmen zumindest in der Entwurfsphase anvertraut werden. Anwendungen im ökonomischen Kontext sind in Tabelle 3.9 aufgelistet: Das Spektrum reicht hier von Aufgaben der Projektplanung und Zeiteinteilung bis hin zu Risikoeinschätzungen und Entscheidungsunterstützung im Aktienhandel. Zudem werden so unterschiedliche Bereiche wie die Planung von Müllabfuhrrouten und die Suche nach guten Kombinationen von Bewerbern und Arbeitgebern mit genetischen Algorithmen untersucht. Im Bereich der Medizin ist das Feld der künstlichen Intelligenz auch durch die Gründung spezieller Zeitschriften (z.B. „Artificial Intelligence in Medicine“) verstärkt zu einer beliebten Quelle für die Entwicklung von Prognose- und Diagnoseverfahren geworden. Auch genetische Algorithmen werden für diese Fragestellungen, aber auch für die Erstellung von Dienstplänen oder die Anpassung von Hörgeräten eingesetzt, wie Tabelle 3.10 zusammenfasst.

Dem Künstlichen näher steht das Feld der Robotik, in dem sowohl Design-Aufgaben wie auch die Steuerung und Kalibrierung von Automaten und Robotern erhebliche Probleme darstellen. Den Einsatz von genetischen Algorithmen für die Bearbeitung dieser Aufgabenstellungen demonstriert Tabelle 3.11.

Zuletzt sind in Tabelle 3.12 noch einige Anwendungen aus diversen Fachgebieten aufgeführt, in denen genetische Algorithmen als Simulation oder zur Unterstützung einer Simulation eingesetzt werden, so dass sie zur Basisstufe einer Meta-Simulation werden. Zu den Anwendungsbereichen gehören künstliche ökonomische Systeme, Aufrüstungssimulationen und Erdbebensimulationen. Zu einer interessanten Verdopplung führt zudem der Einsatz genetischer Algorithmen bei der Simulation evolutionärer Dynamiken. Damit präsentieren sich die Anwendungsfelder in großer Breite. Die Erfahrungen der beschriebenen Disziplinen sollten auch Psychologen und Psychologinnen motivieren können, den methodischen Baukasten zu vergrößern.

3.7.2 Anwendungen von Partikelschwärmen

An dieser Stelle sollen einige Anwendungen von Partikelschwarmalgorithmen dargestellt werden, um die Bandbreite der Einsatzgebiete, aber auch die der eingesetzten Varianten des Algorithmus zu beleuchten. Dabei wird ein besonderer Fokus auf die Setzung der Steuerungsparameter gelegt. Einige Varianten verwenden Erweiterungen des basalen Algorithmus, die im Anhang vorgestellt werden.

3.7.2.1 Statistische Klassifikation

Die Möglichkeit, Partikelschwärme zur statistischen Klassifikation einzusetzen, wird von Sousa, Silva und Neves (2004) diskutiert. Die Autoren wenden sechs verschiedene Partikelschwarm-Algorithmen auf drei Datensätze (*Zoo*, *Breast Cancer*, *Wisconsin-Breast-Cancer*) an, die vom Department of Computer Science, University of Waikato, Hamilton, New Zealand und Department of Information and Computer Science, University of California zur Verfügung gestellt wurden. Dabei wird im Sinne des Michigan-Ansatzes durch ein Individuum jeweils eine Regel kodiert, die sich aus einer Konjunktion von (Variable=Wert)-Bedingungen zusammensetzt. Klassifizierte Fälle werden aus dem Datensatz entfernt. So wird sequenziell ein Regelsatz kreiert, bis weniger als 10% des Datensatzes zur Klassifikation verbleiben, für die eine letzte Regel hinzugenommen wird, die alle verbleibenden Fälle der Majoritätsklasse zuweist. Als Fitnesswert wird das Produkt aus Sensitivität und Spezifizität gewählt mit einer Penalty-Modifikation für Regeln, deren Werte sich außerhalb des zulässigen Bereichs befinden. Anschließend werden unnötige Regeln aus dem Regelsatz entfernt. Regeln werden gelöscht, wenn sie der letzten Regel vorangehen und die gleiche Klasse vorhersagen. Streng genommen könnte die Ökonomie weiter erhöht werden, wenn aufeinanderfolgende Regeln mit gleichem Klassifikationsergbnis zusammengefast würden durch entsprechende Oder-Verknüpfungen.

Die sechs Algorithmen, die in der Studie verwendet werden sind die folgenden:

- Discrete Particle Swarm Optimization (DPSO): Diese Variante operiert in einem nicht-kontinuierlichen Suchraum und nutzt Binärstrings als Repräsentationen

- Constricted Particle Swarm Optimization (CPSO): Diese Variante operiert in einem kontinuierlichen Suchraum unter Verwendung des Simple Constriction Coefficient.

- Linear Decreasing Weight Particle Swarm Optimization (LDWPSO): Diese Variante operiert ebenfalls in einem kontinuierlichen Suchraum. Welches Gewicht sich linear verringert, wird im Artikel nicht genannt.

Zur Parametersetzung findet sich allein die Bemerkung, dass $\varphi_{b_{max}}$ und $\varphi_{n_{max}}$ auf 2,05 gesetzt werden, χ auf 0,72. Dabei bleibt unbekannt, wie genau sich der dritte Algorithmus vom zweiten unterscheidet und welche Nachbarschaftstopologie gewählt wurde, da zwar von Nachbarschaften mit $k = 2$ oder $k = all$ gesprochen wird, die Abkürzung P_g für den besten Punkt in der Nachbarschaft aber auch auf ein globales Modell hinweisen könnte. Zur Initialisierung finden sich keine Angaben, das Abbruchkriterium wird als eine Abfolge von 30 Iterationen angegeben, für die es keine Egebnisverbesserung gibt. Diese Iterationen werden von den Autoren *convergence platform* genannt. Wie sichergestellt wurde, dass der LDWPSO-Ansatz nicht vor Erreichen der Plattform zum Stillstand kommt, wird nicht angegeben.

3.7.2.2 Stromversorgung

Ein Problem im Gebiet des *Least-Cost Generation Expansion Planning* wird von Kannan, Slochanal, Subbaraj und Padhy (2004) mit sechs verschiedenen Schwarmalgorithmen angegangen. Das Problem besteht aus der Frage, welche und wie viele Generatoren langfristig angeschafft werden sollen unter Berücksichtigung von Versorgung und verschiedenen Generatortypen. Das Problem ist nichtlinear, diskret, mit vielen Nebenbedingungen versehen und komplex. Als Bewertungskriterium für Lösungen wird eine zu minimierende Kostenfunktion aufgestellt.

Die Autoren stellen sechs verschiedene PSO-Algorithmen vor:

- *Simple PSO* (PSO): Dieser Algorithmus entspricht dem Standard-Partikelschwarm mit globaler Nachbarschaftstopologie und Verwendung eines Inertia-Faktors, der über die Iterationen linear abnimmt.

- *Constriction Factor Approach* (CFA): In diesem Modell wird anstelle des Inertia-Faktors ein Constriction-Koeffizient verwendet, der auf einen festen Wert gesetzt wird. An anderer Stelle wird trotzdem ein Inertia-Faktor für dieses Modell angegeben, so dass der tatsächlich eingesetzte Algorithmus unklar bleibt.

- *Lbest model*: Bei diesem Modell wird anstelle der globalen Nachbarschaftstopologie ein lokales Modell eingesetzt. Die Anzahl der Nachbarn soll dabei der Problemschwierigkeit entsprechend gewählt werden.

- *Hybrid PSO* (HPSO): Dieses Modell ist eine Mischform aus einem Partikelschwarm und dem Selektionsoperator eines evolutionären Algorithmus. Nach jeder Iteration werden die Positionen der Partikel mit den schlechtesten Lösungen auf die Positionen von Partikeln mit besseren Lösungen gesetzt. Die Geschwindigkeiten werden dabei nicht angepasst, so dass sich die Exploration in verschiedene Richtungen fortsetzen kann.

- *Stretched PSO* (SPSO): Dieser Algorithmus verwendet eine Transformation der Fitnessfunktion, die gezielt dazu dienen soll, dass Partikel lokalen Optima im Suchraum entkommen können.

- *Composite PSO* (C-PSO) mit *Differential Evolution* (DE): Bei diesem Algorithmus wird die Festlegung von $\varphi_{b_{max}}$, $\varphi_{i_{max}}$ und α evolutionär vollzogen. Dabei wird vor jedem Bewegungsschritt eine Population von Parameterwerttriplets erzeugt, und es werden Mutations-, Rekombinations- und Selektionsoperatoren eingesetzt.

Die Initialisierung erfolgt randomisiert sowohl für Positionen als auch für Geschwindigkeiten. Ein Penalty-Faktor wird benutzt, um nicht umsetzbare Lösungen zu bestrafen. Als Abbruchkriterium wird eine feste Anzahl von Iterationen vorgegeben. Die Parametersetzungen sind in präziser Art und Weise für jeden der sechs Algorithm tabellarisch aufgelistet. Dabei werden alle φ_{max}-Werte auf $2,2$ gesetzt, die Populationsgröße auf 20, für den HPSO-Algorithmus auf 50. Der *Constriction Factor* nimmt für den CFA-Algorithmus den Wert $0,8$ an. Der Parameter $\alpha(0)$ hat den Wert $0,8$, der Parameter $\alpha(t_{max})$ den Wert $0,2$. Der Inertia-Faktor nimmt in den Modellen linear ab. Das Lbest-Modell wird mit einer Nachbarschaftsgröße von 3 initialisiert, die Anzahl der Iterationen wird auf 200 festgelegt, für den C-PSO-Algorithmus auf 100.

In der Diskussion der Ergebnisse kommen die Autoren zu dem Schluss, dass das CFA-Modell insgesamt die besten Ergebnisse erbringt. Unter der Bedingung vieler lokaler Minima stellen sich hingegen auch das Lbest-Modell und das SPSO-Modell als vorteilhaft dar.

3.7.2.3 Projektplanung

Eine Aufgabenstellung im Bereich der Projektplanung gehen Zhang, Li, Li und Huang (2005) mit einem Partikelschwarmalgorithmus an. Beim *Project Scheduling* wird festgelegt, welche Aktivitäten zu welchen Zeitpunkten ausgeführt werden müssen, um eine Projektaufgabe zu erfüllen. Die in diesem Bereich eingesetzten Standardverfahren machen die Annahme unbegrenzter Ressourcen, was ihre breite Anwendung verhindert.

Die Autoren vergleichen zwei Patikelschwarmalgorithmen, die sich in ihrer Repräsentation des Problems unterscheiden, einen prioritätszentrierten und einen permutationszentrierten Ansatz. Der prioritätszentrierte Ansatz sucht nach einer optimalen Setzung von Prioritätswerten für Aufgaben. Der Suchraum hat so viele Dimensionen, wie es Aufgaben gibt, und die Positionen entsprechen einer bestimmten Kombination von Prioritätswerten für alle Aufgaben. Im permutationszentrierten Ansatz werden Lösungen repräsentiert durch ein N-Tupel natürlicher Zahlen aufsteigend von 1 bis N.

Für den prioritätszentrierten Ansatz werden beide φ_{max}-Werte auf 1 gesetzt, zudem wird ein neutrales *Inertia*-Gewicht von $\alpha = 1$ und $V_{max} = 1$ eingesetzt. Die Populationsgröße wird auf eine Zahl nahe der Aufgabenzahl gelegt. Der Unterschied im permutationszentrierten Ansatz ist die Wahl eines *Inertia*-Gewichtes von $\alpha = 0,8$ und $V_{max} = Aufgabenzahl$. Der Algorithmus verwendet ein globales Optimum, die Partikelpositionen und Geschwindigkeiten des Anfangszustand werden im zulässigen Bereich randomisiert erzeugt.

In der Beispielaufgabe wird mit einer Population von 40 Partikeln in einem 30-dimensionalen Suchraum gearbeitet über 125 Iterationen. Die Ergebnisse zeigen, dass der permutationsbasierte Ansatz dem prioritätsbasierten Ansatz überlegen ist und auch gegen die Vergleichsalgorithmen bestehen kann.

3.7.2.4 Strukturelle Reliabilität

Dem Problem der Maximierung der strukturellen Reliabilität eines Systems widmet sich eine Studie von Elegbede (2005). Systemfehlerzustände werden dabei als negative Funktionswerte einer Zufallsvariablen mit bekannter Verteilung modelliert, so dass das Integral der Funktion die Wahrscheinlichkeit des Strukturversagens entspricht. Die Dimensionszahl des Vektors ergibt sich aus der Anzahl der Variablen, über die Unsicherheit besteht. Gesucht wird für die Berechnung des Hasofer-Lind-Reliabilitätsindex der Designpunkt, mit dem eine Approximation der Strukturversagenswahrscheinlichkeit möglich wird.

Die Partikel, deren Position Variablenwerten entspricht, werden randomisiert initialisiert mit einer Anfangsgeschwindigkeit von 0. Das Modell ist ein globales Schwarmmodell mit einem Inertia-Faktor von $\alpha = 0,5$. Die Gewichtungsparameter werden auf $\varphi_{b_{max}} = 1,5$ und $\varphi_{n_{max}} = 1,2$ gesetzt. Die Populationsgröße ist 70, die Iterationszahl 990. Zur Kontrolle der Geschwindigkeit wird ein Mechanismus eingeführt, der leider unzureichend beschrieben ist und die Formulierung enthält, dass Geschwindigkeiten sich außerhalb des Suchraums befinden können, die im Artikel kryptisch bleibt. Eine Analyse des Konvergenzverhaltens deutet darauf hin, dass eine Iterationszahl von 300 ausreichend gewesen wäre und die Populationsgröße ab 20 zufriedenstellende und ab 30 fast optimale Ergebnisse erbringt. Die Autoren sind mit den Möglichkeiten des Schwarmalgorithmus zufrieden:

> The PSO algorithm is very efficient to solve global optimization problems with continuous variables. Its use in structural reliability field presents not only the advantage of its facility of implementation, but also the possibility to obtain the design point and the failure probability with a good accuracy. (Elegbede, 2005, S. 185)

3.7.2.5 Weitere Anwendungen

Eine Anwendung im Bereich der Chemietechnik steuern Ourique, Biscaia und Pinto (2002) bei. Hier werden Partikelschwärme eingesetzt, um Regionen mit komplexen dynamischen Phänomenen im Modellparameterraum zu identifizieren, um eine klassische Bifurkationsanalyse vorzubereiten. Die Positionen der Partikel repräsentieren Festlegungen von Modellparametern, die Bewertungsfunktion ergibt sich aus

dem Vergleich von gewünschter und erhaltener Komplexität der Lösung. Dabei können Lösungen stabile Zustände sein, Oszillationen verschiedener Komplexität oder auch Chaos.

Es werden 20 Partikel mit einer globalen Topologie eingesetzt bei einer Iterationszahl von 100 für zweidimensionale Suchräume. Für dreidimensionale Suchräume kommen 40 Partikel zum Einsatz. Die Positionen und Geschwindigkeiten der Partikel zu Beginn des Laufs werden randomisiert erzeugt, wobei die Höhe der Geschwindigkeiten maximal einem Viertel der zulässigen Spannweite für die Parametersetzungen entsprechen darf. Der Inertia-Faktor wird auf $\alpha = 0,9$ gesetzt, $\varphi_{b_{max}}$ und $\varphi_{n_{max}}$ nehmen den Wert 2 an.

Die Autoren kommen zu dem Ergebnis, dass der vorgeschlagene Algorithmus zu einer wesentlichen Zeitersparnis führt bei der Suche nach Regionen mit periodischen oder chaotischen Oszillationen im Modellparameterraum. Dabei wird das Ergebnis des Algorithmus als Richtlinie und erste Orientierung für zeitaufwändigere Analysen genutzt.

Die von Salman, Ahmad und Al-Madani (2002) in Angriff genommene Aufgabe der Zuteilung der Tasks eines Programm zu Prozessoren einer verteilten Rechnerarchitektur ist ein NP-vollständiges Problem. Die Partikelposition im vorgestellten Schwarmmodell entsprechen Zuordnungen von Tasks zu Prozessoren, wobei der Suchraum so viele Dimensionen wie vorhandene Tasks besitzt und die möglichen Werte auf diesen Dimensionen natürliche Zahlen zwischen 1 und der Anzahl der berücksichtigten Prozessoren sind. Dementsprechend repräsentiert ein Partikel die Zuordnung aller Tasks.

Das Modell ist ein globales Schwarmmodell mit einem *Inertia*-Faktor von $\alpha = 0,9$, einer Populationsgröße und maximale Iterationszahl, die der doppelten Taskanzahl entsprechen und Werten von $\varphi_{b_{max}} = 1$ und $\varphi_{n_{max}} = 1$. Die Anfangspopulation wird randomisiert erzeugt. Die Testprobleme haben Taskzahlen zwischen 10 und 169, dementsprechend also Populationsgrößen zwischen 20 und 338. In den Ergebnissen setzt sich der Partikelschwarm gegen einen Genetischen Algorithmus durch und zeichnet sich darüber hinaus durch kürzere Laufzeiten aus.

Mit einem Problem im Bereich des Antennenarray-Designs beschäftigen sich Gies und Rahmat-Samii (2003). Sie benutzen einen Partikelschwarm mit globaler Topologie, 20 Partikeln, $\varphi_{b_{max}} = 2$ und $\varphi_{n_{max}} = 2$. Die beschriebene Geschwindigkeitsdämpfung scheint ein Inertia-Faktor zu sein, der linear von $\alpha(1) = 0,9$ zu $\alpha(t_{max}) = 0,4$ abnimmt. Die Konvergenzanalyse zeigt befriedigende Lösungen nach 300 bis 1000 Zeitschritten, je nach Problemstellung.

Eine clusteranalytische Anwendung präsentieren Van der Merwe und Engelbrecht (2003). Hier wird ein globaler Partikelschwarmalgorithmus mit 10 Partikeln, $\varphi_{b_{max}} = 1,49$, $\varphi_{n_{max}} = 1,49$, $\alpha = 0,72$ und 1000 Zeitschritten eingesetzt. Jeder Partikel repräsentiert dabei eine Menge von Clusterzentroiden. Der Algorithmus wird bei mehreren Clusterproblemen mit dem *K-means*-Algorithmus verglichen. Zusätzlich wird eine hybride Methode eingesetzt, deren Initialisierung durch einen *K-means*-Algorithmus übernommen wird. Die Autoren empfehlen den Partikelschwarmalgorithmus als Alternative zur *K-means*-Methode.

Einen Beitrag zur Genomanalyse leisten Chang, Ratnaweera, Halgamuge und Watson (2004). Bei der Suche nach übereinstimmenden Mustern repräsentieren Partikel rigide oder flexible Proteinsequenzen, die in einem ersten Schritt generiert und in einem zweiten Schritt optimiert werden. Dabei werden 100 Partikel eingesetzt. Leider verhindern syntaktische Fehler im angegebenen Pseudo-Code und Auslassungen von Wertbelegungen bzw. widersprüchliche Angaben eine genauere Beschreibung des verwendeten Algorithmus. Um unmögliche Lösungen zu verhindern, wird ein Partikel bei einem Verlassen des gültigen Intervalls auf die jeweils entgegengesetzte Grenze des zulässigen Intervalls versetzt. Das Verfahren kann nach Ansicht der Autoren mit den üblicherweise eingesetzten (PROSITE und Neuro-Fuzzy-Optimierung) konkurrieren.

Fukuyama, Takayama, Nakanishi und Yoshida (1999) bearbeiten ein Problem in der Kontrolle von Reaktoren mit einem simplen Partikelschwarmmodell mit 10 Agenten und $\varphi_{b_{max}} = 1$ und $\varphi_{n_{max}} = 1$. Das globale Modell konvergiert bei einem neundimensionalen Problem in 44 Iterationen.

Die Platzierung von Kondensatoren zur Minimierung des Energieverlusts untersuchen Yu, Xiong und Wu (2004). Dabei wird ein Partikelschwarmalgorithmus genutzt, um das Ladungsniveau einer Reihe von Kondensatoren einzustellen. Dabei kommt ein Constriction-Faktor von $\chi = 0,729$ zum Einsatz bei $\varphi_{b_{max}} = 2,05$ und $\varphi_{n_{max}} = 2,05$, 50 Partikeln und Iterationen in einem globalen Modell. Die Ergebnisse des Algorithmus werden positiv bewertet:

> From the viewpoint of researchers who employ this algorithm to solve engineering problems, the main attractions of PSO are simple concept, easy parameter tuning, and good convergence.

> Even for a large-scale system, PSO can achieve the global optimal or near-optimal solution only through less than 50 iterations. (Yu et al., 2004, S. 33)

3.7.2.6 Überblick und Diskussion

Die dargestellten Anwendungen von Partikelschwärmen verweisen zum einen auf die Bandbreite des Aufgabenspektrums, das von Partikelschwarmalgorithmen erfolgreich angegangen werden kann, sie demonstrieren aber ebenso deutlich die zur Zeit bestehenden Probleme in der Darstellung von Forschungsergebnissen mit Schwärmen. Wünschenswert wäre zum einen eine Vereinheitlichung der Terminologie. Für den Inertia-Faktor sind beispielsweise die Abkürzungen α, w, W und ω im Einsatz, für das individuelle Beschleunigungsgewicht φ_1, $C1$ und c_1, um nur einige Varianten zu nennen.

Gravierender aber stellt sich das Problem fehlender Angaben dar. Um die Replikation eines Experiments zu ermöglichen ist die Spezifizierung aller Teilschritte des Algorithmus erforderlich. Diese notwendigen Informationen werden aber nicht von allen Autoren berichtet, was zu einem Problem führt, wenn ihr Beitrag produktiv genutzt werden soll. Ebenso wird die Repräsentation in vielen Fällen unzureichend verdeutlicht, so dass auch der fachspezifische Beitrag in die Gefahr gerät, aufgrund fehlender Angaben nicht genutzt werden zu können. Eine Vereinheitlichung der Darstellung sowie eine Standardisierung der notwendigen Angaben scheint dementsprechend unumgänglich, um eine bessere Bündelung und Verzahnung der Forschung zu gewährleisten.

Vergleicht man die gewählten Ansätze (s. Tabelle 3.13 und Tabelle 3.14), wird das Spektrum möglicher Parametersetzungen deutlich. Eine Bestimmung optimaler Parameter ist nicht zu erwarten, es sei denn durch Studien, die künstlich generierte Datensätze extensiv testen (Trelea, 2003). Auch in diesem Fall wäre aber eine Übertragung auf Datensätze mit unbekannter Struktur nicht leistbar. Dementsprechend bleiben die Entscheidungen für Modell und Parameter beim Anwender, dessen Intuition und Abschätzung von Suchraum, Dimensionalität und Problemschwierigkeit die Wahl unterstützen.

Es fällt insgesamt aber auf, dass wenige Partikel für die meisten Suchaufgaben ausreichen und auch die Iterationenzahl in Grenzen bleibt. Da die Laufzeit des Algorithmus abhängig ist von der Repräsentation und dem Aufwand der Lösungsprüfung, kann trotzdem keine generelle Aussage über den Zeitbedarf des Algorithmus gemacht werden, in spezifischen Fällen konnte hingegen die zeitliche Ökonomie von Schwarmalgorithmen gezeigt werden. Bei Vergleichen mit Standardverfahren erzielt der Algorithmus selten schlechtere, in einigen Fällen bessere Ergebnisse, so dass er auch in Anbetracht der Konkurrenz als ernst zu nehmende Alternative zu sehen ist.

Tabelle 3.13: Quellen und Namen der vorgestellten Schwarmalgorithmusapplikationen mit Angaben zu Methode, Parameterwahlen und Besonderheiten

ID	Quelle	Methode	constriction	inertia	Gbest	Lbest	Positionen	Geschwindigkeiten	Besonderheiten
1	Fukuyama et al. (1999)	PSO	nein	nein	ja	nein	random	random	keine
2	Ourique et al. (2002)	PSO	nein	ja	ja	nein	random	random (25%)	keine
3	Salman et al. (2002)	PSO	nein	ja	ja	nein	random	random	keine
4	Gies und Rahmat-Samii (2003)	PSO	nein	ja	ja	nein	random	random	keine
5a	Van der Merwe und Engelbrecht (2003)	PSO	nein	ja	ja	nein	random	random	keine
5b	Van der Merwe und Engelbrecht (2003)	K-means/PSO	nein	ja	ja	nein	K-means	random	Initialisierung durch K-means
6	Chang et al. (2004)	PSO	nein	???	ja	nein	random	random	Reinitialisierungsbeschl.
7a	Kannan et al. (2004)	PSO	nein	ja	ja	nein	random	random	keine
7b	Kannan et al. (2004)	CFA	ja	???	ja	nein	random	random	keine
7c	Kannan et al. (2004)	lbest	nein	ja	nein	ja	random	random	keine
7d	Kannan et al. (2004)	hybrid PSO	nein	ja	ja	nein	random	random	Selektionsoperator
7e	Kannan et al. (2004)	SPSO	nein	ja	ja	nein	random	random	Transformation Fitness
7f	Kannan et al. (2004)	CPSO-DE	nein	ja	ja	nein	random	random	Selektionsoperator
8a	Sousa et al. (2004)	DPSO	ja	nein	???	???	???	???	Diskrete Repräsentation
8b	Sousa et al. (2004)	CPSO	ja	nein	???	???	???	???	keine
8c	Sousa et al. (2004)	LDWPSO	ja	nein	???	???	random	random	Angaben fehlen
9	Yu et al. (2004)	PSO	ja	nein	ja	nein	random	random	keine
10	Elegbede (2005)	PSO	nein	ja	ja	nein	random	0	Geschwindigkeitskontrolle
11a	Zhang et al. (2005)	Prio-PSO	nein	ja	ja	nein	random	random	Prioritätsbasiert
11b	Zhang et al. (2005)	Perm-PSO	nein	ja	ja	nein	random	random	Permutationsbasiert

Tabelle 3.14: Parameterwerte der vorgestellten Schwarmalgorithmusapplikationen

ID	Quelle	Methode	$\varphi_{b_{max}}$	$\varphi_{g_{max}}$	Constriction	V_{max}	Inertia	Populationsgröße	Abbruch	Dimensionen
1	Fukuyama et al. (1999)	PSO	1,00	1,00	n.v.	n.v.	n.v.	10	convergence	9
2	Ourique et al. (2002)	PSO	2,00	2,00	n.v.	n.v.	0,9	20 bis 40	100	2 bis 3
3	Salman et al. (2002)	PSO	1,00	1,00	n.v.	n.v.	0,9	20 bis 338	20 bis 338	10 bis 169
4	Gies und Rahmat-Samii (2003)	PSO	2,00	2,00	n.v.	n.v.	0,9-0,4	20	1500	40
5a	Van der Merwe und Engelbrecht (2003)	PSO	1,49	1,49	n.v.	n.v.	0,72	10	1000	4 bis 39
5b	Van der Merwe und Engelbrecht (2003)	K-means/PSO	1,49	1,49	n.v.	n.v.	0,72	10	1000	4 bis 39
6	Chang et al. (2004)	PSO	???	???	n.v.	n.v.	???	100	???	100
7a	Kannan et al. (2004)	PSO	2,20	2,20	n.v.	n.v.	0,8-0,2	20	200	4 bis 8
7b	Kannan et al. (2004)	CFA	2,20	2,20	0,800	n.v.	???	20	200	4 bis 8
7c	Kannan et al. (2004)	lbest	2,20	2,20	n.v.	n.v.	0,8-0,2	20	200	4 bis 8
7d	Kannan et al. (2004)	hybrid PSO	2,20	2,20	n.v.	n.v.	0,8-0,2	50	200	4 bis 8
7e	Kannan et al. (2004)	SPSO	2,20	2,20	n.v.	n.v.	0,8-0,2	20	200	4 bis 8
7f	Kannan et al. (2004)	CPSO-DE	2,20	2,20	n.v.	n.v.	DE	20	100	4 bis 8
8a	Sousa et al. (2004)	DPSO	2,05	2,05	0,730	n.v.	n.v.	25 bis 300	convergence	???
8b	Sousa et al. (2004)	CPSO	2,05	2,05	0,730	n.v.	n.v.	25 bis 300	convergence	???
8c	Sousa et al. (2004)	LDWPSO	2,05	2,05	0,730	n.v.	n.v.	25 bis 300	50	???
9	Yu et al. (2004)	PSO	2,05	2,05	0,729	n.v.	n.v.	50	50	???
10	Elegbede (2005)	PSO	1,50	1,20	n.v.	call back to center	0,75	70	990	2 bis 6
11a	Zhang et al. (2005)	Prio-PSO	1,00	1,00	n.v.	1	1	40	125	30
11b	Zhang et al. (2005)	Perm-PSO	1,00	1,00	n.v.	1	0,8	40	125	30

A cartoon shows an early Homo sapiens standing in front of a cave. He is calculating the trajectory of a lion's jump and the magnitude of the impulse the lion will have in order to decide what to do. The last picture shows a sated, happy lion.

(Gigerenzer, 2001c, S. 113)

4

Grenzen der Rationalität und das Forschungsprogramm der Adaptiven Werkzeugkiste

Im Gegensatz zu den vorangehenden Kapiteln werden in diesem Kapitel Anleihen an die Philosophie zu erwarten sein. Die Wahl dieser Orientierung ist motiviert aus der Wahrnehmung eines Defizits innerhalb der eigenen Disziplin, in der Rationalität zumeist sehr eingeschränkt betrachtet wird als ein untergeordneter Aspekt des Menschen. Diese Beschränkung selbst wird allerdings selten benannt, so dass Schreier (2003) kritisieren kann:

> Während Rationalität in der Philosophie [...] eine zentrale Kategorie darstellt, die den gesamten abendländischen Strang dieser Disziplin entscheidend geprägt hat [...], ist die Rationalität innerhalb des Faches der Psychologie von vergleichsweise untergeordneter Bedeutung: Der Begriff der Rationalität im engeren Sinne erscheint in der psychologischen Fachliteratur eher selten; fasst man unter 'Rationalität' im weiteren Sinne die verschiedenen Formen menschlichen Denkens, Entscheidens und Problemlösens, so findet sie sich unter dem Stichpunkt 'Kognition' als Teilbereich der Allgemeinen Psychologie wieder, als ein Gegenstand neben anderen wie etwa Emotion, Motivation oder Wahrnehmung, dem aber in keiner Weise herausragende Bedeutung zugemessen wird. Dies überrascht insofern, als es sich bei der Rationalität um eine basale anthropologische Kategorie handelt, die nicht selten zur Abgrenzung des Menschen gegenüber dem Tier herangezogen wurde, für den Gegenstand der Psychologie somit als konstitutiv gelten kann. (S. 107)

Schreier sieht die Ursache des Defizits strukturell in der Geschichte der Disziplin begründet, die im 20. Jahrhundert mit der Akzeptanz des Weberschen Werturteilsfreiheitspostulats einem hochgradig normativen Begriff wie „Rationalität“ skeptisch gegenüber stehen muss. Es könnte hinzugefügt werden, dass normative Diskurse in der Psychologie schnell in den Metaphysikverdacht geraten, so dass ein breiter Rationalitätsdiskurs im Zuge der nicht unproblematischen Operationalisierung und Naturalisierung der eigenen Grundkonzepte nicht zuletzt aufgrund seiner selbstreflexiven Sprengkraft gescheut worden ist. Während nach Schreier (2003) in der neuzeitlichen philosophischen Diskussion einer „grundsätzlichen Fähigkeit des Menschen zur Rationalität“ das Wort geredet wird, gibt und gab es in der Geschichte der Psychologie zahlreiche Tendenzen, diese Annahme zu unterhöhlen oder auszuräumen:

> In der Psychologie [...] ist dem Menschen die Rationalität in verschiedenen Bedeutungsvarianten wiederholt abgesprochen worden: sei es im Rahmen reduktionistischer Menschenbildannahmen wie jenen des Behaviorismus [...] oder der Psychoanalyse [...] - sei es im Rahmen kognitionspsychologischer Untersuchungen, als deren Ergebnis immer wieder konstatiert wird, dass menschliches Denken gemessen an den Normen der Logik und Wahrscheinlichkeitstheorie grundsätzlich fehlerbehaftet und damit irrational sei [...]. (Schreier, 2003, S. 108)

Dieser Umstand wird als Anlass genommen, die Gründe für die unterschiedlichen Standpunkte zu untersuchen, aber dabei zunächst einmal das zur Debatte stehende Konzept zu betrachten, das die beschriebenen Diskurse in schillernden Formen durchzieht und sich trotz seiner grundsätzlich positiven Konnotation in einer normativen Funktion zu verantworten hat. Der Blickwinkel wird in diesem Kapitel weniger in die ökonomische Richtung wandern, da eine große Zahl von Ökonomen das folgende Zitat von Broome (1990) keineswegs als fehlgeleitet bewerten würden:

> And that is my impression of many of he economists working on decision theory. They are looking for a theory that fits behavior. They find the question of rationality uninteresting and perhaps empty. (S. 135)

Psychologen wiederum, die die beschriebene Ansicht teilen, können in diesem Kapitel nur um Verständnis dafür gebeten werden, dass der Arbeit an Begriffen im Sinne einer Grundlegung der wissenschaftlichen Bewegungsfähigkeit und Schnittstelle zwischen den Systemen der Wissenschaft und menschlichen Lebenswelten vom Verfasser ein größerer Stellenwert eingeräumt wird. Zunächst soll ein Überblick gegeben werden über einige Spannungen, die das Konzept der Rationalität betreffen und unterschiedliche - bis hin zu inkompatiblen - Auffassungsarten zur Konsequenz haben.

4.1 Arten der Rationalität

Am Anfang dieser Zusammenstellung zum Konzept der Rationalität soll die Diskussion grob gerahmt werden, indem einige begriffsgeschichtliche Aspekte und Grundunterscheidungen in der Rationalitätsdebatte erläutert werden.

4.1.1 Einige begriffliche Unterscheidungen

In der abendländischen Philosophie des Mittelalters wird eine Spaltung zwischen dem schlussfolgernden, logischen Denken (*ratio*) und dem Vermögen übersinnlicher Erkenntnis (*intellectus*) eingeführt (Karafyllis & Schmidt, 2002). Die *ratio* wurde fortan mit dem Verstand gleichgesetzt, der in ein Spannungsfeld zum Glauben (*fides*) trat, der *intellectus* hingegen mit der Vernunft im Sinne des griechischen *νόυς*: „Die Vernunft vermochte den Gesamtsinn der ratio zu erfassen“ (Karafyllis & Schmidt, 2002, S. 10).

In der Aufklärung wurde der Begriff der Vernunft durch Kant der Tätigkeit des Verstandes zugeordnet, die Verbindung zwischen sensorischen Wahrnehmungen und konzeptuellen Begriffen herzustellen. Dementsprechend wird die Vernunft „zum Kennzeichen der jüngeren philosophischen Traditionen des Denkens und Denkenkönnens“ (Karafyllis & Schmidt, 2002, S. 10) In der Bezugnahme auf „intellektuelle Fähigkeiten“ zeigt sich die Auflösung der ursprünglichen Grenzen im heutigen Sprachgebrauch. Gosepath (2002) bezeichnet den Begriff „Vernunft“ im Kontrast mit „Rationalität“ als das „emphatischere, aber deshalb zu vielen Missverständnissen Anlass gebende Wort“. Auch aus diesem Grund wurde in dieser Arbeit der Begriff „Rationalität“ vorgezogen. Für die Konzeption der Aufklärung zentral ist die Einordnung des Menschen als „vernunft*begabtes*, jedoch nicht selbstverständlich als vernunft*innehabendes* Wesen“ (Karafyllis & Schmidt, 2002, S. 10): So wird „der Mensch sich so selbst zur Aufgabe“, d.h. dem Streben danach, rational zu handeln und zu denken wird eine Erfolgsmöglichkeit eingeräumt, die gleichzeitig wiederum zu einer Verpflichtung führt. Der Begriff des *animal rationabile* wird dem Begriff des *animal rationale* gegenüber gestellt, der Mensch ist nicht von Natur aus notwendigerweise vernünftig, er kann aber vernünftig werden.

Eine Unterscheidung in *theoretische* und *praktische* Rationalität kann bis Aristoteles zurückverfolgt werden (Audi, 2004). Dabei wird die Rationalität von Kognitionen (so etwa Glaubenssätze, aufgrund derer ein wahres Bild der Welt gesucht wird) von der Rationalität von Elementen wie Handlungen getrennt (durch die Menschen zu praktischen Wesen werden). Smith (2004) bezeichnet den theoretischen Bereich als den Bereich der Formierung von Glaubenssätzen, den praktischen Bereich als den Bereich der Formierung von Wünschen (die eine Person zu Handlungen disponieren, bei denen Weltzustände verändert werden können). Gosepath (2002) unterscheidet zwischen einer *epistemischen*, d.h. auf Wahrheit ausgerichteten Rationalität und einer *prudentiellen*, d.h. auf Nützlichkeit ausgerichteten Rationalität. Für ihn bricht die Unterteilung in theoretische und praktische Rationalität allerdings auseinander. Zum einen ist für ihn Wahrheit kein erschöpfendes Kriterium für die theoretische Meinungsbildung:

> Es scheint unsinnig und unmöglich, alles glauben zu wollen, von dem man Grund hat zu glauben, dass es wahr ist. Wir müssen die Proposition auswählen, die wir glauben wollen. Als Auswahlkriterium fungiert dabei das jeweilige Erkenntnisinteresse. Um einen motivierenden Grund zu haben, eine wahre Proposition zu glauben, bedarf es noch zusätzlich eines Interesses an ihrem Informationsgehalt. (S. 46)

Wird Rationalität als ein Dispositionsprädikat verstanden, so sollte diese Einordnung auch für den Begriff der Irrationalität vorgenommen werden. Trotzdem ist das Verhältnis zwischen beiden Begriffen nicht als eine einfache Kontradiktion zu behandeln. Es sollen einige Gründe angeführt werden, warum auch hier von einer Spannung gesprochen wird, die sich auf der begrifflichen Ebene befindet und nicht mit einer Ambiguität in der Zuschreibung im konkreten Fall identifiziert werden kann. Zur Frage des Verhältnisses zwischen Rationalität und Irrationalität soll ein Ansatz Searles (Searle, 2001) vorgestellt werden. Searle bezweifelt einige Postulate, die er als Orthodoxie des Rationalitätsbegriffes versteht. Zum Beispiel bestreitet er, dass rationale Handlungen durch Überzeugungen und Wünsche verursacht werden, eine direkte Verursachung durch diese sieht er lediglich bei - in seinen Worten - „irrationalen Handlungen“. Die Annahme von Willensfreiheit kann nur bei Annahme eines Spalts zwischen den Prozessen des Erwägens und des Handelns aufrechterhalten werden. Tatsächlich sieht er sogar eine dreifache Spaltung: Der erste ist der Spalt zwischen den erwogenen Gründen und der Entscheidung, der zweite zwischen der Entscheidung und der korrespondierenden Handlung, der dritte zwischen dem Beginn einer Handlung und ihrer Fortführung bis zur Vervollständigung. In keinem Schritt sieht er eine kausale Verursachung, die nicht mediatisiert wird mit der Möglichkeit eines alternativen Weges. Die Spaltung sieht er als konstitutiv für einen Begriff von Rationalität:

> An understanding of the gap is essential for the topic of rationality because rationality can only operate in the gap. Though the concept of freedom and the concept of rationality are quite different, the extension of rationality is exactly that of freedom. The simplest argument for this point is that rationality is only possible where irrationality is possible, and that requirement entails the possibility of choosing between various rational options as well as irrational options. The scope of that choice is the gap in question. (Searle, 2001, S. 316f)

Als Beispiel für eine notwendige Zuschreibung von Irrationalität sieht er die *bewusste* Entscheidung eines Menschen, die Hand vor seinen Augen trotz genügenden Abstands und guter Beleuchtung nicht sehen zu wollen, obwohl er aufgrund seiner Sinneswahrnehmung nicht anders kann, als diese Hand zu sehen. Für das Betrachten der Hand selbst gibt es den angesprochenen Spalt für ihn nicht, weshalb es keine Möglichkeit der Zuschreibung von Rationalität oder Irrationalität für die Wahrnehmung gibt. Der Spalt öffnet den Raum für das Konzept sowohl in praktischen als auch in theoretischen Fragestellungen:

> You can only have rationality where you have the possibility of irrationality. And with just sheer, raw perceptions, you do not get rationality or irrationality. They only come into play where you have a gap, where the existence of the intentional phenomena by themselves is not sufficient to cause the outcome, and these are cases where you have to decide what you are going to do or think. (Searle, 2001, S. 320)

In dieser Perspektive gibt es eine Klasse von Phänomenen, bei denen entweder Rationalität oder Irrationalität zugeschrieben werden kann (hier in einer kontradiktorischen Verwendung), die aber nur eine Teilklasse der Phänomene darstellt, für die andere Autoren die Begriffe verwenden. Oaksford und Chater (1993) urteilen für den Bereich der Strategieanwendung:

> It is only possible to condemn people as irrational for not using a particular strategy if they *could* use it. To think otherwise, would be like condemning us because we can not breathe under water even though we do not possess gills. (S. 51)

In ähnlicher Weise grenzt Davidson (2004) das Irrationale vom Nicht-Rationalen ab. Allerdings findet er ein Paradox in der Idee irrationaler Phänomene:

> The idea of an irrational action, belief, intention, inference, or emotion is paradoxical. For the irrational is not merely the non-rational, which lies outside the ambit of the rational; irrationality is a failure within the house of reason. (S. 169)

Für ihn kann nur eine rationale Kreatur auch irrational sein. Diese Position ähnelt der Idee Searles, dass eine Möglichkeit für eine rationale Handlung bestehen muss, denn Davidson sieht den Versuch der Quadratur des Kreises etwa nur dann als irrational, wenn das Wissen vorliegt, dass es unmöglich ist. Das eigentliche Paradox ergibt sich nun für Davidson aus dem Problem der Erklärung:

> The underlying paradox of irrationality, from which no theory can entirely escape, is this: if we explain it too well, we turn it into a concealed form of rationality; while if we assign incoherence too glibly, we merely compromise our ability to diganose irrationality by withdrawing the background of rationality to justify any diagnosis at all. (S .184)

Wird Irrationalität unerklärbar und damit auch unvorhersagbar, wird der Beurteilende in seiner diagnostischen Fähigkeit für Davidson suspekt, da er nicht ausschließen kann, in seiner Diagnose selbst irrational zu sein, wird Irrationalität hingegen wieder in das Gebiet mentaler Verursachung zurückgeholt, so wäre die Handlung unter Berücksichtigung der Umstände ihrer Genese wieder als rational, da erklärbar und verursacht zu klassifizieren. Der von Searle vorgeschlagene Spalt würde in dieser Perspektive einen definitorischen Kraftakt darstellen, der Freiheit postuliert, die einem Verursachungsschema zuwider läuft. Gigerenzer (2001b) betont den Unterschied zwischen nichtrationalen Theorien der Entscheidungsfindung und „irrationalem Entscheiden“. Eine nichtrationale Theorie der Entscheidungsfindung, die Zeit- und Wissensgrenzen einräumt und Emotionen berücksichtigt, muss keinen Agenten beschreiben, der schlechte Entscheidungen trifft. Nichtrationale Theorien können bei Fehlen einer im engeren Sinne rationalen Strategie auch normativ interpretiert werden, zudem kann auf die Zuschreibung von Irrationalität ohne Konflikt verzichtet werden. Als Reaktion auf überzogene Anforderungen an die Rationalität des Menschen definiert Elster das Konzept der Hyperrationalität:

> As in Kant's critique, the first task of reason is to recognize its own limitations and draw the boundaries within which it can operate. The irrational belief in the omnipotence of reason is what I call hyperrationality. [S. 17](Elster, 1989)

Eine bewusst rational im Sinne von instrumentell rational ausgeführte Handlung kann den mit ihr intendierten Effekt sogar verhindern, da ein Übermaß an Willen ähnliche Effekte wie Willensschwäche haben kann. Hyperrationalität kommt für Elster auch dann ins Spiel, wenn Wünsche abgelehnt werden, deren Erfüllung Freude bereitet, nur weil der Prozess ihrer Formung nicht rational erklärt werden kann. Dabei ist Rationalität für ihn immer noch privilegiert, weil Menschen rational sein wollen, sie im Gespräch die Annahme von Rationalität zur Ermöglichung eines Austauschs treffen müssen und auch zur Zuschreibung von Irrationalität Rationalität notwendig sind:

> The explanatory privilege of rationality rests on two grounds. [...], rationality is presupposed by any competing theory of motivation, whereas rationality itself does not presuppose anything else. On grounds of parsimony, therefore, we should begin by assuming nothing but rationality. Also, while rationality may have its problems, the opposition is in even worse shape. The dictum that you can't beat something with nothing applies here, with some modification. (S. 42f)

Damit wird letztlich ein asymmetrisches Verhältnis von Rationalität und Irrationalität beschrieben. Zur Frage, ob Rationalität als ein universales oder ein spezifisches Konzept verstanden werden sollte, kann zunächst einmal mit Schreier (2003) formuliert werden:

> 'Rational' wird aufgefasst als ein Dispositionsprädikat; mit verschiedenen Rationalitätsbegriffen verbindet sich jeweils die Spezifikation der Kriterien, die erfüllt sein müssen, um einem Menschen das Prädikat 'rational' zuschreiben zu können [...]. (S.108)

Die Einnahme dieser Perspektive leistet noch keine Positionierung gegenüber der einleitenden Frage, denn die zur Debatte stehenden Begriffe könnten in Konkurrenz zueinander stehen und jeweils universalen Anspruch haben, es wird hier zunächst nur eingeräumt, dass es keine einheitliche Konzeption von Rationalität gibt, weder einer universalen noch einer oder mehrerer spezifischer Rationalitäten. Spezifisch könnte in diesem Fall bedeuten, dass es entweder Individuen gibt, für die abweichende Rationalitätskonzepte

formuliert werden können oder müssen, oder Situationen, Kontexte oder Umgebungen, in denen nicht die gleiche Form von Rationalität Gültigkeit hat. Ein universaler Standpunkt würde eine Binnendifferenzierung nicht bis zum Punkt der Wesensverschiedenheit aushalten. Aber auch hier gibt es Auffassungen zwischen den Polen:

> Eine uniforme Verständnisweise von Rationalität deutet sich genauso wenig an wie eine heterogene Pluralität. Rationalität ist weder Einheit noch Vielheit. Vielmehr ist Rationalität kontextuell und perspektivisch im Zugang, aber individuell verbindlich in Genese, Geltung und Anwendung. (Karafyllis & Schmidt, 2002, S. 13)

Der Kontext wird in diesem Ansatz allerdings im Folgenden stärker betont:

> Rationalität wird gerade in ihrem Kontextbezug als zentral und leitend für das Menschsein und das Gesellschaftsein im Zukunftshorizont angesehen. Rationalitäten sind zugangsspezifisch, sie entstehen im Kontext, in einer Lebenspraxis mit anderen, im Dialog eines Handlungsvollzugs, in Bezug auf ein Gegenüber in einer gemeinsamen Situation. (S. 13)

Gleichzeitig wird die Existenz kontextueller oder perspektivisch heterogener Rationalitäten nicht als unproblematisch betrachtet, da zumindest eine Fiktion oder Leitvorstellung von universaler Rationalität gewünscht oder angenommen wird und das Aufscheinen der Spaltungen nicht als Selbstverständlichkeit hingenommen werden kann.

4.1.2 Die Problematik von Normen der Rationalität

Im Folgenden soll geprüft werden, welche Rolle bestimmten vorgeschlagenen Normen zugesprochen werden kann für die Frage, ob menschliches Verhalten als rational oder irrational deklariert werden kann. Für Eisenführ und Weber (2003) gibt es eine beeindruckende Liste von Aspekten menschlichen Verhaltens, die nicht unter „rationales Verhalten“ fallen, in alphabetischer Sortierung ohne Anspruch auf Vollständigkeit: Ambiguitätseinstellung, *Anchoring and Adjustment*, *Availability* Bias, Bandbreiteneffekt, *Base Rate Fallacy*, Besitztumseffekt, *Choice vs. Matching*-Anomalie, *Disappointment*-Effekte, *Framing*-Effekte, *Gambler's Fallacy*, *Hindsight Bias*, *Illusion of Control*, *Mental Accounting*, *Omission*-Bias, *Overconfidence-Bias*, *Preference Reversal*, Referenzpunkt-Effekt, *Regret*-Effekte,Repräsentativitäts-Heuristik, *Resolution of Uncertainty*, Sicherheitseffekt, *Splitting*-Bias, Status quo-Bias, *Sunk Costs*, Überschätzung kleiner Wahrscheinlichkeiten, Verlustaversion, *Winner's Curse*. Schreier (2003) beobachtet einen Wandel in der Einschätzung der kognitiven Fähigkeiten des Menschen in der zweiten Hälfte des letzten Jahrhunderts:

> Aus dem intuitiven Wahrscheinlichkeitstheoretiker wurde ein Tollpatsch, der kein syllogistisches Schlagloch, keinen entscheidungstheoretischen Stolperstein auslässt, so dass sich die Bewertungsrichtung der psychologischen Rationalitätsforschung gegenüber den 1960er Jahren gerade umkehrte. Die Frage lautete häufig nicht mehr, weshalb der prinzipiell durchaus rationale Mensch in bestimmten Situationen und bei bestimmten Problemstellungen einem Irrtum unterliegt, sondern wie es bei einem kognitiv so inkompetentem Wesen wie dem Menschen eigentlich gelingt, sich überhaupt im Alltag zurechtzufinden [...]. (S. 112)

Es steht hier nicht eine Detaildiskussion der beschriebenen angenommenen Effekte im Vordergrund, sondern eine Prüfung der Basis für die Diagnose von Irrationalität, die zu einer Aufnahme der Effekte in diese Liste geführt haben mag.

Antworten auf die Frage, nach welchen Kriterien menschliches Denken als rational oder irrational bewertet werden kann, haben bisher keine konsensfähigen Positionen erzeugt, wenn das gesamte Gebiet der Kognitionsforschung betrachtet wird. Um Konsens und Dissens voneinander trennen zu können, soll hier ein Streit zwischen zwei prominenten Lagern dokumentiert werden, deren Forschungsarbeiten in dieser Arbeit vorgestellt worden sind, den Vertretern des Forschungsprogramms von Kahneman und Tversky auf der einen Seite und Gigerenzer mit der ABC-Gruppe auf der anderen Seite. Bei einer chronologischen Sortierung der Argumente fällt auf, dass zu Beginn der 90er Jahre ein Angriff auf die Grundthesen des *Heuristics-and-Biases*-Programms stattfindet, auf den 1996 eine Reaktion seitens der Angegriffenen erfolgt (Kahneman & Tversky, 1996), auf die wiederum von Gigerenzer (1996) geantwortet wird. Diese Phase ist einer der wenigen Momente von Interaktivität in der Debatte, die im Anschluss wieder in größtenteils unverbundenen Publikationen ausgetragen wird.

4.1.2.1 Gigerenzers Kritik am Fehlerkonzept von Kahneman und Tversky

Tversky und Kahneman (1983) nennen drei Bedingungen, damit von einem Fehler und nicht nur von einem Missverständnis gesprochen werden kann:

1. Ein nichttrivialer Fehler wird zugegeben, der in ähnlichen Situationen wiederholt auftreten würde.
2. Der Fehler ist ein konzeptueller Fehler, kein rein verbaler oder technischer Fehler.
3. Die richtige Lösung oder eine Lösungsprozedur zu ihrer Ermittlung hätten gefunden werden können.

Caverni und Péris (1990) bemerken, dass ein Bias in der künstlichen Laborsituation entstehen, aber in einem informationsreichen Kontext verschwinden kann, ähnlich wie Wahrnehmungsillusionen bei der Interaktion mit Objekten verschwinden. Der Angriff Gigerenzers muss als fundamentaler eingestuft werden und als substantiver in dem Sinne, dass nicht die unterstützenden empirischen Belege für eine These (methodisch) kritisiert werden, sondern Aspekte der These selbst auf konzeptueller Ebene hinterfragt werden. Eine Vermittlung zwischen den beiden Positionen kann deshalb größtenteils kategoriell nicht auf experimenteller Ebene erfolgen. Zentral für den Disput ist die Rolle von Normen, die zur Bewertung der Qualität menschlichen Denkens und zur Diagnose von Fehlern und Verzerrungen im menschlichen Denken als relevant betrachtet werden.

Nach Gigerenzer (1993) sind die Gesetze der Logik und der Wahrscheinlichkeitsrechnung schlechte Kandidaten, um normative Geltung beanspruchen zu können. Diese kritisierte Setzung hat eine geschichtliche Tradition: für Laplace war die Wahrscheinlichkeitsrechnung nichts als die Reduktion eines allgemeinen Verständnisses auf einen Kalkulus, für Boole waren die Gesetze von Logik, Algebra und der Wahrscheinlichkeitsrechung aus den Gesetzen des menschlichen Denkens ableitbar. Auch Inhelder und Piaget setzen den propositionalen logischen Kalkulus als Norm für menschliches Denken an. Im Heuristics-and-Biases-Ansatz muss die enge Kopplung allerdings aufgegeben werden, denn menschliches Denken könnte schließlich nicht gegen die eigenen Gesetze verstoßen: Die Gesetze der Logik und Wahrscheinlichkeitsrechnung werden dementsprechend als normative Standards gesetzt, gegen die menschliches Denken empirisch verstoßen kann.

Ein erster Kritikpunkt von Gigerenzer (1993) ist die Problematik einer Definition von mentalen Algorithmen, die nicht zwischen dem Algorithmus im engeren Sinne als Abfolge von Schritten und der Informationsrepräsentation trennt. Zahleninformationen können beispielsweise auf verschiedene Art und Weise repräsentiert werden:

> Not only are algoritms tuned to particular representations, but different representations make explicit different features of the same information. (S. 289)

Als zweite relevante Differenzierung wird von Gigerenzer (1993) auf den Unterschied zwischen einem subjektiven „degree of belief“ und objektiven Häufigkeiten verwiesen. Die Bayesianische Auffassung von subjektiver Wahrscheinlichkeit kann kontrastiert werden mit der frequentistischen Auffassung von Wahrscheinlichkeit als relativer (objektiver) Häufigkeit, die im 19. und 20. Jahrhundert zur dominanten Auffassung wurde. Die Angabe von Wahrscheinlichkeiten für einzelne (nicht-wiederholte) Ereignisse ist nur bei einer subjektiven Auffassung von Wahrscheinlichkeit sinnvoll, für den Frequentisten ist die Angabe einer Wahrscheinlichkeit hier überhaupt nicht möglich.

In einer evolutionären Spekulation zieht Gigerenzer die Konsequenz, dass eine Kapazität oder ein Algorithmus, der durch Prozesse der natürlichen Selektion entwickelt worden ist (wie z.B. Lernmechanismen in der Form von Habituation, Sensitivierung oder operanter Konditionierung), mit hoher Wahrscheinlichkeit auf eine ökologisch relevante Repräsentation von Information abgestimmt worden sein sollte. Diese Betrachtung würde den Zuschnitt auf Ereignishäufigkeiten in sequenzieller Kodierung (wie „3 von 20“) wahrscheinlicher machen als den Zuschnitt auf Prozentzahlen oder Wahrscheinlichkeiten von Einzelereignissen, die erst in Jahrtausenden des Umgangs mit Schriftlichkeit und Numerik entwickelt werden konnten. Zudem enthält diese Kodierung durch die Berücksichtigung von Stichprobengrößen und unter Umständen sogar Referenzklassen mehr Information.

Für den Frequentisten (wie Neyman oder von Mises) ist die Angabe einer Wahrscheinlichkeit für die Gültigkeit einer Aussage, die sich auf Eigenschaften eines einzelnen Individuums bezieht, nicht zulässig, die Diagnose eines Fehlers im Umgang mit subjektiven Wahrscheinlichkeiten ist deshalb nicht voraussetzungslos:

> One should be careful not to evaluate reasoning against some norm, unless subjects have been commited to that particular norm. ... What has been labeled the 'conjunction fallacy' here does *not* violate the laws of probability. It only looks so from *one* interpretation of probability. (Gigerenzer, 1993, S. 293)

Angegriffen wird also die fehlende Berücksichtigung von Repräsentationsformaten und die implizite Annahme der Gültigkeit von Normen, für die es geschichtlich, wissenschaftlich, statistisch oder psychologisch keinen zwingenden Grund gibt. In diesem Abschnitt kann deshalb die angegriffene Form von Normativität als Erstellung künstlicher Normen betrachtet werden. Künstlich sind diese Normen, weil sie weder durch den Kontext gestützt sind, noch im Rahmen eines diskursiven kommunikativen Prozesses etabliert werden konnten. Ein Fehler im Sinne eines Verstoßes gegen eine künstliche Norm ist dementsprechend ein ebenso künstlicher Fehler:

> [...] it should be clear that the single most trenchant conclusion reached by heuristics-and-biases program, namely that people are all too bad at reasoning, is itself, to a large degree, an illusion fostered by all-too-narrow norms of sound reasoning. (Gigerenzer, 1997, S. 206)

4.1.2.2 Der Streit

Die Angriffe auf das Programm führten 1996 zu einer scharf formulierten Gegenreaktion von Kahneman und Tversky (1996). Zur Legitimierung der Beschäftigung mit Verzerrungen des Denkens führen sie drei Hauptargumente an:

> There are several reasons for studying judgmental or perceptual biases. First, they are of interest in their own right. Second, they can have practical implications (e.g. to clinical judgment or intuitive forecasting). Third, the study of systematic error can illuminate the psychological processes that underlie perception and judgment. (S. 582)

Das erste Argument ist wenig trennscharf, da wenigen Gegenständen der Forschung eine potenzielle Bedeutung im eigenen Kontext abgesprochen werden kann. Die beiden anderen Argumente sind interessanter: Die Bedeutung von Fehlern wird in Problemen im Rahmen von praktischen Entscheidungssituationen gesehen, zum anderen wird ein Einblick in die psychologischen Prozesse der Wahrnehmung und der Entscheidung erhofft. Die Kritik Gigerenzers wird äußerst entschieden zurückgewiesen:

> The position described by Gigerenzer is indeed easy to refute, but it bears little resemblance to ours. It is useful to remember that the refutation of a caricature can be no more than a caricature of refutation. (S. 584)

Die Autoren wehren sich gegen den Vorwurf der fehlenden Kontextspezifität ihrer Heuristiken. Gigerenzers Position fassen sie dabei ihrerseits wie folgt zusammen:

> Gigerenzer argues that because the concept of subjective probability is controversial in statistics, there is no normative basis for diagnosing such judgments as wrong or biased. (Kahneman & Tversky, 1996, S. 582)

Sie sehen dabei ihre eigenen Arbeiten nicht von der Kritik getroffen, denn

> ... the normative standards to which we have compared intuitive judgments have been eclectic and often objective. (S. 583)

Weiterhin werten sie die Relevanz des Verweises auf alternative Konzeptionen der Wahrscheinlichkeit ab, mit einem Verweis auf das Laienverständnis von Wahrscheinlichkeiten. Dabei werfen sie Gigerenzer gleichzeitig einen normativen Agnostizismus vor, da ihrer Ansicht nach bei ihm keine Kriterien verbleiben, die zur Bewertung von Urteilen verbleiben:

> Although Gigerenzer invokes the meaninglessness argument with great conviction, his position on the issue is problematic. On the one hand, he surely does not regard statements of subjective probability as meaningless; he has even collected such judgments from subjects. On the other hand, he invokes the argument that subjective probabilities are meaningless to deny that these judgments are subject to any normative standards. This position, which may be described as normative agnosticism, is unreasonably permissive. Is it not a mistake for a speaker to assign probabilities of .99 both to an event and to its complement? We think that such judgments should be treated as mistaken; they violate accepted constraints on the use of probability statements in everyday discourse. (S. 585f.)

Der Vorwurf bezüglich der Wahrscheinlichkeitsaussagen wird insbesondere noch einmal verschärft im Kontext der Anwendung logischer Regeln, wie der Konjunktionsregel, die besagt, dass die Wahrscheinlichkeit einer Konjunktion von Ereignisse niemals größer sein kann als die Wahrscheinlichkeit eines der Einzelereignisse. Titelgebend für ihre Antwort ist die Insistenz bezüglich der „Realität“ kognitiver Illusionen. In Analogie zu visuellen Illusionen wird zugestanden, dass es Situationen gibt, in denen kein Fehler auftritt:

> It is in the nature of both visual and cognitive illusions that there are conditions under which the correct answer is made transparent. The Müller-Lyer illusion, for example, „disappears“ when the two figures are embedded in a rectangular frame, but this observation does not make the illusion less interesting. (S. 586)

Zusammenfassend sehen sie maximal eine gravierend eingeschränkte Version von Gigerenzers Kritik als stichhaltig an:

> [...] a charitable interpretation of his position is that he has overstated his case by omitting relevant quantifiers. Thus, some cognitive illusions (not all) are sometimes reduced (not made to disappear) in judgments of frequency. (S. 589)

Gigerenzer (1996) reagiert mit einem Gegenangriff, der dabei auf drei Kategorien von Aussagen fokussiert. Zum einen spricht er Kahneman und Tversky den Erfolg in dem von ihnen deklarierten Bemühen ab, kognitive Prozesse zu verstehen, die Urteile produzieren. Weiterhin kritisiert er erneut die Verwendung künstlicher Normen. Zuletzt wendet er sich explizit gegen den Vorwurf des normativen Agnostizismus.

Zu den grundsätzlichen Vorwürfen gehören zum einen die Kritik am engen Charakter der Normen, die „kognitive Illusionen“ generieren, und zum anderen die Problematisierung des theoretischen Status der vom *Heuristics and Biases*-Programm betrachteten Heuristiken:

> [...] vague heuristics have directed attention away from detailed models of cognitive processes and toward post-hoc accounts of alleged errors. (S. 592)

Insbesondere wird hier vorgeworfen, dass sich die kritisierten Heuristiken zum Teil auf One-Word-Labels reduzieren lassen, deren Erklärungsleistung allenfalls in der Neubeschreibung des zu erklärenden Phänomens liegt. Insbesondere wird das Fehlen von Prozessmodellen kritisiert, so dass auch für die Erklärung des Verschwindens von Illusionen lediglich einfache Umschreibungen angeführt werden. Hier nimmt Gigerenzer explizit Bezug auf die von Kahneman und Tversky (1996) verwendeten Formulierungen[1]:

> Recall Molière's parody of the Aristotelian doctrine of substantial forms: Why does opium make you sleepy? Because of its dormative properties. Why does a frequency interpretation cause more correct answers? Because „the correct answer is made transparent“ [...]. Why is that? Because of „a salient cue that makes the correct answer obvious“ [...]. Researchers are no closer to understanding which cues are more „salient“ than others, much less the underlying process that makes them so. (Gigerenzer, 1996, S. 594)

[1] Gegen die Satire Molières entgegnet allerdings Turner (2003) in Bezugnahme auf Leo Strauss:

> [...] if opium did not have dormitive powers it would not have been able to produce sleep. The claim is not arbitrary. At best, it is classificatory. Such classifications pointed to what, for the most part, were genuine explanatory problems. (S. 23)

Ist der Prozess unbekannt, wird es schwierig, die Angemessenheit von normativen Setzungen zu überprüfen, ohne starke Annahmen in Bezug auf den Kontext der Entscheidungssituation treffen zu müssen[2]. Dementsprechend zweifelt Gigerenzer die Angemessenheit eines experimentellen Paradigmas an, dass mit unlegitimierten bzw. künstlichen Normen operiert und somit Zahlenreihen erzeugt, deren Bedeutung sich nicht aus den Zahlen selbst erschließt:

> I am concerned with understanding the processes and do not believe that counting studies in which people do or do not conform to norms leads to much. If one knows the process, one can design any number of studies wherein people will or will not do well. (S. 594)

Die typische Modellkonstruktion von Statistikern sieht er als eine Abfolge von Analyseschritten: Einer Analyse des Inhalts eines Problems führt zu der Entwicklung von Annahmen, auf deren Grundlage schließlich ein statistisches Modell konstruiert wird. Der Inhalt des Problems bildet also den Anfangspunkt, das formale Modell das Ergebnis. Die Konstruktion eines typischen Experiments im Heuristics-and-Biases-Ansatz fasst er nicht ohne Ironie im Kontrast dazu in folgender Weise zusammen:

> The heuristics-and-biases program starts at the opposite end. A convenient statistical principle, such as the conjunction rule or Bayes's rule, is chosen as normative, and some real-world content is filled in afterward, on the assumption that only structure matters. The content of the problem is not analyzed in building a normative model, nor are the specific assumptions people make about the situation. (S. 592)

Als Gegenbild zum Vorwurf des normativen Agnostizismus entsteht das Bild eines indoktrinierenden Monotheismus seitens Kahneman und Tversky, der von ihm zurückgewiesen wird:

> Kahneman and Tversky [...] have subscribed to a controversial doctrine that indiscriminately evaluates all statements about single events by the laws of probability. I do not object to their philosophy, everyone is free to have one. But I do object to imposing it as a norm of sound reasoning on the participants in psychological experiments and as a criterion by which to diagnose cognitive illusions. (S. 593)

Den Verweis auf das Allgemeinverständnis von Wahrscheinlichkeit in der Öffentlichkeit sieht er als paradoxes Argument, da eine normative Grundlegung auf Basis der empirischen Analyse einer Praxis die normative Kritisierbarkeit eben jener Praxis ausschließen muss. Werden die Gesetze der Wahrscheinlichkeitsrechnung und Logik als unverrückbar betrachtet, muss ihnen ein A-Priori-Status zugesprochen werden, ähnlich wie es bei Wedgwood (1999) diskutiert wird. Für Gigerenzer ist die Auswirkung des Kontextes nicht allein darin zu sehen, dass er determiniert, ob Verletzungen gegen *die* Norm auftreten oder nicht. Die Rolle des Kontextes muss in seiner Betrachtung ergänzt werden um die Bestimmung, welche Normen überhaupt zur Bewertung von Handlungen als relevant betrachtet werden können:

> Content-blind norms are appropriate for textbook problems in probability theory, where the content is only decorative, but they are not appropriate either for evaluating human judgment or as a research tool to uncover the underlying process. (S. 593)

Im Unterschied zur Stellung von Computerprogrammen bei Herbert Simon und auch in der ABC-Gruppe findet Gigerenzer im kritisierten Forschungsprogramm keine programmierbaren Algorithmen und zweifelt fundamental an der Nützlichkeit der gewonnenen Ergebnisse:

> After three decades of research, the heuristics-and-biases program has generated only nebulous, if plausible, proposals for simple mechanisms of (poor) reasoning [...]. (Goldstein & Gigerenzer, 1999, S. 28)

[2] Zu Einwort-Erklärungen äußert sich später Gigerenzer (1998) auch in folgender Weise:

> One-word explanations derive their seductive power from the fact that almost every observation can be called upon as an example. (Gigerenzer, 1998, S. 197)

4.1.2.3 Wacklige Prüfsteine menschlicher Rationalität

Das Verhältnis zwischen rationalen Modellen und ihnen impliziten Normen und Verhalten ist nicht unproblematisch, nicht immer wird das Verhalten als durch die Norm kritisierbar bewertet:

> When a rational model fails to describe behavior, a different rational model, not different behavior, might be called for. (McKenzie, 2003, S. 405)

Gigerenzer (2006) kritisiert eine Forschungsmethodik, die mit einer vom Kontext getrennten Norm beginnt, die abstrakten Terme willkürlich mit Semantik und Pragmatik verbindet, um dann eine Abweichung von der Norm als „kognitive Illusion“ deklarieren zu können.

Gigerenzer (2001a) verweist auf das Problem der Interpretation von Einzelfallwahrscheinlichkeiten durch Laien. Er berichtet den Fall eines Mediziners, dessen Patienten bei der Verschreibung von Prozac im Allgemeinen sehr betroffen auf die Information reagierten, dass sexuelle Probleme mit einer Wahrscheinlichkeit von 30 bis 50 Prozent zu erwarten seien. Insbesondere reagierte jener Teil der Patienten betroffen, der diese Angabe so interpretierte, dass von nun an 30 bis 50 Prozent ihrer sexuellen Aktivitäten von diesen Problemen betroffen sein würden:

> The Prozac example illustrates *one* problem with statements about the probability of single events: a reference class is not specified. People who think in terms of concrete cases will fill a class in, consciously or unconsciously, from their own perspective. The example also indicates that when reference classes differ, single-event probabilities and frequencies may be systematically different. (S. 94)

In vielen Fällen kann durch eine Veränderung von Wahrscheinlichkeitsangaben von subjektiven Wahrscheinlichkeiten hin zu expliziten Häufigkeiten das Verständnis und der Umgang erleichtert werden (Hertwig & Hoffrage, 2002; Hoffrage & Gigerenzer, 2004). Die scheinbar paradoxen Konsequenzen der Vorenthaltung einer Referenzklasse werden in den Ergebnissen eines Experiments von Birnbaum (1999) deutlich:

> One group saw the number 9 and was asked to rate on a 10-point scale how large this number was. Another group was asked to rate how large the number 221 was. On average, the number 9 was rated larger than the number 221. (Gigerenzer & Dieckmann, 2006, S. 19)

Es kann ebenfalls prinzipiell gefragt werden, ob sich bspw. die „Gesetze der Logik“ als adäquate Normen für menschliches Verhalten qualifizieren können: Lowe (1993) versucht einzurahmen, was genau eine „auf den Gesetzen der Logik beruhende“ Norm sein kann, an der menschliches Denken gemessen werden soll. Erst einmal wird selten zwischen syntaktischen und semantischen logischen Ansätzen differenziert: Erstere stützen sich auf Formalisierungen in der Objektsprache, letztere fokussieren auf die Meta-Sprache (Wahrheitswertzuweisungen etc.). Formale Logik lässt sich nicht allein auf syntaktische Methoden reduzieren, wie es in vielen Fällen versucht wird. Zudem betont Lowe:

> ...the purpose of formal logic is to build upon our basic logical competence by constructing a set of tools which will enable us to *extend* our grasp of logical relationships... . (S. 222)

Das Instrumentarium der formalen Logik ist nicht intendiert, menschliches Denken zu spiegeln, sondern eine natürliche Grundkompetenz zu erweitern. Demnach wäre es für Lowe vermessen anzunehmen, dass menschliche Versuchspersonen im Allgemeinen Techniken der formalen Logik für einfache Schlüsse verwenden sollten, für die Kompetenzen erforderlich sind, die nur an Universitäten unterrichtet werden - die aristotelische Syllogistik hat nur noch geringe Ähnlichkeit mit modernen logischen Kalkülen. Für Aufgaben, die im gewöhnlichen Sprachgebrauch als „logische Probleme“ deklariert werden, ist nicht die Beherrschung formaler Logik notwendig, sondern allenfalls ein Umgang mit Wahrheitsbedingungen einfacher Sätze für einfache Schlüsse. Eine Basiskompetenz hierfür erwartet Lowe (1993) bei jedem kompetenten Denker, allerdings sieht er die Wahl der Methoden für die Erweiterung der Basiskompetenzen pragmatisch motiviert: Menschen wählen opportunistisch zwischen Diagrammen, Analogien, syntaktischen Methoden, Symbolismus und schlichtem Raten, ohne dass bei fehlendem formalen Training Konsistenz auch nur innerhalb eines einzelnen Subjekts zu erwarten wäre. Gleichzeitig muss das Denken nicht automatisch

erfolgreich sein, denn gerade weil es Probleme damit gab, bestimmte Resultate zu erzielen, wurde schließlich die formale Logik entwickelt. Die Vorstellung einer unbewussten Ausstattung mit Logikprogrammen analog zu Logiklehrbüchern hält er dementsprechend für abwegig, egal ob in der Form des natürlichen Schließens oder semantischer Tafeln. Für Boole ist der Inhalt logischer Gesetze erfahrbar im Studium mentaler Inferenz, für Frege sind die logischen Gesetze A-Priori-Prinzipien. Die Debatte zwischen Logizisten und Konstruktivisten hinsichtlich der Kategorie des Verhältnisses zwischen logischen Normen und menschlichen Denken ist bei weitem nicht in Konsens aufgegangen, so dass die Rede vom normativen Status der „Gesetze der Logik“ als fragwürdig zu betrachten ist.

Priest (2001) übt fundamentale Kritik an einem naiven Verständnis dieser „Gesetze“, die für ihn das Vermächtnis einer in dieser Hinsicht unzeitgemäßen Verehrung Aristoteles darstellen. Für Vorsokratiker wie Heraklit war der Glaube an Widersprüche durchaus legitim, wie etwa die Vorstellung, sowohl in den gleichen Fluss zu steigen, wie auch nicht in den gleichen Fluss zu steigen. Erst mit Aristoteles Gesetzen der Logik, die bis in das letzte Jahrhundert Bestand hatten, wurde dieser Glaube geächtet, und dies auch nicht im Bereich der Rezeption seiner Texte. Priest (2001) verweist auf ein aus der gleichen Zeit (und im diskutierten Sinne auch nicht aus der gleichen Zeit) überlieferten Standardkonzept indischer Logik, in dem zu jeder Zeit vier Möglichkeiten für eine Sichtweise betrachtet werden mussten, die als die vier Ecken (*catushkoti*) bezeichnet wurden: eine Sichtweise kann wahr und nur wahr sein, sie kann falsch und nur falsch sein, sie kann aber auch wahr und falsch zugleich oder aber weder wahr noch falsch sein. Als Konsequenz folgert Priest (2001):

> ... we may conclude that if figures in the history of Western philosophy held that it was irrational to believe a contradiction, this view was itself irrationally held. [...] Wise people, as Hume put it, apportion their beliefs according to the evidence. Those who subscribed to the orthodox view, were not, then, wise. They instantiated a pathology of reason. (S. 286)

Gegen den Einwand, dass aus einer Kontradiktion alles abgeleitet werden kann, verweist er auf die Unterstellung der Gesetze moderner Standardlogik für diesen Einwand, da bei der Verwendung parakonsistenter Logiken dieser Umstand nicht zutrifft. Die Rede von *der* Logik im Singular ist also ebenfalls irreführend, synchron und diachron sind zahlreiche unterschiedliche Konzeptionen und Kalküle entwickelt worden. Auch in Bezug auf logische Kalküle kann von Polysemie ausgegangen werden. Wenige Eigenschaften sind allen Konzepten gemeinsam, die Idee einer universalen Logik ist ohne eine explizite Definition des damit Gemeinten leer. Logik stellt ein von Menschen entwickeltes Hilfsmittel dar zur formalen Analyse spezieller Probleme, für die alternative Mittel nicht geeignet sind, sie kann keinen A-Priori-Status beanspruchen, und sie ist nicht das einzige Hilfsmittel zur Bearbeitung von Problemen, die als „logische Probleme“ bezeichnet werden.

4.2 Bounded Rationality bei Herbert Simon

Ein wichtiger Ausgangspunkt für ein Verständnis des Ansatzes der ABC-Gruppe ist Herbert Simons Konzept der *begrenzten Rationalität*. Eine Zusammenfassung des Konzepts „Bounded Rationality“ durch Herbert Simon selbst findet sich in Simon (2000):

> Bounded rationality is simply the idea that the choices people make are determined not only by some consistent overall goal and the properties of the external world, but also by the knowledge that decision makers do and don't have of the world, their ability or inability to evoke that knowledge when it is relevant, to work out the consequences of their actions, to conjure up possible courses of action, to cope with uncertainty (including uncertainty from the possible responses of other actors), and to adjucate among their many competing wants. Rationality is bounded because these abilities are severely limited. Consequently, rational behavior in the real world is as much determined by the „inner environment“ of people's minds, both their memory contents and their processes, as by the „outer environment“ of the world on which they act, and which acts on them. (S. 25)

Das implizit aufgebaute Gegenkonzept der Unbounded Rationality beinhaltet damit zum einen eine Überschätzung der Fähigkeiten von Entscheidern, zum anderen wird grundsätzlich das Problem einer Beschränkung der Perspektive auf die äußeren Gegebenheiten der Entscheidungsaufgabe aufgeworfen. Die

zusätzlichen „inneren Komponenten“ sind hier Wissen, Wissensabruf, der Umgang mit Unsicherheit und die Vermittlung zwischen konfligierenden Wünschen. Damit wird das innere Problem als ein Problem der Informationsverarbeitung und der Induktion charakterisiert. Gefordert wird deshalb vor allem ein methodisches Umdenken, da sich die „innere Welt“ der Entscheider nicht der distanzierten Beobachtung erschließt, weil auf die Unterstellung einer allwissenden Rationalität verzichtet wird: Simon (1979) sieht im Verzicht auf die Einfachheit in Theorien der begrenzten Rationalität gleichzeitig aber auch den gewünschten Verzicht auf starke und zweifelhafte Annahmen über die Fähigkeiten menschlicher Agenten, Ökonomen müssen sich dieser Veränderung methodisch anpassen:

> [...] economists must be trained in methods of observing ad interviewing, of taking and analysing verbal think-aloud protocols, of extracting information about decision processes from written records, and of drawing conclusions reliably from multiple studies of these kinds. (Simon, 2000, S. 36)

Die Annahme einer prinzipiellen Zugänglichkeit der relevanten Verhaltensaspekte liegt auch in Simons Konzeption der Wesensnähe von menschlichen und künstlichen informationsverarbeitenden Systemen begründet[3]. Gerade die Imkompatibilität zwischen der Theorie des subjektiv erwarteten Nutzens und menschlichen Entscheidungsproblemen veranlasst ihn, diese in Platons Himmel der Ideen zu verbannen, nicht aufgrund eines konzeptuellen, sondern aufgrund eines empirischen Mangels (Simon, 1993). Er wendet sich explizit gegen die impliziten entgrenzenden Rationalitätsannahmen der von ihm rekonstruierten neoklassischen Ökonomie:

> In its treatment of rationality, neoclassical economics differs from the other social sciences in three main respects: *(a)* in its silence about the content of goals and values; *(b)* in its postulating global consistency of behavior; and *(c)* in its postulating „one world“ - that behavior is objectively rational in relation to its total environment, including both present and future environment as the actor moves through time. (Simon, 1986, S. S210)

Gerade in diesem Punkt sieht er eine Abweichung zu den übrigen Sozialwissenschaften, die er in deren empirischem Zugang begründet sieht:

> In contrast, the other social sciences, in their treatment of rationality, *(a)* seek to determine empirically the content of goals and values; *(b)* seek to determine the processes, individual and social, whereby selected aspects of reality are noticed and postulated as the „givens“ (factual bases) for reasoning about actions; *(c)* seek to determine the computational strategies that are used in reasoning, so that very limited information-processing capabilities can cope with complex realities; and *(d)* seek to describe and explain the ways in which nonrational processes (e.g. motivations, emotions, and sensory stimuli) influence the focus of attention and the definition of the situation that set the factual givens for the rational processes. (Simon, 1986, S. S210)

Auch wenn die empirische Erfassung des 'Inhalts von Zielen und Werten' ein begrifflich verdächtiges Unterfangen scheint, das eine reduktionistische Lesart zu diagnostizieren scheint, wird erneut die Art der Ergänzung der Standardmethoden im Sinnes eines Imports aus den Verhaltenswissenschaften formuliert. Das Fundament der Hartnäckigkeit vieler Entscheidungsprobleme wird von Simon (1978) in der Komplexität der Welt gesehen, der eine Theorie der Rationalität gerecht werden muss, gleichzeitig zieht Simon aber nicht die Konsequenz, dass Problemlösungsversuche und Entscheidungen deshalb zum Scheitern verurteilt sind. Allerdings sind für ihn Bestrebungen nutzlos, die versuchen, Optima aufzuspüren oder *alle vorhandenen* Informationen zu verarbeiten. [4]:

> In a world where information is relatively scarce, and where problems for decision are few and simple, information is almost always a positive good. In a world where attention is a

[3] Künstliche Intelligenz stellt für Simon (1995) eine empirische Disziplin dar.

[4] Der Begriff vorhandener Information in der Umgebung ist erklärungsbedürftig, er steht dem Informationsbegriff der Informationstheorie nahe und kann aufgrund der Reifizierung subjektiver Strukturierungen nicht als unproblematisch im Argumentationskontext gelesen werden.

> major scarce resource, information may be an expensive luxury, for it may turn our attention from what is important to what is unimportant. We cannot afford to attend to information simply because it is there. (Simon, 1978, S. 13)

In diesem Sinne werden Informationsprobleme nicht als Probleme des Mangels an Information formuliert, sondern als Problem der Auswahl von Information, der Suche von relevanten Informationen in einem Überangebot, und der Versuch einer *vollständigen* Suche ist zum Scheitern verurteilt:

> When a great space of possibilities is to be explored [...], search becomes very selective. It is then guided by various rules of thumb, or heuristics, some of which are specific to particular tasks, but some of which are more general [...]. (Simon, 1990, S. 9)

Als Konsequenz bleiben Menschen nur Formen der begrenzten Rationalität, deren Bestandteile im Folgenden diskutiert werden sollen.

4.2.1 Heuristiken und Bounded Rationality

An dieser Schlüsselstelle positioniert Herbert Simon die Rolle von Heuristiken, die er als notwendige Daumenregeln für den Umgang mit Komplexität anführt - nicht im Sinne einer kritisierbaren Vernachlässigung von Informationen, sondern als einzig gangbaren Weg in Anbetracht der Umstände:

> Complexity is deep in the nature of things, and discovering tolerable approximation procedure and heuristics that permit huge spaces to be searched very selectively lies at the heart of intelligence, whether human or artificial. (Simon, 1978, S. 12)

Die alternative Idee einer universellen Berechenbarkeit und Optimierbarkeit weist er mit dem Beispiel des Schachspiels zurück, dass sich trotz seiner einfachen Struktur der Berechenbarkeit weitgehend entzieht (Simon, 1990). Die Einschränkungen, die Simon für menschliche Informationsverarbeitung betrachtet, sind für ihn immer auch Einschränkungen künstlicher Informationsverarbeitung, wenngleich das Ausmaß der Rechenleistung divergieren mag mit differenziellen Konsequenzen für die Problemschwierigkeit:

> *Because of the limits on their computing speeds and power, intelligent systems must use approximate methods to handle most tasks. Their rationality is bounded.* (Simon, 1990, S. 6)

Als Vorschläge für Heuristiken finden sich bei Simon weniger Vorschläge formaler Modelle als vielmehr Verweise auf mögliche Komponenten dieser Modelle. Als eine solche wird die Wiedererkennung vorgeschlagen, die Simon als Basis für Expertentum identifiziert in Bereichen von Schach bis hin zur medizinischen Entscheidungsfindung (Simon, 1990):

> We can regard intuition as a phenomenon that has been rather thoroughly explained: It is achieved through acts of recognition. (Simon, 1990, S. 9)

Die vollmundige Ankündigung der Reduzierbarkeit von Intuition auf Wiedererkennung soll an dieser Stelle nicht geprüft werden[5], für Simon gilt diese als erwiesen:

> Solving problems by responding to cues that are visible only to experts is sometimes called solving them by 'intuition'. A better label would be 'solving by recognition'. Intuition consists simply in noting features in a situation that index useful information. (Simon, 1996a, S. 86)

Expertentum lässt sich damit quantitativ über die wiedererkannte Informationsbasis fassen, demnach sind mindestens 50 000 Chunks für eine Weltklasseleistung erforderlich, sei es im Schachspiel, in der Musikkomposition, Malerei, Schwimmen oder Tennis[6]. Andere von Simon vorgeschlagene Heuristiken, sind Mittel-Ziel-Analysen oder Mustererkennung, die für ihn kompatibel sind mit Bounded Rationality:

[5] Bei einer großzügigen Definition von Intuition würde ein beträchtlicher Teil menschlichen Empfindens auf objektive Informationsverarbeitungsprozesse reduziert, und in der Tat spielt Simon wiederholt mit eben dieser Suggestion.

[6] Simon (1996a) verweist auf die Tatsache, dass Mozarts Musikausbildung mit vier Jahren begann, er damit auch die für Expertentum notwendige Trainingszeit von ungefähr zehn Jahren durchlaufen hat.

> [...] *They are not optimizing techniques, but methods for arriving at satisfactory solutions with modest amounts of computation.* They do not exhaust, but they typify, what we have been learning about human cognition, and they go a long way toward explaining how an organism with rather modest computational capabilities can adapt to a world that is very complex indeed. (Simon, 1990, S. 11)

Heuristiken werden damit verantwortlich gemacht für einen beträchtlichen Teil, wenn auch nicht die Gesamtheit menschlicher Kognition und sie werden Optimierungstechniken gegenübergestellt. Weizenbaum (1977) weist auf die Vorarbeit des Mathematikers George Polya hin, der mit dem 1945 erschienenen Buch „How to Solve It“ den Begriff der Heuristik für das mathematische Problemlösen stark macht. Polya hält die Suche nach „unfehlbaren Regeln“ nicht mehr für „seriöse Forschung“. Für ihn sind Heuristiken Verfahren, die auf Verallgemeinerung gerichtet sind, die unabhängig vom Gegenstand auf alle Problemtypen anwendbar sind. Zum Status des Konzepts „Bounded Rationality“ äußert sich Simon (2000) in folgender Weise:

> We can call bounded rationality a theory, although it is hard to say just what empirical predictions it makes. Alternatively, we can call it a representation: a suitable framework for economic theories, where the product of decision cannot be predicted without knowledge of the processes. Choosing bounded rationality for our representation of economic phenomena has substantial consequences for theory and the ways of discovering and testing it. (S. 26)

Der zentrale Kontrast zu klassischen Theorien wird in der Betonung des Prozesses der Entscheidungsfindung gesehen, der wie oben ausgeführt eine neue Methodologie der Untersuchung erfordert.

4.2.2 Ameisen, Gelatine und Scherenklingen

Eines der Hauptziele, die bereits in Simons frühen Texten formuliert werden, ist der Versuch, rationales Verhalten nicht dekontextualisiert, sondern in Abhängigkeit von organismischen Fähigkeiten und Eigenschaften der Umgebung zu betrachten (Simon, 1955). Die Einbeziehung der organismischen Begrenzungen ergibt sich als notwendig aus Plausibilitätsüberlegungen, ein Ausblenden der Umgebung inklusive von Aufgabenstellungen und Zielen wiederum würde ein ebenso seltsames Modell erstellen, das lediglich noch autistische Monaden modellieren könnte. Simon (1996b) betrachtet eine Analogie mit dem Lauf einer Ameise über einen Strand. Eine Analyse der Laufspur der Ameise deckt ein trügerisch komplexes Muster auf:

> Viewed as a geometric figure, the ant's path is irregular, complex, hard to describe. But its complexity is really a complexity in the surface of the beach, not a complexity in the ant. On that same beach another small creature with a home at that same place as the ant might well follow a very similar path. (Simon, 1996b, S.51f)

Bei einer Ameise kann nicht von einem übermäßig komplexen Organismus ausgegangen werden. Eine Rekonstruktion des Enstehungsprozesses des komplexen Pfades weist auf den Umstand hin, dass dieser Pfad nicht von der Ameise losgelöst von der Beschaffenheit der Umgebung kreiert wurde, sondern im Gegenteil sich aus Reaktionen auf deren Beschaffenheit zusammensetzt:

> An ant, viewed as a behaving system, is quite simple. The apparent complexity of its behavior over time is largely a reflection of the complexity of the environment in which it finds itself. (S. 52)

Die entscheidende Analogie wird von Simon durch die Ersetzung von „Mensch“ für „Ameise“ gebildet, so dass er die Hypothese erhält:

> Human beings, viewed as behaving systems, are quite simple. The apparent complexity of our behavior over time is largely a reflection of the complexity of the environment in which we find ourselves. (Simon, 1996b, S. 53)

Im folgenden Text beschränkt sich Simon (1996b) dann auf die Betrachtung des „denkenden Menschen“, allerdings nur aus methodischen Gründen:

> Now I should like to hedge my bets a little. Instead of trying to consider the „whole person“, fully equipped with glands and viscera, I should like to limit the discussion to Homo sapiens, „thinking person“. I myself believe that the hypothesis holds even for the whole person, but it may be more prudent to divide the difficulties at the outset, and analyze only cognition rather than behavior in general. (S. 53)

Seine Grundhaltung zur aufgestellten Hypothese spiegelt die Bemerkung wider, dass er einen kategoriellen Unterschied in der Wegfindung von Menschen und Ameisen nur aus der Perspektive des menschlichen Stolzes begründet findet (Simon, 1996b). Simon unterteilt die Betrachtung des Verhaltens eines künstlichen oder adaptiven Systems in die getrennte Betrachtung der „äußeren“ und der „inneren“ Umgebung. Mit der äußeren Umgebung sind situationale und physikalische Bedingungen gemeint, aber auch Aufgabenstellungen und externe Ziele, die „innere“ Umgebung ist ein irreführender Begriff, denn gemeint ist die innere Struktur und Organisation des Systems, es wird dementsprechend nichts umgeben:

> An artifact can be thought of as a meeting point - an „interface“ in today's terms - between an „inner“ environment, the substance and organization of the artifact itself, and an „outer“ environment, the surroundings in which it operates. (Simon, 1996b, S. 6)

Diese Grenzziehung[7] bietet aus Simons Perspektive Vorteile für die Analyse des Systems:

> [...] the first advantage of dividing outer from inner environment in studying an adaptive or artificial system is that we can often predict behavior from knowledge of the system's goals and its outer environment, with only minimal assumptions about the inner environment. (Simon, 1996b, S. 8)

Gleichzeitig ist aber auch eine gegenläufige Beziehung denkbar:

> There is often a corresponding advantage in the division from the standpoint of the inner environment. In very many cases whether a particular system will achieve a particular goal or adaptation depends on only a few characteristics of the outer environment and not at all on the detail of that environment. (Simon, 1996b, S. 8)

Der erste Fall wäre genau dann gegeben, wenn das System so designt worden ist, dass es das Ziel in jedem Fall erreichen kann, sämtliche potenziellen Hindernisse oder Störungen also überwinden kann. Der zweite Fall ist dann gegeben, wenn ein inneres Ziel verfolgt wird, das von der äußeren Umgebung abgekoppelt sein soll, eine homöostatische innere Temperaturregulation wäre ein derartiges Beispiel, hier werden alle Störungen von außen im Inneren kompensiert. Im ersten Fall ist eine Analyse der „äußeren“ Umgebung ausreichend für die Erklärung des Verhaltens, im zweiten Fall eine Analyse der „inneren“ Umgebung. Der Zusammenhang zu rationalen Konzeptionen des Menschen ergibt sich bei der Parallelbetrachtung von optimal designten Artefakten und dem perfekt rationalen Menschen (Simon, 1996b). Simon betrachtet den menschlichen Geist explizit als adaptives System, was die Parallele ermöglicht:

> The human mind is an adaptive system. It chooses behaviors in the light of its goals, and as appropriate to the particular context in which it is working. (Simon, 1992, S. 156)

Bereits in einer früheren Metapher - in der ökonomische Akteure mit einer Schüssel Gelatine bzw. „Wackelpudding“ verbunden werden - kritisierte Simon die klassische ökonomische Theorie:

[7] Erneut führt ein Kunstgriff zur Eliminierung von möglicherweise unauflösbarer Subjektivität, indem das Innerste zur Umgebung erklärt wird, und somit ein Blickwinkel zugleich Inneres und Äußeres erfasst, also einer „View from Nowhere“ entspricht, wie der Titel eines Buches von Nagel es umschreibt. Ähnlich operieren Einhorn und Hogarth (1981), wenn sie die inneren Eigenschaften des Organismus als „Kontext“ einführen:

> While context has typically been defined in terms of task variables, it is clear from the above examples that it is also a function of what the person brings to the task in the way of prior experience via learning, and biological limitations on attention, memory, and the like, that affect learning. Therefore, the elements of a psychological theory of decision making must include a concern for task stucture, the representation of the task, and the information processing capabilities of the organism. (S. 61f)

> If we wish to know the form gelatin will take when it solidifies, we do not study the gelatin; we study the shape of the mold in which we are going to pour it. In the same way, the economist who wishes to predict behavior studies the environment in which the behavior takes place, for the rational economic actor will behave in whatever way is appropriate to maximize utility in that environment. Hence (assuming the utility function to be given in advance), this maximizing behavior is purely a function of the environment, and quite independent of the actor.“ (Simon, 1990, S. 6)

Diese Metapher zeigt die vordergründige Eleganz der Optimierungsannahmen, denn bei ihrer Gültigkeit kann die gesamte „Innenseite“ des Verhaltens ausgeblendet werden, der mühsame Umweg durch die Psychologie bliebe erspart (vgl. Simon, 2000). Im gleichen Sinne könnte oder müsste sich die Psychologie auf Aspekte der Umgebung beschränken, um Verhalten zu erklären:

> The shape of a gelatin dessert cannot be predicted from the properties of gelatin, but only from the shape the mold into which it was poured. If people were perfectly adaptable, psychology would need only to study the environments in which behavior takes place. (Simon, 1992, S. 156)

Simon sieht dabei eine Anpassung an Erfordernisse eines Problems als gleichbedeutend mit Intelligenz[8]:

> The same strategy can be used to construct a psychology of thinking. If we wish to know how an intelligent person will behave in the face of a particular problem, we can investigate the requirements of the problem. Intelligence consists precisely in responding to these requirements. (Simon, 1990, S. 6)

Im Allgemeinen stellen sich die Dinge allerdings weniger einfach dar, selbst beim Design von Artefakten ist eine perfekte Anpassung an Aufgabe und Umgebung selten erreichbar:

> Often we shall have to be satisfied with meeting the design objectives only approximately. Then the properties of the inner system will „show through“. That is the behavior of the system will only partly respond to the task environment; partly it will respond to the limiting properties of the inner system. (Simon, 1996b, S. 12)

Die innere Struktur wird in dieser Anschauung genau dann sichtbar, wenn die Anpassung an die äußere Struktur nicht perfekt ist, das funktional designte Artefakt weist keine an sich wertvollen inneren Eigenschaften auf, sondern wird allein in Bezug auf die Aufgabenerfüllung bewertet. In diesem Sinne wird die Sichtbarkeit seiner inneren Eigenschaften zum Problem[9]. Eine rationale Analyse im Sinne der Annahme von fehlerfreier Problemlösung funktioniert beim Menschen in ähnlicher Weise nur unzureichend. Probleme werden selten exakt gelöst, sondern eher ungefähr, das Verhalten eines Systems begrenzter Rationalität kann für Simon (1990) nur durch eine Theorie der Systemprozesse und Umgebungen beschrieben, erklärt und vorhergesagt werden. Aus dieser Betrachtung leitet Simon (1990) die Idee der Scherenmetapher ab, die die oben angestellten Überlegungen zusammenfasst:

> Human rational behavior [...] is shaped by a scissors whose two blades are the structure of task environments and the computational capabilities of the actor. (S. 7)

Dabei hat Simon den Begriff der Umgebung zuvor in folgender Weise definiert:

[8] Dabei werden selbstverständliche Aspekte intelligenten Verhaltens ausgeklammert wie das Unterscheiden zwischen Problemen und Nichtproblemen, sowie die Einschätzung der Dringlichkeit und Relevanz von Problemen und den Nebenfolgen ihrer Lösung, dementsprechend muss die Formulierung ’“consists precisely“ zurückgewiesen werden. Gleichzeitig wird die Grenze zwischen der Fiktion einer „objektiven“ Situation und einer interpretierten, erlebten Situation verwischt.

[9] An dieser Stelle wird die Klippe zwischen funktionalen Artefakten und Menschen offenkundig. Simon (1992) verweist auf einige weitere Differenzen, die bei der Analyse menschlichen Verhalten von Bedeutung sind:

> In most real choice situations there is a multiplicity of goals, often partly conflicting and even incommensurable. A simple example is the trade-off between speed and accuracy. Unless we know their relative importance, we cannot select an optimal behavior. Nor are the alternatives from which the actor might choose usually known in advance [...].(S. 156)

> ... the term *environment* is ambiguous. We are not interested in describing some physically objective world in its totality, but only those aspects of the totality that have relevance as the „life space“ of the organism considered. Hence, what we call the „environment“ will depend upon the „needs,“ „drives,“ or „goals“ of the organism, and upon its perceptual apparatus. (Simon, 1956, S. 130)

Verschiedene Organismen finden dementsprechend in der gleichen Situation unterschiedliche Umgebungen vor [10]. Ähnlich wie die Klingen einer Schere zusammen benötigt werden, um Papier zu schneiden, ist die Analyse menschlichen Verhaltens in dieser Perspektive zum Scheitern verurteilt, wenn entweder - wie bei der Unterstellung perfekter Rationalität - nur die Eigenschaften der Aufgabe und Umgebung betrachtet werden oder andererseits der Organismus unabhängig von seiner Umgebung analysiert wird, Situation und Ziele und komputationale Mittel müssen als Prämissen für die Einschätzung der Rationalität eines Verhaltens gelten (Simon, 1986).

Die Rolle der Umwelt ist in der Konzeption Simons aber alles andere als nur die Quelle von Problemstellungen. Die notwendigen Vereinfachungen, die für ein adaptives Verhalten in komplexen Situationen notwendig sind, sind ebenfalls nicht isoliert auf der Organismenseite zu verorten:

> Now if an organism is confronted with the problem of behaving approximately rationally, or adaptively, in a particular environment, the kinds of simplifications that are suitable may depend not only on the characteristics - sensory, neural, and other - of the organism, but equally upon the structure of the environment. (Simon, 1956, S. 129f)

In der Konsequenz stellen Umgebungen damit eine ebenso notwendige Quelle für die Analyse rationalen Verhaltens dar wie die Organismen selbst, da mögliche Vereinfachungen von Entscheidungsverfahren aus ihnen abgeleitet werden können. Der methodische Fokus liegt bei Simon dabei auf eben dieser Einfachheit, wie das Ameisenbeispiel zu Beginn dieses Abschnitts verdeutlichen konnte. Nahe liegend ist deshalb für ihn die Fragestellung:

> How simple a set of choice mechanisms can we postulate and still obtain the gross features of observed adaptive choice behavior? (Simon, 1956, S. 129)

Diese Perspektive hat damit deutliche Implikationen für empirische Forschungszugänge. Simon sieht Rationalität dabei nicht auf Überlebensfunktionen oder im engeren Sinne biologische Funktionen von Organismen beschränkt. Es geht viel mehr um funktionales Verhalten:

> Behaviors are functional if they contribute to certain goals, where these goals may be the pleasure or satisfaction of an individual or the guarantee of food or shelter for the members of a society. (Simon, 1978, S. 3)

Unter diesem Blickwinkel können beispielsweise auch Institutionen betrachtet werden, die er - nicht unproblematisch - als Mittel zum Erreichen sozialer Bedürfnisse interpretiert. Als Kriterium für die Rationalität der Strukturen kann hier also die Anpassung an die Erfordernisse der Situation ausgemacht werden.

4.3 Ökologische Rationalität und die adaptive Werkzeugkiste

4.3.1 Bounded Rationality im Forschungsprogramm der ABC-Gruppe

Die Arbeiten der *Adaptive Behavior and Cognition (ABC)*-Gruppe am Max-Planck-Institut für Bildungsforschung in Berlin beziehen sich explizit auf ein Konzept von *Bounded Rationality* im Sinne Simons. Dabei werden beide Klassen der von Simon betrachteten Beschränkungen berücksichtigt:

> The term *bounded* can refer to constraints in the environment, such as information costs, and to constraints in the mind, such as limited memory [...]. (Gigerenzer, 2004a, S. 65)

[10] Hiermit wird zumindest für den Vergleich zwischen Spezies die Betrachtung subjektiver Perspektiven möglich, trotzdem ist diese Betrachtung bei Annahme identischer Bedürfnisse, Triebe und Ziele von Menschen mit der oben kritisierten erzwungenen Objektivität kompatibel.

Weiterhin werden von Simon die Absage an Optimierungsmodelle, der Fokus auf psychologisch realistische Modelle, die informationstheoretische Grundbetrachtung und die Hypothese der Serialität menschlicher Informationsverarbeitung übernommen:

> Begrenzte Rationalität hat also zwei verzahnte Komponenten: die Grenzen menschlicher Verarbeitungsfähigkeit und die Informationsstrukturen der Umgebungen, in denen Menschen Urteile und Entscheidungen treffen müssen. [..] Die erste Komponente fordert, dass jegliches Modell menschlichen Denkens auf realistischen Annahmen über menschliche Verarbeitungskapazitäten basieren muss, um psychologisch plausibel zu sein. Begrenztes Gedächtnis, unvollständiges Wissen und serielle Verarbeitung sind unleugbare Konstituenten menschlicher Rationalität. Menschen sind nicht kognitiv omnipotent, und optimale Strategien sind im Regelfall ein unerreichbares Ideal. (Hoffrage, Hertwig & Gigerenzer, 2005, S. 70)

Zentral für diesen Ansatz ist die Betonung der Problematik der fehlenden Berechenbarkeit optimaler Ergebnisse in vielen realistischen Entscheidungssituation:

> Models of bounded rationality recognize that humans often have limited information, time, and computational capacities when making judgments or decisions. Given these constraints, the optimal solution is often unattainable. Moreover, many problems are too complex to solve within a reasonable amount of time, even if all the relevant information is available and the most powerful computers are used. (Hoffrage & Reimer, 2004, S. 441)

Aufgrund der schieren Komplexität relevanter Problemstellung wird deshalb zur Wahrung der Handlungsfähigkeit allein eine selektive und deshalb unvollständige Suche gesehen:

> The key feature of bounded rationality is limited information search, which requires some kind of stopping rule. Note that here search can refer either to search for alternatives (e.g. mates), or search for each alternative's values on particular cues (e.g. a potential mate's age, sense of humor, etc.). (Chase, Hertwig & Gigerenzer, 1998, S. 208)

Gigerenzer und Selten (2001) setzen drei Klassen von einfachen Regeln als maßgeblich für realistische, menschliche Entscheidungsprozesse an. Der Anspruch ist dabei nicht eine vollständige Theorie der Rationalität, sondern eine Charakterisierung wesentlicher Merkmale vieler abgeleiteter Modelle:

1. *Simple search rules.* The process of search is modeled on step-by-step procedures, where a piece of information is acquired, or an adjustment is made [...], and then the process is repeated until it is stopped.

2. *Simple stopping rules.* Search is terminated by simple stopping rules [...]. Simple stopping rules do not involve optimization calculations [...] to determine the optimal stopping point.

3. *Simple decision rules.* After search is stopped and a limited amount of information has been acquired, a simple decision rule is applied [...]. (S.8)

Die Suchregeln sind Beispiele für serielle Informationsverarbeitung, die einfachen Abbruchregeln sind kompatibel mit einem *Satisficing*-Ansatz. Zuletzt werden einfache Entscheidungsregeln gefordert, die keine komplexen Berechnungen verlangen und von Menschen ohne technische Unterstützung angewandt werden können. Weiterhin werden auch Grenzen betrachtet, die durch die Umgebung selbst gesetzt werden. Todd, Fiddick und Krauss (2000) nennen zwei Beispiele für Einschränkungen der beschriebenen Art, die sich auf Organismen auswirken: Strukturelle Unsicherheit und Zeitdruck. Unsicherheit im Sinne fehlender Stabilität fordert Robustheit der Mechanismen, die vom Organismus in einer instabilen Welt eingesetzt werden. Zum anderen wird berücksichtigt, dass sich ein Entscheidungspunkt zeitlich nicht bis in die Unendlichkeit dehnen lässt, die Zeit, die auf Such- und Entscheidungsprozesse verwandt wird, führt zu Energieverbrauch und damit im Allgemeinen zu einer Verschlechterung der Situation für den Organismus, so dass ein Vorgehen mit einer schnellen Suche, das aufgrund weniger Informationen zu frugalen Entscheidungen führt, begünstigt wird.

In dieser Betrachtung werden an die Entscheidungsprozesse von Organismen durch Eigenschaften der natürlichen Umgebung drei Forderungen gestellt: Sie sollen einfach, schnell und erfolgreich sein. Eine

weitere Komponente, die den Ansatz der ABC-Gruppe auszeichnet und mit der Betrachtung Simons verbindet, ist die Betonung der adaptiven Leistungen gegenüber einer Herausstellung von Fehlerquellen bei der Betrachtung menschlichen Verhaltens. Die Idee von Optimalität selbst wird einer Prüfung durch die Welt ausgesetzt, um Idealen den Boden zu entziehen, die auf einer Vorstellung von Qualitätsmaßstäben basieren, die lediglich in kruden Modellwelten erreicht werden können. Aus diesem Grund wird wiederholt das als „optimal“ postulierte als defizitär im Zusammenprall mit realistischen Umwelten herausgestellt.

Gigerenzer (2005b) weist beispielsweise auf die Problematik eines perfekten Gedächtnisses hin. Ein als Ideal ersonnenes System erweist sich als dysfunktional, wenn berücksichtigt wird, dass menschliche Informationsverarbeitung bei Prozessen wie dem Spracherwerb vom Vergessen profitieren kann[11]:

> Is perfect memory desirable, without error? The answer seems to be no. The „sins“ of our memory seem to be good errors, that is, by-products („spandrels“) of a system adapted to the demands of our environments [...]. In this view, forgetting prevents the sheer mass of details stored in an unlimited memory from critically slowing down and inhibiting the retrieval of the few important experiences. Too much memory would impair the mind's ability to abstract, to infer, and to learn (Gigerenzer, 2005a, S. 211)

Die Betonung des Zusammenspiels von Organismus und Aufgabe bzw. Umgebung beinhaltet, dass eine kontextfreie Bewertung von Prozessen nur in wenigen Fällen sinnvoll scheint (etwa bei klar dominierten Strategien und Verankerung der Dominanzbeziehung in Relationen der Umgebung).

4.3.2 Die Ökologische Rationalität der ABC-Gruppe

Auch das Konzept der Ökologischen Rationalität wird auf die Rationalitätskonzeption Simons zurückgeführt, insbesondere auf die mit der Scherenmetapher implizierte Verbindung zwischen inneren und äußeren Strukturen:

> Just as one cannot understand how scissors cut by looking only at one blade, one will not undersand human behavior by studying cognition or the environment alone. (Gigerenzer, 2006, S. 119)

Damit wird sowohl der Beschäftigung mit Kognition allein eine Absage erteilt als auch einer Eliminierung des Geistes durch Ersetzung mit der Umgebung, wie es bei Brunswik oder Gibson anklingt (Todd et al., 2000). Die Umgebung wird also deshalb zum Fokus des Interesses, um in der Verbindung zu Strukturen des Geistes den Prozess menschlicher Kognition analysieren zu können:

> Die zweite Komponente der Scheren-Metapher impliziert, dass man das Innere der „Black Box“ menschlicher Kognitionen erhellen kann, indem man konsequent den Blick nach außen auf die Informationsstrukturen richtet. (Hoffrage et al., 2005, S. 70)

Die Informationsstrukturen der Umwelt werden deshalb als Untersuchungsgegenstand interessant, weil angenommen wird, dass stabile Strukturen das Funktionieren einfacher innerer Mechanismen ermöglichen, da auf eine Struktur zurückgegriffen werden kann, die nicht selbst erzeugt werden muss. Die Umgebung leistet damit einen Teil der Arbeit selbst (Todd et al., 2000). Die Einhaltung der in der Diskussion von *Bounded Rationality* genannten durch die Umgebung gesetzten Grenzen wird in dieser Betrachtung durch stabile Eigenschaften der Umgebung erleichtert. Diese positive Abstimmung zwischen Informationsstrukturen der Umgebung und Informationsverarbeitungsstrukturen des Geistes wird als *Ökologische Rationalität* bezeichnet. Bei Chase et al. (1998) werden Bounded Rationality und Ökologische Rationalität als separate Konzepte, die dem Geist als Attribut zugesprochen werden können, gegenübergestellt:

[11] Als kennzeichnend für Darstellungen der ABC-Gruppe ist zudem die auch im Zitat erkennbare rhetorische Strategie, den orthodoxen Sprachgebrauch ad absurdum zu führen. Das in der Einleitungsfrage des folgenden Zitats als „perfekt“ bezeichnete Gedächtnis wird als „suboptimal“ ausgewiesen, der Widerspruch löst sich auf, wenn hervorgehoben wird, dass sich der Betrachtungswinkel verändert, wenn zunächst „perfekt“ auf eine kontextabstrahierte Fehlerfreiheit bezogen wird und später dann Probleme des „perfekten Gedächtnis“ in Bezug auf eine spezifische Anwendung aufgedeckt werden. Dabei wird zwar auf das Gedächtnis zurückgegriffen, die Anwendung lässt sich aber nicht auf Gedächtnisprozesse reduzieren. In diesem Sinne ist das fehlerhafte Gedächtnis für die Aufgabenstellung „perfekter“ als das „perfekte“ Gedächtnis, was einen ironischen Betrachter zu der Schlussfolgerung verleiten könnte, dass im Beispiel an Optimalitätsannahmen auf einer zweiten Ebene festgehalten wird, indem alles unternommen wird, um mögliche Schwächen als Stärken auszuweisen.

> We argue that to discover how the mind works, and how well, we need to understand how the mind functions under its own constraints - its bounded rationality - and how it exploits the structure of the social and physical environments in which it must reach its goals - its ecological rationality. By adding these perspectives to our theoretical scope, we broaden and deepen our vision of rationality. (S. 212)

Diese Auffassung ist in den Texten der ABC-Gruppe aber nicht einheitlich, da zuweilen *Bounded Rationality* als übergreifendes bzw. als grundlegendes Konzept verstanden wird und ökologische Rationalität eher spezifischen Informationsverarbeitungsstrukturen zugesprochen wird, also bspw. eine Eigenschaften von Heuristiken ist:

> The „rationality“ of domain-specific heuristics is not in optimization, omniscience, or consistency. Their success (and failure) is in their degree of adaptation to the structure of environments, both physical and social. The study of the match between heuristics and environmental structures is the study of ecological rationality. (Gigerenzer, 2001d, S. 38)

Aus einer Passung zwischen Umweltstrukturen und Strukturen von Heuristiken resultiert, dass Heuristiken erfolgreich in diesen Umwelten und Umgebungen angewandt werden können. Während ökologische Rationalität letztendlich in den meisten Fällen als Eigenschaft von Informationsverarbeitungsprozessen bezeichnet werden, folgt daraus aber eben gerade nicht die Beschränkung auf eine isolierte Betrachtung dieser Prozesse. Um die Passung zwischen Heuristiken und ihrer Umgebung zu analysieren, ist ein breiterer Fokus notwendig:

> Models of ecological rationality describe the structure and the representation of information in actual environments and their match with mental strategies, such as boundedly rational heuristics. (Hoffrage & Reimer, 2004, S. 442)

Die ökologische Rationalität beruht nicht notwendigerweise auf einer einzigen Passungsdimension, unter Umständen kann ihre Grundlage in mehrere Aspekte aufgeteilt werden. Ein Aspekt der ökologischen Rationalität von einfachen Heuristiken ist die im letzten Abschnitt angesprochene Robustheit in unsicheren Umgebungen:

> Seen through the eyes of a model, information structures in the ecology contain noise that does not transfer to new situations or objects. By being simple, fast and frugal heuristics have - loosely speaking - a lower noise-to-information ration than complex strategies. (Hoffrage & Reimer, 2004, S. 450)

Trotz einer möglicherweise heterogenen Grundlage wird ökologische Rationalität trotzdem als eindimensionale Bewertungsdimension verstanden. Die Bestimmung ökologischer Rationalität soll dabei auf jeden Fall ordinale Vergleiche zwischen Heuristiken zulassen, zum Teil auch Messungen auf stärkeren Skalenniveaus ermöglichen:

> Ecological rationality can be expressed in comparative terms: a given heuristic performs better in environment A than in B. [...] Ecological rationality can also be analyzed in quantitative terms: a heuristic will make 80 percent correct predictions in environment E, and requires only 30 percent of the information. (Gigerenzer, 2006, S.123f)

Mit dieser Setzung unterscheidet sich ökologische Rationalität von anderen Rationalitätsbestimmungen. Wird Rationalität als Optimalität verstanden, dann wird sie zu einem dichotomen Konzept: Eine Handlung, eine Prozedur, eine Überzeugung ist entweder rational oder irrational, wenn Rationalität als Konzept auf sie angewandt werden kann. *Bounded Rationality* im Gegensatz dazu ist je nach Lesart entweder ein Modell, das qualitative universale Hypothesen über menschliches Denken und Handeln aufstellt, eine These, in der eine stärkere Form der Rationalität negiert wird, oder aber, was dem hier vorgestellten Konzept der ökologischen Rationalität am nächsten käme, ein graduelles Erfolgsmaß bei Berücksichtigung der Umwelt, in der ein Verhalten gezeigt wird. Da Rationalität hier einer Prozedur zugesprochen wird, stellt sich zunächst nicht die Frage, ob Menschen prinzipiell ökologisch rational sind oder nicht, da Menschen in verschiedenen Umgebungen verschiedene unterschiedlich ökologisch rationale Strategien

anwenden können. [12] Die These der begrenzten Rationalität menschlichen Denkens und Verhaltens ist also nicht deckungsgleich mit ökologischer Rationalität. Ein Bezug ergibt sich allerdings aus der Feststellung, dass bestimmte Umgebungen den Erfolg von Verfahren verhindern, die auf der Grundlage von Normen operieren, die eine unbegrenzte Rationalität implizieren. Wäre die Welt eine andere, so würde eine höhere ökologische Rationalität auch anderen Verfahren zugeschrieben werden. Dementsprechend sind Umgebungen eine nichtreduzierbare Komponente ökologischer Rationalität, so dass präzise gesprochen, nicht eine isolierte Heuristik zum adäquaten Träger des Konzeptes erklärt werden kann, sondern mindestens Heuristik und Umgebung spezifiziert werden müssen [13]:

> The rationality of heuristics is not logical, but ecological. Ecological rationality implies that a heuristic is not good or bad, rational or irrational per se, only relative to an environment. It can exploit certain structures of an environment or change an environment. (Gigerenzer, 2004a, S. 64)

Eine Bündelung der Aussagen über eine allgemeinere Form der Rationalität wird durch die Hypothese der adaptiven Toolbox ermöglicht, die im folgenden Abschnitt vorgestellt werden soll.

4.3.3 Die Adaptive Toolbox

Ketelaar und Todd (2001) stellen die Hypothese auf, dass sich Heuristiken als Elemente menschlicher Informationsverarbeitung aus basalen Bausteinen im Laufe der Evolution entwickeln konnten, weil sich die aus adaptiven Bausteinen geformten Strategien in den Umgebungen der Vergangenheit bewährt haben:

> This bottom-up approach to the construction of fast and frugal heuristics starts with psychological plausibility and ends with adaptive behaviour in a particular problem domain. Determining the initial set of building blocks to work with should be simpler than finding an optimal solution in the top-down approach. Finding suitable combinations of fast and frugal building blocks for a particular problem may often prove challenging - we do not yet know hot to specify this (re)search process - but trial-and-error methods (including heuristic-guided search) and exploratory shuffling will work here in a way that is impossible with the mathematical derivations required by the top-down method. (S. 206)

In dieser Konzeption ist ein großer Teil der von Menschen eingesetzten Heuristiken nicht mit ideosynkratisch entwickelte Ad-Hoc-Lösungen zu verwechseln. Gigerenzer spricht sogar von Heuristiken als speziesbestimmend für den Menschen und formt das Bild des *Homo Heuristicus* als theoretischer Gegenspieler zum *Homo Oeconomicus*:

> The question of the rationality of *Homo Heuristicus* concerns the question of ecological rationality. A heuristic is not good or bad, rational or irrational, in itself, but only relative to an environment, just as adaptations are context bound. (Gigerenzer, 2004b, S. 403)

Es wird eine Sammlung von Lösungsstrategien angenommen, die im Rahmen der gesetzten Grenzen zu vernünftigen, nicht aber notwendigerweise zu optimalen Ergebnissen verhelfen. In diesem Sinne sind

[12] Wird ökologische Rationalität in der beschriebenen Art und Weise verstanden, dann ist die Sprache von einer ökologisch rationalen Strategie im gleichen Sinne irreführend wie die Sprache von einem intelligenten Verhalten. In beiden Fällen wird eine Dichotomie suggeriert, die sich bei einem metrischen Konzept nur künstlich erzeugen lässt, ohne dass der notwendige Cutoff für die Grenzziehung geliefert wird. Zieht man die Grenze zwischen intelligentem und nicht intelligentem Verhalten etwa zwischen Lebewesen und Mineralien, Menschen und Tieren, bestimmten Menschen und anderen Menschen, bei einer Teilgruppe von Verhaltensweisen bestimmter Menschen etc., resultieren jeweils andere Bedeutung des Begriffes „intelligentes Verhalten“. In der gleichen Art und Weise ist die Zuschreibung von „ökologischer Rationalität“ unterdeterminiert. Ansonsten ist die Nähe zur Optimierung wieder hergestellt, was sich bspw. in der (fehlerhaften) Rekonstruktion ökologischer Rationalität durch Funke (2003) zeigt:

> Ökologisch rational heißt eine Heuristik dann, wenn sie an die Struktur ihrer Umwelt angepasst ist [...], indem sie die dort vorhandenen Hinweisreize maximal ausschöpft. (S. 104)

[13] Diese Klarheit der Zuschreibung wird nicht in allen Publikationen der ABC-Gruppe aufrechterhalten. Zuweilen kommt es zu einer Verselbständigung, so dass ökologische Rationalität entweder isolierten Heuristiken oder dem Menschen in seiner Gesamtheit zugesprochen wird.

die Heuristiken Modelle von *Bounded Rationality* (Hoffrage & Reimer, 2004). Dieser Sammlung wird die Bezeichnung *Adaptive Werkzeugkiste* bzw. *Adaptive Toolbox* gegeben[14]. Für Gigerenzer und Todd (1999) stellt diese Konzeption des Geistes eine Komponente der zweiten Revolution nach der probabilistischen Revolution dar, die zu einem erneuten Umdenken in der Positionierung des Menschen in der Welt führen könnte:

> The second revolution [...] concerns the way minds deal with an uncertain world. We propose replacing the image of an omniscient mind computing intricate probabilities and utilities with that of a bounded mind reaching into an adaptive toolbox filled with fast and frugal heuristics. (S. 5)

Heuristiken sind in dieser Konzeption Werkzeuge des Geistes, die nicht die Beherrschung psychologisch implausibler Kompetenzen verlangen, sondern in der Anwendung für Menschen einfach sind. Diese Werkzeuge werden als *fast and frugal heuristics* bezeichnet, Einfachheit gehört dabei zu den leitbildgebenden Idealen des Forschungsansatzes:

> The pleasing ideal of a universal calculus may have distracted researchers in many fields from exploring the contents of the adaptive toolbox. However, there is also another sense of beauty: the aesthetics of simplicity. There is a sense of wonder in how simplicity can produce robustness and accuracy in an overwhelmingly complex world. (Gigerenzer & Selten, 2001, S. 11)

Dabei sollen weder die Probleme trivialisiert werden, noch die Komplexität der Implementierung einfacher Heuristiken ignoriert werden. Zur Frage der Implementierung verweisen sie auf die Diskrepanz zwischen der einfachen Beschreibbarkeit des Mechanismus der Gesichtserkennung und der sehr komplexen neuronalen Realisierung des Mechanismus. Zur Frage komplexer Probleme argumentieren Todd et al. (2000):

> The question adressed in *Simple Heuristics* is not whether there are simple *problems*, but whether, given even a complex problem, there might be simple *solutions*. The point of exploring ecological rationality is to see to what extent simple mechanisms can be useful and adaptive even in complex environments. (Todd et al., 2000, S. 380)

Nach Gigerenzer (2001d) gibt es drei Prämissen des Forschungsprogramms der adaptiven Werkzeugkiste:

1. **Psychologische Plausibilität**: Ziel des Programms ist ein Verständnis wie echte Menschen, Bienen, Affen etc. Entscheidungen treffen, die über begrenzte Ressourcen an Zeit, Wissen, und Speicher verfügen.
2. **Bereichspezifität**: Anders als ihre zum Teil bereichsübergreifend konzipierten Bausteine sollen Heuristiken spezifisch auf einen Anwendungsbereich zugeschnitten sein und dementsprechend auch nur in diesem Bereich genutzt werden.
3. **Ökologische Rationalität**: Die *Adaptive Toolbox* wird als ökologisch rationale Sammlung von Heuristiken verstanden, ihre Bestandteile werden unter dem Aspekt der ökologischen Rationalität untersucht.

Modelle der begrenzten Rationalität sollen auf dem Verhaltensrepertoire - kognitiv, emotional und sozial - aufgebaut werden, das einer Spezies zur Verfügung steht. Hertwig und Hoffrage (2001a) sehen die

[14] Damit wird eine Abgrenzung von der evolutionspsychologisch gebräuchlichen Metapher des Geistes als Schweizer Taschenmesser vorgenommen:

> The possibility of recombining parts into new strategies is central to the toolbox metaphor, but not to the related image of the mind as a Swiss army knife [...]. Both views emphasize the existence of many domain-specific strategies (rather than a single general-purpose calculus); but the toolbox highlights the possiblity of recombining tools and nesting heuristics. (Todd, Gigerenzer & the ABC Research group, 2000, S. 770)

Grenzen der Intelligenz koextensiv mit den Grenzen der *Adaptive Toolbox*. Die Einfachheit der Heuristiken wird durch die Komplexität der Welt nicht behindert, sondern durch den systematischen Anteil der Komplexität ermöglicht. In diesem Sinne kann der Inhalt der angenommenen Werkzeugkiste nur durch eine Beschäftigung mit den Strukturen der Umgebungen und Aufgaben sinnvoll exploriert werden. Dabei wird allerdings nicht davon ausgegangen, das gesamte Spektrum menschlicher Kognition auf diesem Weg erklären zu können:

> By letting the world do some of the work - by relying on the presence of particular useful information patterns - the decision mechanisms themselves can be simpler; hence our focus on simple, fast and frugal heuristics (This is certainly not to say that all of cognition is handled by such heuristics, but we expect that a large proportion of the inferences and choices people make can be explained through their use.) (Todd & Gigerenzer, 2003, S. 148)

Einer der Vorteile einer informationstheoretischen Perspektive auf die Welt ist die Möglichkeit der Abstraktion, da Informationsstrukturen verschiedenster inhaltlicher Bereiche als zueinander isomorph bestimmt werden können und dementsprechend die gleiche Heuristik in sehr unterschiedlichen Bereichen zur Anwendung gebracht werden:

> [...] matching heuristics to environment *stucture* does not mean that every new *environment* or problem demands a new heuristic - the simplicity of these mechanisms implies that they can often be used in multiple, similarly structured domains with just a change in the information they employ [...]. (Todd et al., 2000, S. 379)

Die Formulierung von Ökologischer Rationalität als graduelles Konzept macht deutlich, dass nicht von einer perfekten Lösung aller von Menschen bearbeiteten Probleme ausgegangen wird. Das Auftreten von Fehlern wird also als notwendige Konsequenz der Struktur von Welt und Heuristiken nicht negiert, gleichzeitig wird aber in dieser Hinsicht kein systematischer Nachteil gegenüber alternativen Verfahren angenommen:

> Simple heuristics are, by their very nature and by the nature of an uncertain world, often wrong, and often in systematic ways, inevitably producing errors just as do standard normative procedures. The question is whether or not heuristics are *more* often wrong than other decision mechanisms that require more information and more computation. Our research has shown that this is often not the case - simple heuristics can match the performance of more complex mechanisms in many situations. (Todd et al., 2000, S. 379)

Es wird bspw. von Todd und Gigerenzer (1999) unterstrichen, dass die Werkzeugkiste nicht als halbseidene Lösung in einem Raum mit dominierenden Konkurrenten verstanden wird. Zuletzt wird gerade in der Analyse von Fehlern ein Hauptzugang zur Analyse der Funktionsweise von Heuristiken identifiziert. Die Heuristiken der Adaptiven Werkzeugkiste werden als Prozessmodelle vorgeschlagen, beinhalten also im Gegensatz zu Modellen mit Optimialitätsannahmen explizite Aussagen über den Prozess der Informationsverarbeitung, weshalb die Analyse von Fehlern in besonderer Weise relevant wird (Gigerenzer, 2004a).

Die Hypothese der adaptiven Werkzeugkiste und das Konzept der ökologischen Rationalität werden von Gigerenzer (2006) aufeinander bezogen, indem er die adaptive Werkzeugkiste als deskriptives Schlüsselkonzept beschreibt, die ökologische Rationalität als normatives Schlüsselkonzept. Die adaptive Werkzeugkiste stellt die Heuristiken bereit, für die das Konzept ökologischer Rationalität erst anwendbar ist, da sie als Prozessmodelle einer entsprechenden Prüfung unterzogen werden können:

> Heuristics work in real-world environments of natural complexity, where an optimal strategy is typically unknown. The study of ecological rationality answers the question: In what environments will a given heuristic work? Where will it fail? Note that this normative question can only be answered if there is a process model of the heuristic in the first place, and the results are gained by proof or simulation. (Gigerenzer, 2006, S. 119)

Hoffrage und Reimer (2004) ordnen die Beziehung zwischen *fast and frugal heuristics* und früheren Rationalitätskonzeptionen in folgender Weise:

> [Fast and frugal heuristics] are - consistent with traditional views about what constitutes a heuristic - useful tools for finding reasonable solutions to a given problem. In contrast to models of unbounded rationality, they are psychologically plausible process models that take the limits of human information processing into account. In contrast to models that optimize under constraints, they consist of simple building blocks, which dispense with unrealistic assumptions that inconspiciously introduce complex calculations when it comes to specifying stopping criteria for information search. In contrast to heuristics proposed in the heuristics-and-biases-program, they are formalized enough to allow for computer simulations to determine their fit in a particular real-world environment. (S. 457)

4.3.4 Das Forschungsprogramm der ABC-Gruppe

Als wesentliches Alleinstellungsmerkmal des Programms der ABC-Gruppe muss der Fokus auf ökologische Rationalität genannt werden, der eine Auflösung des Paradoxes anbietet, wie realistische und damit einfache Strukturen in einer komplexen Welt Überlebensfähigkeit ermöglichen können, also wie einfache Heuristiken für Menschen rational sein können:

> Given their ecological rationality, which exploits the structure of the environment in which they function, fast and frugal heuristics do not need to sacrifice accuracy for speed and frugality. Thus, it can be rational not to use information even when it is available: Simplicity does not imply irrationality. (Hoffrage & Reimer, 2004, S. 457)

Für die Untersuchung ökologischer Rationalität wird ein interdisziplinärer Ansatz gewählt, der Methodologien aus experimenteller Psychologe, der Computermodellierung, mathematischer Analyse, evolutionärer Biologie und kognitiver Anthropologie zu integrieren sucht (Todd & Gigerenzer, 2003). Die Menge der genannten Disziplinen liefert Ansätze für die Erforschung von Heuristiken, die als positiv bewertete kognitive Werkzeuge und nicht als Schwachstellen bewertet werden. Die Funktionen von Heuristiken werden dabei präzise spezifiziert, was vor allem durch Computermodellierung und distinkten Prinzipien für Informationssuche, Stoppregeln und Entscheidungsregeln erreicht wird (Goldstein & Gigerenzer, 1999). Als Komponenten des Forschungsprogramms werden die folgenden genannt:

> The program of studying boundedly rational heuristics involves (a) designing computational models of candidate simple heuristics, (b) analyzing the environmental structures in which they perform well, (c) testing their performance in real-world environments, and (d) determining whether and when people really use these heuristics. (Goldstein & Gigerenzer, 1999, S. 16)

Ein komputationales Modell wird dabei als Prozessmodell definiert, dass eine Umsetzung als Computerprogramm ermöglichen würde [15]. Dabei werden Computermodelle als Ausdruck der Transparenz der Heuristiken verstanden, die eine gewünscht nüchterne Perspektive versprechen:

> Transparent models of fast and frugal heuristics avoid misunderstanding and mystification of the processes involved, even if they do sacrifice some of the allure of the unknown. (Gigerenzer & Todd, 1999, S. 18)

Fünf Jahre später, nach Exploration der genannten Schritte sowie Identifizierung und Analyse einiger der Kandidaten für Heuristiken in der Toolbox, klingt die Formulierung des Programms bereits zuversichtlicher, die Menge der untersuchten Heuristiken ist klarer umrissen, die Belege für die Möglichkeit einer erfolgreichen Anwendung sind gesammelt:

> Studies on fast and frugal heuristics include (a) computer simulations to explore the performance of the heuristics in a given environment, in particular in real-world environments [...]; (b) the use of mathematical or analytical methods to explore when and why they perform

[15] Diese Forderung ist insofern elliptisch, als dass für die Anwendung einer programmierten Heuristik auch die Umgebung in ein Format gebracht werden muss, das mit den Schnittstellen des Programms kompatibel ist. Demenstsprechend ist ein Heuristikcode ohne ein Weltmodell nicht anwendbar, und die Gültigkeit und Angemessenheit des (im Allgemeinen informationstheoretischen) Weltmodells ist eine Voraussetzung für die Gültigkeit und Angemessenheit der Simulationen.

> as well as they do (eventually supported by simulations, in particular in artificial environments in which information structures are systematically varied [...]); and (c) experimental and observational studies to explore whether and when people actually use these heuristics [...]. (Hoffrage & Reimer, 2004, S. 442)

Bröder (2001) greift das Forschungsprogramm mit dem Argument an, dass die Relevanz für die psychologische Forschung eher bescheiden gewesen sei, da die vorgeschlagenen Heuristiken nicht als empirische testbare Modelle menschlicher Entscheidungen formuliert worden seien und eine empirische Überprüfung fehle. In einer Reaktion auf Bröder (2001) wenden Hertwig und Hoffrage (2001b) ein, dass eine Studie von Bröder selbst und andere Studien einen Ansatzpunkt für die geforderte Brücke zwischen Theorie und Empirie darstellen können (Bröder, 2000) und schätzen die Testbarkeit der vorgeschlagenen Heuristiken positiver ein. In der Tat zeigen die folgenden Jahre eine Akkumulation von empirischen Studien zu den Heuristiken der adaptiven Werkzeugkiste (vgl. auch Bröder, 2005). Am Beispiel von zwei prominenten Mitgliedern der Familie der *fast and frugal heuristics* sollen in den folgenden Abschnitten Elemente des adaptiven Werkzeugkastens näher vorgestellt werden. Dabei werden Simulationen, analytische Ergebnisse und experimentelle Studien zu zwei zentralen Heuristiken der Adaptiven Werkzeugkiste vorgestellt, der *Recognition Heuristic* und der *Take The Best-Heuristic*[16].

4.4 Die Rekognitions-Heuristik

Die Rekognitions-Heuristik oder *Recognition*-Heuristik (Goldstein & Gigerenzer, 1999, 2002) ist die einfachste der im Rahmen der *Adaptive Toolbox* diskutierten Heuristiken. Sie wird eingesetzt, um ordinale Vergleiche zwischen zwei Alternativen hinsichtlich der Ausprägung eines bestimmten Kriteriums zu entscheiden ohne direkten Zugriff auf die Kriteriumswerte selbst. Wird von zwei Objekten genau ein Objekt wiedererkannt, so wird die Entscheidung getroffen, dass dieses Objekt den höheren Kriteriumswert aufweist. Dabei lassen sich zwei Stufen der Wiedererkennung unterscheiden, auf der ersten Stufe handelt es sich um eine reine Wiedererkennung, ohne dass zusätzliche Informationen über das Objekt bekannt sind bei der zweiten Stufe liegen über die Wiedererkennung hinausgehende Kenntnisse vor. Dementsprechend kann die Heuristik als Entscheidungsbaum mit zwei Ebenen dargestellt werden (vgl. Abbildung 4.1).

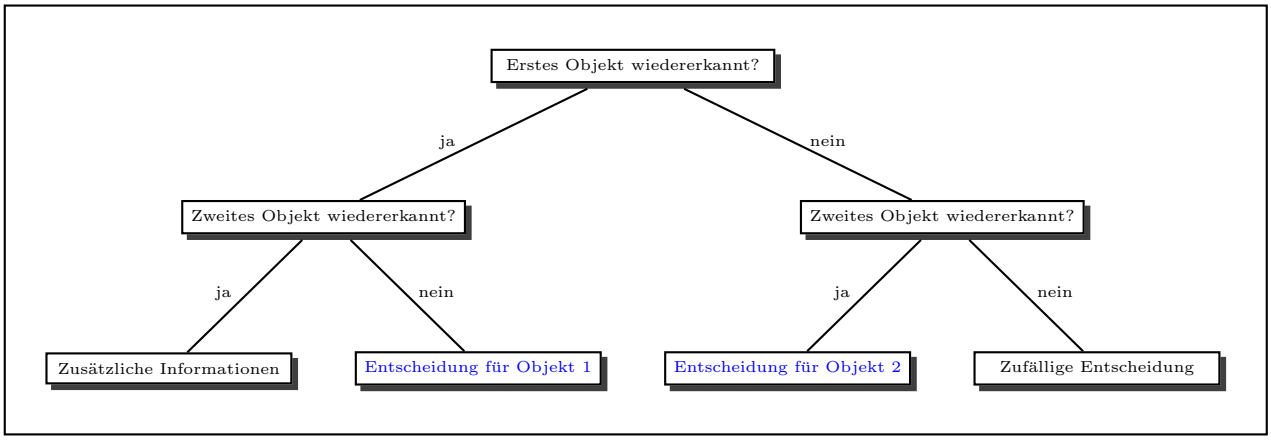


Abbildung 4.1: Entscheidungsbaum bei Anwendung der Rekognitionsheuristik

Als Antwort auf die Frage, welches von zwei Objekten den höheren Kriteriumswert aufweist, wird im Falle der Wiedererkennung nur eines der beiden Objekte, nämlich das wiedererkannte Objekt ausgewählt. Wird keines der beiden Objekte wiedererkannt, so muss geraten werden, werden beide Objekte wiedererkannt, können unter Umständen zusätzliche Informationen genutzt werden. In der Definition wird nur in solchen Fällen von einer Anwendung der Heuristik gesprochen, wenn genau ein Objekt in dem zur Entscheidung stehenden Objektpaar identifiziert werden kann. Zählt man die Frage nach dem Wiedererkennen der Objekte selbst als im Sinne einer Suchprozedur zur Heuristik gehörend, so wird auch in den

[16] Auf eine Darstellung einer Vielzahl von weiteren im Programm bearbeiteten Heuristiken wird an dieser Stelle verzichtet, unter anderem auf eine Diskussion der *Priority Heuristic* (Brandstätter, Gigerenzer & Hertwig, 2006), die als Alternative zur *Prospect Theory* entwickelt wurde. Dies soll ausdrücklich betonen, dass das Programm nicht auf die Auswahl der diskutierten Heuristiken reduziert werden kann.

beiden anderen Fällen zumindest von einer versuchten Anwendung gesprochen werden können. Formal lassen sich die Probleme, auf die die Heuristik angewendet werden kann, als Kategorisierungsprobleme bezeichnen (Martignon & Hoffrage, 2002): Ein Objektpaar (a, b) soll bezüglich des Kriteriums X kategorisiert werden als $X_a > X_b$ oder $X_b > X_a$. Die Matrix aller Objekte der untersuchten Klasse mit allen Informationen über die Objekte heißt *Environment*. Variablen sollen auch als *Cues* bezeichnet werden. Für die Rekognitionsheuristik wird ein einziger Cue benötigt: die Wiedererkennungsinformation.

4.4.1 Bedingungen für die ökologische Rationalität der Heuristik

Die Heuristik ist domänenspezifisch, sie lässt sich nicht auf jede Fragestellung erfolgreich anwenden. Entscheidend für die Möglichkeit, von der Anwendung der Heuristik zu profitieren, ist eine positive Korrelation zwischen Wiedererkennung und Kriterium. Das Zustandekommen dieser Korrelation kann über einen oder mehrere Mediatoren erfolgen, die einen Beobachter mit den Objekten oder auch nur den Namen der Objekte konfrontieren, wobei die Häufigkeit dieser Konfrontationen wiederum in positivem Zusammenhang mit den Werten des Kriteriums stehen muss. In diesem Fall kann aus dem Wiedererkennen der Objekte auf die Höhe der Ausprägung des Kriteriums zurückgeschlossen werden. Die Korrelation zwischen Kriterium und Konfrontationshäufigkeit wird von Goldstein und Gigerenzer (1999) als ökologische Korrelation, die Korrelation zwischen dieser Häufigkeit und der Wiedererkennung als Ersatzkorrelation bezeichnet, während der inferierte Zusammenhang zwischen Wiedererkennen und Kriteriumswerten Rekognitionsvalidität genannt wird (s. Abbildung 4.2). Die Rekognitionsvalidität α wird genauer definiert als

$$\alpha \equiv \frac{n_k}{n_k + n_f}, \tag{4.1}$$

wobei n_k die Anzahl korrekter Inferenzen darstellt über alle Objektpaare hinweg gezählt, bei denen genau ein Objekt erkannt wird, und n_f die entsprechende Anzahl falscher Inferenzen. Die dargestellte Nutzung eines vermittelten Zusammenhangs nennen die Autoren die ökologische Rationalität der Recognition-Heuristik. Als Beispiel nennen die Autoren unter anderem das Dreieck, dass aus der finan-

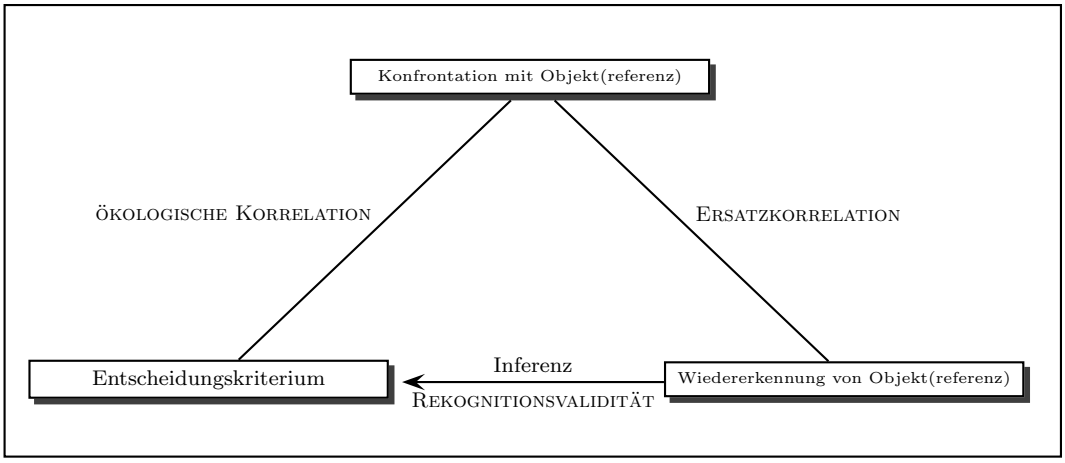


Abbildung 4.2: Darstellung der ökologischen Rationalität der Recognition-Heuristik, modifiziert nach Goldstein und Gigerenzer (1999)

ziellen Ausstattung von Universitäten (Kriterium), der Anzahl der Nennungen in Zeitungen (Mediator) und der Wiedererkennung der Namen durch einen Beobachter gebildet wird. Die Betrachtung von Wiedererkennung als binäre Variable grenzt die Rekognitionsheuristik von der Verfügbarkeis-Heuristik aus dem Heuristics-and-Biases-Programm und verwandten Konzepten ab.

Die Heuristik gehört zu den *fast and frugal heuristics*, die auf psychologischen Kapazitäten und heuristischen Prinzipien beruhen,und deren Geschwindigkeit durch einfache Entscheidungsmodelle erreicht wird, für die nicht die Schätzung zahlreicher Parameter erforderlich ist. Ihre Frugalität erwächst durch die Suffizienz geringer Informationsmengen. Die Rekognitionsheuristik ist in dieser Hinsicht extrem, da sie lediglich eine einzige binär kodierbare Information für jedes Objekt benötigt, um zu einer Entscheidung in einem einzigen Schritt zu gelangen. Dementsprechend verkörpert sie die Prinzipien des *One-Reason*

Decision Making und des *Ignorance-Based Decision Making* (Goldstein & Gigerenzer, 1999) (s.u.)[17]. Bröder (2001) wirft der Betrachtung der ökologischen Rationalität allerdings die Aufdeckung einer Trivialität vor:

> Es ist in der Rückschau eine ausgemachte Trivialität (bzw. offensichtliche Tautologie), dass Rekognition immer dann erfolgreich ein Kriterium vorhersagen kann, wenn sie mit ihm korreliert ist. Interessant an den Beispielen ist jedoch, dass auf diese schlichte Tatsache erst einmal jemand kommen musste (nämlich die ABC Research Group). (S. 160)

Allerdings kann gegen diese nüchterne Betrachtungsweise eingewandt werden, dass die Feststellung einer Korrelation zwischen Kriterium und Rekognition empirischen Gehalt haben *kann*.

4.4.2 Analytische Ergebnisse

Im folgenden sollen einige analytische Ergebnisse zur Rekognitionsheuristik präsentiert werden, dabei wird zunächst die Vorhersageleistung der Heuristik formal hergeleitet. Eine weitere Untersuchung ist dem *Less is More*-Effekt gewidmet.

4.4.2.1 Analyse der Vorhersagefähigkeit

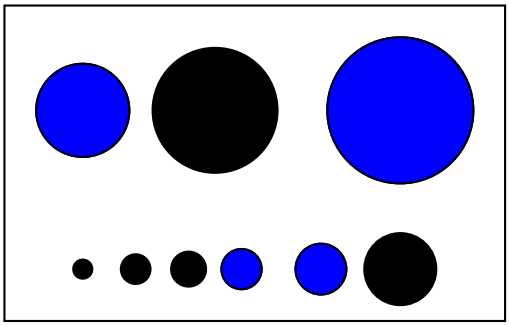


Abbildung 4.3: Beispielobjekte für eine Analyse der Vorhersageleistung der Rekognitionsheuristik

Die Herleitung einer Formel zur Berechnung der Genauigkeit der Rekognitionsheuristik soll an einem Beispiel nachvollzogen werden. Gegeben seien die in Abbildung 4.3 dargestellten Kreise als Objektmenge. Als Kriterium soll der Radius bzw. der Flächeninhalt der Kreise dienen, die Wiedererkennung sei durch die Farbe der Kreise gekennzeichnet. Blaue Kreise stellen wiedererkannte Objekte dar, schwarze Kreise repräsentieren Objekte, die nicht wiedererkannt wurden. Das Beispiel ist selbstverständlich als künstliches Beispiel zu verstehen, da die Kriteriumswerte visuell vorliegen.

Bei N Objekten gibt es zunächst einmal $\frac{N(N-1)}{2}$ verschiedene mögliche Paarvergleiche, im Beispiel $\frac{9(9-1)}{2} = 36$. Die Anzahl der wiedererkannten Objekte soll als r bezeichnet werden (im Beispiel $r = 4$). Vergleiche, in denen nur das erste Objekt wiedererkannt wird sollen genauso behandelt werden wie Vergleiche, in denen nur das zweite Objekt wiedererkannt wird, dementsprechend reduzieren sich die vier Bedingungen, die in den Endknoten in Abbildung 4.1 dargestellt wurden, für die Analyse auf drei. Die drei verbleibenden Kategorien sind für das Beispiel in Abbildung 4.4 getrennt voneinander dargestellt.

[17] Abweichend vom Wortlaut des Textes, aber den Beispielen zu Folge im Sinne des Textes, soll aber noch einmal differenziert werden zwischen dem Wiedererkennen der Objekte selbst und dem Wiedererkennen von Referenzen auf Objekte. So ist das Objekt Stanford University nicht mit dem Namen „Stanford University“ identisch. Als Mediatoren kommen Konfrontationen mit beiden Kategorien in Frage, entscheidend ist die Stimmigkeit mit dem Modus der Abfrage. So könnte ein Bild der Universität keine Wiedererkennung auslösen, auch wenn der Name der Universität oft gelesen wurde. In ähnlicher Weise wird hier auf die Duplizierung der Umgebung in eine unbekannte und eine zugängliche Umgebung als Behälter für Kriterium und Mediator abgewichen, da nicht die Zugänglichkeit der Umgebung, sondern die Zugänglichkeit der Kriterienwerte die entscheidende Rolle spielt. Wären die Kriterienwerte direkt zugänglich, könnte die Entscheidung zwischen den Objekten ebenso direkt getroffen werden. In vielen Fällen kann zudem mehr als ein einzelner Mediator für die Anwendung der Heuristik relevant sein, was die Begriffe der ökologischen Korrelation und der Ersatzkorrelation unschärfer werden lässt. Für einen einzelnen gut definierten Mediator lassen sich dementsprechend Korrelationen berechnen, bei mehreren Mediatoren können komplexere Beziehungsgefüge auftreten.

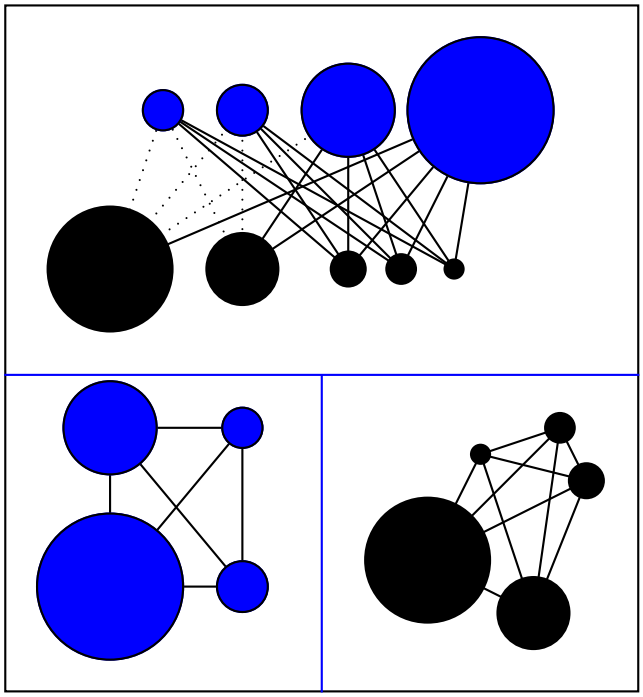


Abbildung 4.4: Aufteilung der Paarvergleiche nach Rekognition der Beispielobjekte

Der Anteil der Paarvergleiche, bei denen keines der beiden Objekte erkannt wird ergibt sich als

$$a_{00} = \frac{(N-r)(N-r-1)}{N(N-1)}, \tag{4.2}$$

der Anteil der Vergleiche, bei denen beide Objekte erkannt werden, ergibt sich als

$$a_{11} = \frac{r(r-1)}{N(N-1)}, \tag{4.3}$$

der Anteil der Vergleiche, bei denen genau eines der beiden Objekte erkannt wird, ergibt sich als

$$a_{10} = \frac{r(N-r)2}{N(N-1)}. \tag{4.4}$$

Im Beispiel sind die Anteile $a_{00} = \frac{10}{36}$, $a_{11} = \frac{6}{36}$ und $a_{10} = \frac{20}{36}$. Neben der bereits definierten Rekognitionsvalidität α müssen zur Berechnung der Genauigkeit zwei weitere Werte festgesetzt werden, die Wissensvalidität β und die implizite Entscheidungskompetenz γ. Dabei stellt β die Wahrscheinlichkeit dar, dass bei einem Vergleich zwischen zwei wiedererkannten Objekten das Objekt mit dem höheren Kriteriumswert korrekt identifiziert wird[18]. Der Parameter γ ist in der Formulierung von Goldstein und Gigerenzer (1999) nicht enthalten. Hier wird er als die Wahrscheinlichkeit betrachtet, dass bei einem Vergleich zwischen nicht wiedererkannten Objekten die richtige Wahl getroffen wird. Die ursprüngliche Formulierung ergibt sich für $\gamma = \frac{1}{2}$. Hier wird aber davon ausgegangen, dass beispielsweise aus den Namen von Objekten selbst (möglicherweise auch irreführende) Informationen extrahiert werden, so dass die Wahrscheinlichkeit der korrekten Auswahl nach oben oder unten abweichen kann. So könnte etwa bei der Frage, ob der (fiktive) Ort Klein-Nauernstein oder der (fiktive) Ort Groß-Königsburg mehr Einwohner hat, mit höherer Wahrscheinlichkeit eine Entscheidung zugunsten des zweiten Ortes zu erwarten sein, auch wenn keiner der Namen wiedererkannt werden kann. Im Beispiel lässt sich zumindest α bestimmen als $\alpha = \frac{15}{20} = 0,75$, da 5 von 20 Vergleichen zwischen wiedererkannten und nicht wiedererkannten Objekten

[18] Die Bezeichnung „Wissen“ ist in Bezug auf diesen Parameter etwas unglücklich, wenn berücksichtigt wird, dass die Heuristik auch für Kriteriumswerte zur Anwendung gebracht werden kann, die erst zeitlich versetzt erhoben werden können, so dass der Parameter allgemeiner eher die Kompetenz beschreibt, aufgrund des Wissens über die wiedererkannten Objekte eine gute Entscheidung zu treffen. Unter Umständen kann dieser Parameter aber nicht zum Zeitpunkt der Entscheidung bestimmt werden.

bei Anwendung der Heuristik zu falschen Entscheidungen führen. Allgemein berechnet sich die Genauigkeit der Heuristik im Sinne des Anteils korrekter Inferenzen bei N Objekten und r wiedererkannten Objekten als

$$g(N,r) = \alpha a_{10} + \beta a_{11} + \gamma a_{00} = \alpha \frac{r(N-r)2}{N(N-1)} + \beta \frac{r(r-1)}{N(N-1)} + \gamma \frac{(N-r)(N-r-1)}{N(N-1)}. \tag{4.5}$$

Dabei bleibt zu betrachten, ob α, β und γ entweder als anwendungsspezifische Konstanten oder als Ausprägungen von Zufallsvariablen zu verstehen sind, was auch vom Status der Kriteriumsvariablen abhängt, deren Ausprägungen wiederum in komplexer Art und Weise von denen unzugänglicher weiterer Variablen oder Zufallsprozessen abhängen können. Aus diesem Grund ist Gleichung 4.5 zur Spezifikation der Inferenzgenauigkeit nur dann ausreichend, wenn entweder die exakten Werte für die Parameter für spezifische Kombinationen aus r und N bekannt sind oder die Parameterverteilungen bestimmt werden können. Der ermittelte Wert $g(N,r)$ lässt sich auch nur dann als Angabe für die Genauigkeit der Heuristik verstehen, wenn die erweiterte Variante, also der gesamte oben entwickelte Entscheidungsbaum, als Anwendungsbereich der Heuristik verstanden wird, denn bei ausschließlicher Betrachtung der Fälle, in denen genau ein Objekt wiedererkannt wird, ist die Genauigkeit bereits durch α gegeben.

4.4.2.2 Der Less-is-More Effekt

Ein interessantes Ergebnis, das bei der Analyse der Funktionsweise der Rekognitionsheuristik gefunden werden kann, ist der so genannte „Less-is-More“ Effekt. Als Beispiel verwenden Goldstein und Gigerenzer (1999) die Betrachtung der Familie MacAllister mit drei Brüdern, die mit der Aufgabe konfrontiert werden, Entscheidungen bezüglich der Größe deutscher Städte zu treffen[19]. Dabei hat der jüngste Bruder nie von Deutschland gehört und kennt keine deutschen Städte, der mittlere Bruder erkennt 25, der älteste Bruder kennt alle 50 zur Diskussion stehenden deutschen Städte. Geht man bei den Brüdern von $\gamma_1 = \gamma_2 = \gamma_3 = 0,5$ aus und setzt man $\alpha_2 = 0,8$, zudem $\beta_2 = \beta_3 = 0,6$, dann ist bei Anwendung der Rekognitionsheuristik das Ergebnis des mittleren Bruders das beste der drei Brüder, denn keiner der beiden anderen kann die Heuristik im engeren Sinne zur Anwendung bringen, und der Vorteil gegenüber den Brüdern ergibt sich aus $\alpha_2 > \beta_3$ und $\alpha_2 > \gamma_1$ mit $\beta_2 = \beta_3$ und $\gamma_2 = \gamma_1$. Setzt man N konstant, so ergibt sich

$$\begin{aligned} g(n) &= \frac{1}{N^2-N}((2nN-2n^2)\alpha + (N^2-2nN+n^2-N+n)\gamma + (n^2-n)\beta) \\ &= \frac{1}{N^2-N}(n^2(-2\alpha+\beta+\gamma) + n(2N\alpha + (1-2N)\gamma - \beta) + (N^2-N)\gamma). \end{aligned} \tag{4.6}$$

Damit ist

$$g'(n) = \frac{2}{N^2-N}(-2\alpha+\beta+\gamma)n + \frac{1}{N^2-N}(2N\alpha + (1-2N)\gamma - \beta). \tag{4.7}$$

Ein mögliches Maximum der Funktion ergibt sich bei $g'(n) = 0$, also

$$\frac{2}{N^2-N}(-2\alpha+\beta+\gamma)n + \frac{1}{N^2-N}(2N\alpha + (1-2N)\gamma - \beta) = 0, \tag{4.8}$$

bzw. für $N > 1$ und $\alpha \neq \frac{\beta+\gamma}{2}$

$$n = \frac{-2N\alpha + \beta + (2N-1)\gamma}{-4\alpha + 2\beta + 2\gamma}. \tag{4.9}$$

Dabei ist

$$g''(n) = \frac{2}{N^2-N}(-2\alpha+\beta+\gamma) \tag{4.10}$$

und aus $g''(n_0) < 0$ folgt

$$\frac{2}{N^2-N}(-2\alpha+\beta+\gamma) < 0 \tag{4.11}$$

bzw.

$$\alpha > \frac{\beta+\gamma}{2}. \tag{4.12}$$

[19] In anderen Varianten wird die ansonsten bedeutungsgleiche Geschichte dreier französischer Schwestern erzählt.

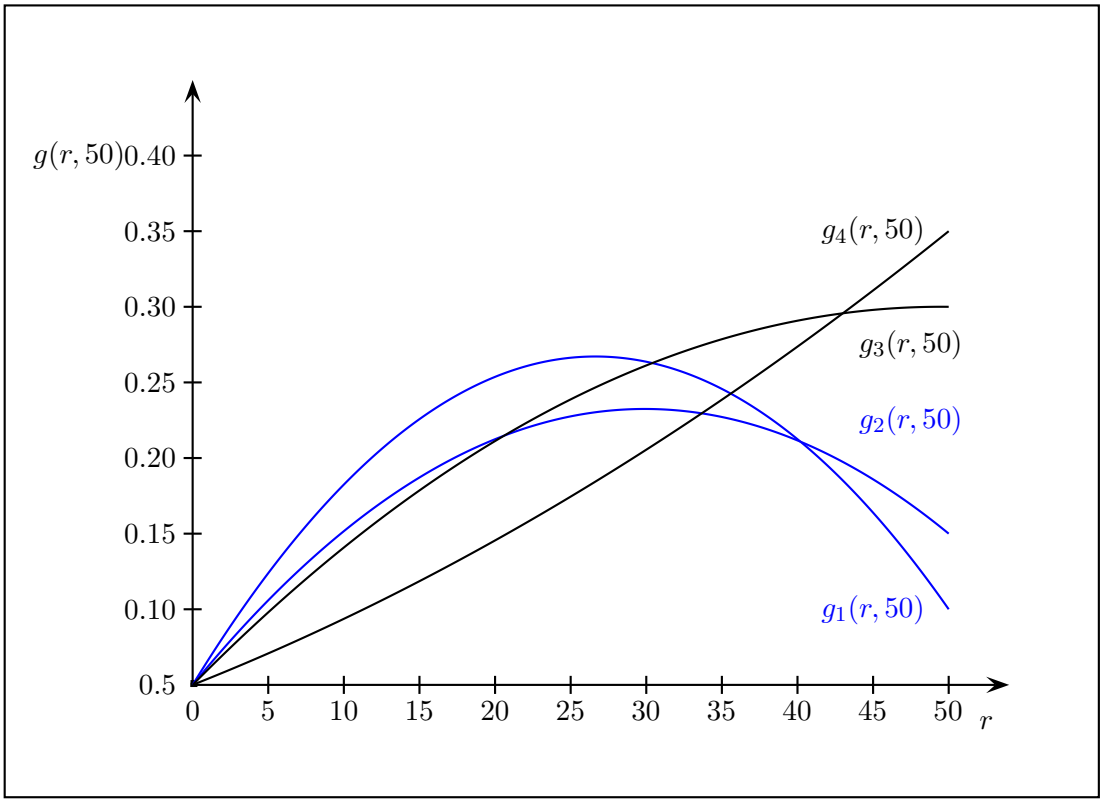


Abbildung 4.5: Illustration des Less-Is-More Effekts bei 50 Vergleichsobjekten und vier Kombinationen von α und β bei $\gamma = 0.5$: Der Effekt tritt auf für g_1 und g_2, da $\alpha_1 = 0.9 > \beta_1 = 0.55$ und $\alpha_2 = 0.8 > \beta_2 = 0.6$. Er tritt nicht auf bei den Funktionen g_3 ($\alpha_3 = \beta_3 = 0.75$) und g_4 ($\alpha_4 = 0.6 < \beta_4 = 0.8$).

Ist also die Rekognitionsvalidität größer als der Mittelwert der anderen beiden Parameter, gibt es ein Maximum an der Stelle

$$n_{max} = \frac{-2N\alpha + \beta + (2N-1)\gamma}{-4\alpha + 2\beta + 2\gamma}. \tag{4.13}$$

Dabei ist noch ungeklärt, ob sich dieser Wert innerhalb des zulässigen Wertespektrums von n befindet. Setzt man beispielsweise $\beta = \alpha > \gamma$, so ergibt sich

$$\begin{aligned} n_{max} &= \frac{-2N\alpha + \alpha + (2N-1)\gamma}{-4\alpha + 2\alpha + 2\gamma} \\ &= \frac{-2N(\alpha-\gamma) + (\alpha-\gamma)}{-2(\alpha-\gamma)} \\ &= N - \frac{1}{2}. \end{aligned} \tag{4.14}$$

.

Dieser kontraintuitive Befund erklärt sich aus der Betrachtung der Komponenten $(N-n)(N-n-1)$, die für $0 < N-n < 1$ negative Werte annimmt. Da n in der Aufgabenstellung aber nur ganzzahlige Werte zwischen 0 und N annehmen kann, würde kein Less-is-More-Effekt sichtbar. Da ein solcher Effekt außer für den trivialen Fall, in dem $\gamma > \beta$ (der sicherlich auch den Titel tragen könnte), nur zu erwarten ist, wenn $\alpha > \beta$, soll die Differenz x mit $\beta = \alpha - x$ berechnet werden, für die ein Maximum an der Stelle $n_{max} = N - y$ erreicht wird. Setzt man $n_{max} = N - y$, ergibt sich aus Gleichung 4.13

$$N - y = \frac{-2N\alpha + \alpha - x + (2N-1)\gamma}{-4\alpha + 2(\alpha - x) + 2\gamma}. \tag{4.15}$$

und dementsprechend aufgelöst für y

$$y = N - \frac{(1-2N)\alpha - x + (2N-1)\gamma}{-2\alpha - 2x + 2\gamma}, \tag{4.16}$$

bzw.

$$y = \frac{-2N\alpha - 2Nx + 2N\gamma + (2N-1)\alpha + x + (1-2N)\gamma}{-2\alpha - 2x + 2\gamma} \tag{4.17}$$

und damit

$$y = \frac{-\alpha + (1-2N)x + \gamma}{-2\alpha - 2x + 2\gamma}. \tag{4.18}$$

Die Position des Maximums ist damit damit abhängig von α, γ und der Differenz zwischen α und β. Verändert man die Fragestellung und sucht nach den Bedingungen, für die der Abstand zwischen Maximalwert N und n_{max} mindestens z beträgt, also $y \geq z$ ergibt sich

$$\frac{-\alpha + (1-2N)x + \gamma}{-2\alpha - 2x + 2\gamma} \geq z. \tag{4.19}$$

Löst man diesen Wert nach x auf, erhält man

$$-2z\alpha - 2zx + 2z\gamma \geq -\alpha + (1-2N)x + \gamma, \tag{4.20}$$

bzw.

$$(1-2z)\alpha + (2z-1)\gamma \geq (1+2z-2N)x \tag{4.21}$$

und schließlich

$$x \geq \frac{2z-1}{2N-2z-1}(\alpha - \gamma). \tag{4.22}$$

Damit lässt sich die Frage beantworten, wie groß der Abstand zwischen α und β mindestens sein muss, damit der Less-is-More-Effekt auftreten kann. Die Bedingung $y \geq 1$ ergibt sich für $z = 1$, also Für $z = 1$:

$$x \geq \frac{1}{2N-3}(\alpha - \gamma). \tag{4.23}$$

Für $\gamma = 0,5$ und $\alpha = 0,8$ ergibt sich zum Beispiel bei $N = 50$

$$x \geq \frac{1}{2 \cdot 50 - 3}(0,8 - 0,5) = \frac{3}{970} \approx 0,003. \tag{4.24}$$

Damit stellt sich der Effekt ein bei $\beta \approx 0,797$.

Die angestellten Berechnungen stellen eine Verallgemeinerung der Ergebnisse von (Goldstein & Gigerenzer, 2002) dar, die das Auftreten von Less-is-More-Erffekten für $\alpha > \beta$ definieren.

Reimer und Katsikopoulos (2004) demonstrieren, dass der Anteil der wiedererkannten Stadtzahlen, für die das Genauigkeitsergebnis über dem Genauigkeitsergebnis für vollständige Bekanntheit liegt (für den Fall dass $\alpha < \beta$), mit α steigt und mit β sinkt. Letzteres zeigt sich auch bei einer näheren Betrachtung von Gleichung 4.18.

4.4.3 Empirische Studien zur Rekognitionsheuristik

4.4.3.1 Städtevergleiche

In einer Befragung wurde 22 amerikanischen Teilnehmern die Aufgabe gestellt, die 25 bzw. 30 größten Städte Deutschlands in Namenspaarvergleichen hinsichtlich ihrer Bevölkerungszahl zu vergleichen (Goldstein & Gigerenzer, 1999). Gleichzeitig wurde die Menge der bekannten Städtenamen erfasst. Durch die unterschiedliche Größe der Mengen der bekannten Städte gibt es bei Vorgabe aller 300 bzw. 435 Vergleiche schwankende Zahlen für die Anzahl der Möglichkeiten, die Rekognitionsheuristik anzuwenden. Dabei gingen in den relevanten Fällen bei den Teilnehmern zwischen 73% und 100% der Wahlen (bei einem Median von 93%) mit den Wahlen bei Anwendung der Rekognitionsheuristik konform. Dementsprechend resümieren Goldstein und Gigerenzer (1999):

> This simple test of the recognition heuristic showed that people adhere to it the vast majority of the time. (S. 50)

Dies ist zwar stark formuliert, da lediglich die Ähnlichkeit der Wahlergebnisse ohne Beobachtung des Wahlprozesses untersucht werden konnte und alternative Modelle ähnliche Vorhersagen machen könnten, trotzdem ist die Übereinstimmung der Wahlentscheidungen sicherlich eine notwendige Bedingung für die Angemessenheit der Rekognitionsheuristik als deskriptives Modell. Die Versuchspersonen erhielten in einer Verschärfung der Testbedingungen die Auskunft, dass neun der dreißig größten Städte Fußballmannschaften besitzen und dass in 78% der möglichen Vergleiche zwischen Städten mit und ohne Fußballmannschaft die Stadt mit Mannschaft die höhere Bevölkerungszahl aufweist[20]. Die Information bezüglich des Besitzes einer Fußballmannschaft wurde nun für acht nur vorgeblich randomisiert gezogene Städte bestimmt, die für jede Versuchsperson aus vier in einer Vorstudie eher bekannten und vier eher unbekannten Städten mit Fußballmannschaft bestanden. Trotz erfolgsabhängiger Bezahlung wichen die meisten Wahlen der 21 Teilnehmer auch bei konfligierenden Zusatzinormationen nicht von den durch die Rekognitionsheuristik prognostizierten Wahlen ab.

Zur Überprüfung des Less-is-More Effekts wurden 52 Studenten der University of Chicago gebeten, die 22 größten amerikanischen und die 22 größten deutschen Städte getrennt voneinander in Zufallsziehungen paarweise zu vergleichen. Obwohl hier definitives Wissen um Teile der amerikanischen Stadtgrößenreihenfolge vorausgesetzt werden kann, waren die Ergebnisse für die deutschen Städte im Median besser als für die amerikanischen Städte (73% versus 71%). Für 26 Studenten wurde zusätzlich die Anzahl der erkannten Städte erhoben. Dieser Gruppe erkannte durschnittlich 12 deutsche Städte, und ihre Wahlen stimmten in durchschnittlich 89% der relevanten Fälle mit den Prognosen der Rekognitionsheuristik überein.

Bei der Suche nach möglichen Mediatoren wurde die Hypothese überprüft, ob die Frequenz der Stadtnennungen in Zeitungen als ein solcher fungieren kann. Hierzu wurde getrennt für alle deutschen Städte mit mehr als 100000 Einwohnern die Anzahl der Artikel in der *Chicago Tribune* zwischen 1985 und 1997 gezählt, in denen das Wort „Germany“ zusammen mit Stadtnamen in jeder möglichen Variante enthalten war. Die Spearman-Korrelationen zwischen der Wiedererkennung des Stadtnamens durch 67 Studenten in Chicago und der Anzahl der den Stadtnamen enthaltenen Artikel zum einen (Surrogatkorrelation) und der Korrelation zwischen Artikelanzahl und Einwohnerzahl (ökologische Korrelation) ergaben sich als $r_s = .79$ und $r_o = 0,70$, während die Korrelation zwischen Wiedererkennung und Einwohnerzahl bei $r_r = 0,60$ lag. Eine analoge Studie über die Nennungen amerikanischer Städte in der Wochenzeitschrift *Die Zeit* und deutschen Teilnehmern kommt zu ähnlichen Ergebnissen bei etwas höheren Korrelationen. Da die Medienpräsenz der Städte mit der Wiedererkennung höher korreliert ist als mit der Einwohnerzahl, kommt sie als Mediator in Frage[21]. Die Autoren berichten weiterhin von experimentellen Studien, in denen der Erfolg von britischen Fußballteams durch türkische Studenten nahezu ebenso gut vorhergesagt wurden wie durch britische Studenten, oder die Einschätzung des Berühmtheitsgrades von Personen durch reine Präsentation der Namen manipuliert werden konnte.

Eine Anwendung im Bereich der Entscheidungsfindung durch Gruppen untersuchen Reimer und Katsikopoulos (2004) an einer Stichprobe von 90 Studierenden der Freien Universität Berlin. Dabei zeigt sich zum einen ein größerer Einfluss der Anwender der Rekognitionsheuristik:

> *It is the members that use the recognition heuristic that are more influential in the process of combining judgments.* (S. 1020)

Die ökologische Validität der Heuristik zeigt sich weiterhin in der besseren Genauigkeit der Gruppenentscheidungen ,die auf einer Mehrheit von Personen beruhen, die die Rekognitionsheuristik anwenden können, im Vergleich zu Mehrheitsentscheidungen, die auf einer Mehrheit von Personen beruhen, die auf Wissen zurückgreifen können.

[20] Bei 189 möglichen Vergleichen sind dies allerdings 147,42 Vergleiche, so dass gerundete Zahlen vorliegen müssen.

[21] Es soll hier angemerkt werden, dass die Verbesserung der Vorhersagefähigkeit nur die notwendige Bedingung für einen Mediator für Wiedererkennung ist, da durchaus mehrere Mediatoren in komplexem Zusammenspiel die Wiedererkennbarkeit einer Stadt bestimmen könnten. Eine genauere Erklärung ließe sich allerdings nur durch geeignete experimentelle Studien finden. Zudem ist die gewählte Operationalisierung von Medienpräsenz nicht für alle Versuchspersonen in gleicher Weise relevant. Dementsprechend kann die Formulierung von (Goldstein & Gigerenzer, 1999):

> These results suggest that individual recogntion is more in tune with the media than with the actual environment, which indicates that city name recognition may come largely from the media. True population size is unknown to most people, but they can rely on mere recognition to make a fairly accurate guess. (S.55)

als plausible, aber nicht als zwingend aus den Daten hervor gehende Interpretation eingeschätzt werden, da allein die Korrelation zwischen Wiedererkennung und Medienpräsenz als über Drittvariablen vermittelt gedacht werden kann, die ihrerseits Ansprüche auf eine Rolle als Mediator stellen könnten.

4.4.3.2 Aktienmarkt

Borges, Goldstein, Ortmann und Gigerenzer (1999) berichten über den Vergleich zwischen der Rekognitionsheuristik und verschiedenen Benchmarks für die Aktienauswahl. Dabei wurde eine Variante der Rekognitionsheuristik zum Einsatz gebracht:

> A generalization of the recognition heuristic for choosing a subset of objects from a larger set, which can be applied to investment decisions, is: *When choosing a subset of objects from a larger set, choose the subset of recognized objects.* (S. 61)

Insgesamt 480 Laien und Experten in Deutschland und den USA wurden bezüglich insgesamt 798 deutschen und amerikanischen Unternehmen hinsichtlich der Wiedererkennung der Firmennamen befragt. Dabei erkannten amerikanische Experten die meisten Firmen, überraschenderweise gefolgt von amerikanischen Laien, auf den dritten und vierten Platz kamen deutschen Experten und deutschen Laien. Um die Leistung der Rekognitionsheuristik in Bezug auf die Aktienauswahl zu testen, wurden Portfolios zusammengestellt, basierend auf den Rekognitionsinformationen: Für jede der vier Gruppen wurde ein Portfolio mit Aktien von zehn Firmen des eigenen Landes zusammengestellt, die von mindestens 90% der Stichprobe erkannt wurden, und dazu ein Portfolio mit den zehn am häufigsten wiedererkannten Aktien von Firmen des anderen Landes. Die Erträge der Portfolios wurden verglichen mit den Erträgen von Portfolios nicht wiedererkannter Firmen, den Marktindizes, Zufallsportfolios und *mutual funds*. Dabei setzten zwei der Autoren eigenes Geld auf ein durch die Angaben von Laien generiertes Portfolio. In den sechs Monaten der Studie erbrachte das amerikanische Portfolio auf der Basis der Wiedererkennung durch deutsche Laien einen Zuwachs von 47% gegenüber einem Zuwachs von 13% für die nicht wiedererkannten Firmen. Bei amerikanischen Firmen liegen diese Werte bei 57% und 14% respektive. Die beiden Portfolios liegen in ihren Erträgen dabei über dem DAX-Index mit einem Zuwachs von 37%. Die Ergebnisse für die amerikanischen Gruppen und Aktien sind weniger deutlich, am schwächsten sind die Egebnisse der durch die Einschätzung von amerikanischen Laien und Experten erstellten Portfolios mit US-Firmen. Bei der direkten Befragung, welche Aktien für eine Investition ausgewählt werden sollten, lagen die Ergebnisse der Laien für deutsche Aktien sogar höher als die von Experten, deren Auswahl zu Geldverlust geführt hätte. Insgesamt stellen die Ergebnisse einen Beleg für eine gute ökologische Rationalität der Rekognitionsheuristik in einer angewandten Fragestellung dar:

> The stock market may be a complex real-world environment in which lack of recognition is not completely random, but rather systematic and informative, in investments, there may be wisdom in ignorance. (Borges et al., 1999, S. 72)

Die Zuwächse bei den Rekognitions-Portfolios übertreffen dabei auch die Zuwächse von zwei zum Vergleich ausgewählten *mutual funds* in sechs von acht Tests. Vor einer undifferenzierten Übernahme der Strategie für die eigene Vermögensvermehrung muss allerdings gewarnt werden. Boyd (2001) demonstriert durch eine Anwendung der Heuristik in einem so genannten Bärenmarkt (im Gegensatz zu einem Bullenmarkt) mögliche Bedingungen für den Anwendungserfolg:

> The ABC Group's tests support the use of the heuristic in a bull market environment, but this study, conducted in a bear market, reaches the opposite conclusion. We find that a high degree of company name recognition can lead to disappointing investment results in a down market, and it can also be beat by pure ignorance. Virtually the only finding of the ABC Group's study that we match here is that Americans are not very good at picking American stocks to beat the market (Boyd, 2001, S. 155).

In ähnlicher Weise relativiert Over (2000) den erwiesenen Erfolg der Heuristik:

> Financial experts would reply that high recognition and high capitalisation tend to go together, and that high-capitalisation stocks happened to do well in the bull market of those months. The experts have to try to anticipate changes in the market, and do well in the longer term, when the market is not so homogeneous. (Over, 2000, S. 188)

Einige weitere kritische Studien zur Rekognitionsheuristik sollen dargestellt werden.

4.4.3.3 Kritische Studien zur Rekognitionsheuristik

Oppenheimer (2003) widmet sich in zwei Studien der Hypothese, dass Rekognition in den früheren Studien zur Rekognitionsheuristik mit anderen Cues konfundiert worden sei. In einem ersten Experiment lässt er zwischen Städten hinsichtlich der Stadtgröße auswählen, verwendet dabei aber bekannte Städte aus der näheren Umgebung, die als kleine Städte bekannt sein sollten. In der Tat neigen Versuchspersonen dazu, im Vergleich mit anderen unbekannten Städten, die anderen Städte als größer einzuschätzen als die bekannten (kleinen) Städte:

> People are clearly using information beyond recognition when making judgments about city size. These results are intriguing because they are not predicted by the recongition heuristic. (S. B4-B5)

In einem zweiten Experiment verwendet er bekannte Stadtnamen, deren Bekanntheitsgrad auf besonderen, aber verschiedenen Gründen beruht, wie Tschernobyl (Nuklearkatastrophe), New Haven (Standort der Yale University) und Timbuktu (Sprichwort). Nur eine der als bekannt vorausgesetzten Städte ist aus dem Grund bekannt, eine lokale Nachbarstadt zu sein. Die verwendeten Städte wurden nur in 40% der Trials als größer eingeschätzt, wenn sie mit nicht wiedererkannten Städten gepaart wurden. Dementsprechend urteilt Oppenheimer (2003):

> An important empirical challenge in testing fast and frugal accounts of cognition is presenting people with problems that are sufficiently complex to allow for complex reasoning to occur. If we don't deal with the complexity that exists in real problems, we may discover a misleadingly simple story about cognitive processes. (S. B8)

Newell und Shanks (2004) untersuchen die Frage, ob der Rekognitionsinformation ein besonderer Status in der Informationsverarbeitung zugewiesen werden kann. In einem Forced-Choice-Experiment mit zwei Alternativen wurde eine Investitionsentscheidung in eine von zwei angebotenen Firmen gefordert. Verschiedene Firmennamen wurden wiederholt, um Wiedererkennung dieser Namen zu indizieren, zusätzlich konnten Beurteilungen dreier unterschiedlich guter Berater gekauft werden. Die Autoren finden in den Ergebnissen das Muster, dass die Rekognitionsinformation bei hoher Validität von Teilnehmern genutzt wird, bei geringer Validität aber ignoriert wird. Auch in Fällen, in denen die Rekognitionsheuristik angewandt werden könne, wird nicht auf den Erwerb zusätzlicher Informationen verzichtet. Einen bescheideneren Hinweis auf einen besonderen Status der Wiedererkennung sehen die Autoren in der Beobachtung, dass die Nützlichkeit des Cues von den Teilnehmern im Allgemeinen überbewertet wurde. Die Autoren halten fest:

> [...] recognition, like any other indicator of performance, needs to be considered carefully and interpreted, rather than used simply to determine an inference. Importantly, our results suggest that, given the appropriate learning environment, this is indeed what people do. (Newell & Shanks, 2004, S. 934)

Richter und Späth (2006) untersuchen in einem ersten Experiment das Entscheidungsverhalten bei der Aufgabe, die Größe von Tierpopulationen einzuschätzen. Dabei wurden Tiernamen verwendet, die von Probanden entweder nicht wiedererkannt wurden, wiedererkannt wurden ohne weitere Informationen oder wiedererkannt wurden bei Verfügbarkeit weiterer aufgabenrelevanter Information. In den Ergebnissen sehen sie die Tendenz, dass Rekognitionsinformation einen großen Einfluss auf die Auswahl hat, der bei Verfügbarkeit zusätzlicher Information aber stark reduziert wird. In diesem Sinne wird Rekognitionsinformation von Versuchspersonen scheinbar kompensatorisch verwendet. In einem zweiten Experiment sollte die Sicherheit von Fluggesellschaften eingeschätzt werden, wobei positive und negative Zusatzinformationen bei bekannten Fluglinien zur Verfügung stehen. Im Ergebnis findet sich erneut ein Einfluss der Rekognitionsinformation, der aber mit der Anzahl negativer Zusatzinformationen monoton absinkt. In einem dritten Experiment wurden Städtevergleiche durchgeführt in Anlehnung an das Versuchsdesign von Goldstein und Gigerenzer (2002). Dabei wurde allerdings an Stelle der Zusatzinformation über das Vorhandensein einer Fußballmannschaft die Information über das Vorhandensein eines Flughafens gegeben. Im Kontrast zu den Ergebnissen von Goldstein und Gigerenzer (2002) finden sie einen höheren Anteil von Entscheidungen gegen die Vorhersage der Rekognitionsheuristik, was sie durch die höhere subjektiv

wahrgenommen Validität des Cues erklären und als Unterstützung für die These der kompensatorischen Nutzung von Rekognitionsinformation deuten.

Pachur und Hertwig (2006) liefern in einer experimentellen Studie allerdings Belege dafür, dass Inferenzen in Konsistenz mit der Rekognitionsheuristik schneller vollzogen werden als inkonsistente Inferenzen und unter Zeitdruck häufiger sind. Sie argumentieren, dass der Abruf von Rekognitionsinformation dem Abruf probabilistischer Cues zeitlich vorausgehe, so dass Rekognitionsinformation vielleicht nicht ein allein determinierender Einfluss, aber dennoch ein hervorgehobener Stellenwert zugesprochen werden kann.

4.5 Die Take-The-Best-Heuristik INTRO

Die *Take-The-Best*-Heuristik (TTB) ist ebenfalls für ordinale Paarvergleiche anwendbar und basiert auf der *Recognition*-Heuristik, bietet aber eine differenziertere Entscheidungsfindung für den Fall, dass beide Vergleichsobjekte bekannt sind (Gigerenzer & Goldstein, 1999). Für alle anderen Fälle ist das Ergebnis von TTB äquivalent zum Ergebnis der Recognition-Heuristik. Gigerenzer und Goldstein (1996b) rahmt die Heuristik in den Kontext von Herbert Simons Konzept des *Satisficing* ein:

> This fast and computationally simple algorithm is a model of bounded rationality rather than of classical rationality. There is a close parallel with Simon's concept of „satisficing“: The Take The Best algorithm stops search after the first discriminating cue is found, just as Simon's satisficing algorithm stops search after the first option that meets an aspiration level. (S. 654)

Der Algorithmus kann als ein nicht-kompensatorischer Algorithmus beschrieben werden, denn Entscheidungen können durch einzelne Cuewerte determiniert werden, die nicht mehr durch Ausprägungen anderer Prädiktoren kompensierbar sind. Der TTB-Algorithmus soll in seiner Grundform und einigen Varianten präsentiert werden.

4.5.1 Darstellung des Algorithmus

4.5.1.1 Der Standardalgorithmus

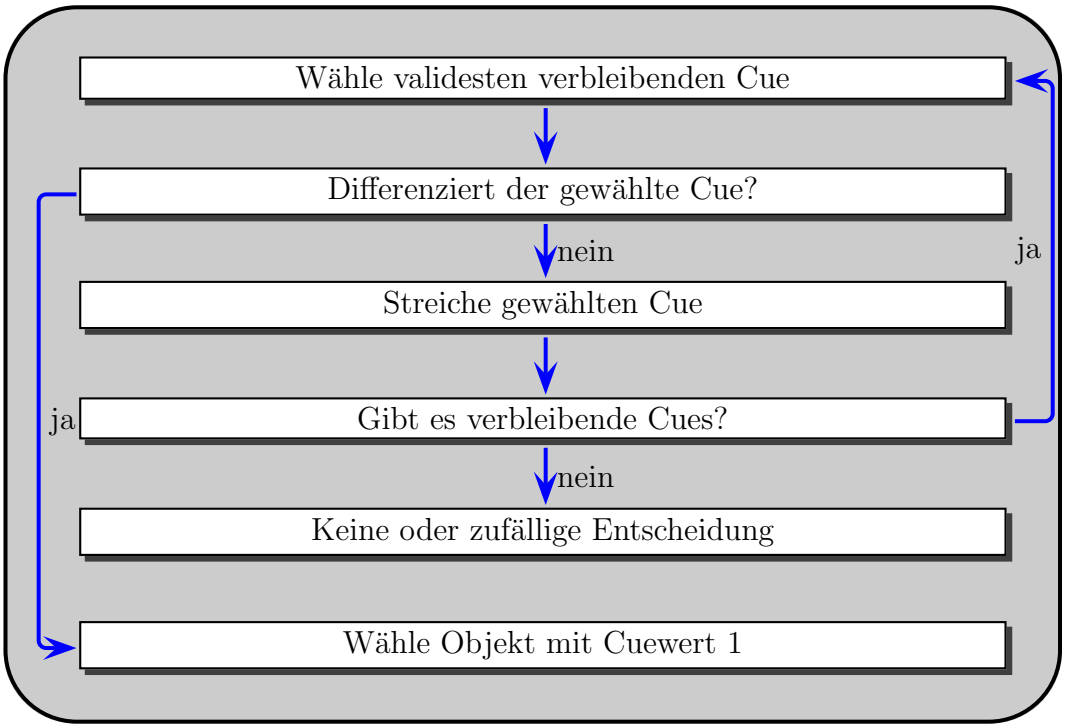


Abbildung 4.6: Schritte des TTB-Algorithmus

Ähnlich wie das oben dargestellte Verfahren verwendet die Heuristik eine Suchprozedur mit einfacher Abbruch- und einfacher Entscheidungsregel. Für den Fall, dass beide Objekte bekannt sind, werden neben

der Rekognitionsinformation weitere Kenntnisse über die Objekte abgerufen, die als „Cues“ bezeichnet werden. Im allgemeinen werden binäre - also dichotome oder dichotomisierte - Variablen als Cues betrachtet, die diagnostische Information bezüglich des Kriteriums enthalten. Dabei wird die *ökologische Validität* eines Cues bestimmt als die relative Häufigkeit von Objektpaaren in Bezug auf alle zulässigen Objektpaare, mit der das Objekt mit höherem Cuewert einen höheren Kriteriumswert aufweist als das Objekt mit niedrigerem oder fehlenden Cuewert. Die in diesem Sinne zulässigen Objektpaare bilden eine Teilmenge der möglichen Objektpaare: Es sind jene Paare, bei denen genau ein Objekt den Cuewert 1 aufweist. Die Entscheidung soll aufgrund eines Cues getroffen werden, für den die beiden Objekte unterschiedliche Werte aufweisen. Aus diesem Grund werden die Werte aller Cues so transformiert, dass der Wert 1 mit einer höheren Krtiteriumsausprägung und der Wert 0 mit einer geringeren Kriteriumsausprägung assoziiert ist, so dass die Entscheidung immer zugunsten des Objekts mit einem Cuewert von 1 getroffen werden kann.. Gigerenzer und Goldstein (1999) sprechen dabei von positiven bzw. negativen Cue-Werten.

Bezeichnet man mit $T_{i,ab}$ die Gesamtzahl der möglichen Paarvergleiche, bei denen das erste Objekt für Cue i den Cuewert a und das zweite Objekt den Cuewert b aufweist, mit $A_{i,ab}$ die Anzahl der Paarvergleiche, bei denen zusätzlich das erste Objekt den höheren Kriteriumswert aufweist, mit $B_{i,ab}$ die Anzahl der Paarvergleiche, bei denen zusätzlich das zweite Objekt den höheren Kriteriumswert zugeordnet hat und zuletzt mit $G_{i,ab}$ die Anzahl der Paarvergleiche bei denen beide Objekte den gleichen Kriteriumswert aufweisen, dann gilt, dass $T_{i,ab} = A_{i,ab} + B_{i,ab} + G_{i,ab}$ und die ökologische Validität des Cues i ist bestimmt als

$$v_i = \begin{cases} \frac{A_{i,10}+B_{i,01}}{T_{i,10}+T_{i,01}} & , T_{i,10} + T_{i,01} > 0 \\ 0 & , T_{i,10} + T_{i,01} = 0. \end{cases} \tag{4.25}$$

Konventionell werden die Cuewerte 1 und 0 so vergeben, dass v_i maximal wird, d.h., der Cue wird ggf. umgepolt. Gibt es keine Objektpaare mit identischem Kriteriumswert, so summieren sich die Validitäten vor und nach Umpolung zu 1, so dass nur eine der beiden berechnet werden muss. Dabei sind Cues mit einer Validität ungleich $0,5$ hilfreich für die Diskriminierung, da die Berücksichtigung der Cuewerte ansonsten keinen Vorteil gegenüber einer Zufallsauswahl bietet.

Da nicht alle bekannten oder konstruierbaren Cues berücksichtigt werden können, ist eine Regel erforderlich zur Bestimmung, in welcher Reihenfolge die Cues abgefragt werden sollen. Prinzipiell stehen hier mehrere Regeln zur Auswahl. In der TTB-Heuristik wird die Validität von Cues herangezogen: Die Cues werden in absteigender Reihenfolge ihrer ökologischen Validität sortiert oder genauer: Solange die Abbruchregel nicht erfüllt ist, wird jeweils der Cue untersucht, dessen Validität die höchste unter den verbleibenden Cues ist. Damit wird die Reihenfolge implizit erzeugt und auch nur so weit wie erforderlich. Da die Entscheidungsregel für ein gegebenes Objektpaar nur für Cues definiert ist, bei denen genau ein Objekt den Cuewert 1 hat, muss die Suche so lange fortgesetzt werden, bis ein solcher Cue gefunden worden ist oder kein Cue mehr zur Verfügung steht.

An dieser Stelle kann der Algorithmus der Heuristik vollständig spezifiziert werden (s. Abbildung 4.6). Die Heuristik zählt zur Klasse der lexikographischen Vergleichsstrategien. Für eine Entscheidung reicht ein einfacher lexikographischer Vergleich zwischen den Zeichenketten aus, die durch Kodierung der Objektprofile abhängig von der Cueausprägung mit einer '1' für positive Cuewerte und einer '0' für negative oder unbekannte Cuewerte in Reihenfolge absteigender Validität der Cues gebildet werden. Der Zeichen-String $'1000000'$ ist lexikografisch größer als der Zeichenstring $'0111111'$, das mit dem ersten String assoziierte Objekt wird gewählt, obwohl alle Cue-Werte nach dem ersten negativ sind (Martignon & Hoffrage, 2002), die Heuristik ist nichtkompensatorisch. Der Vorteil nicht-kompensatorischer Strategien im Vergleich zu kompensatorischen Strategien - wie linearen Modellen oder Bayes-Methoden - liegt damit auf der Hand: Nicht alle Informationen müssen aus der Umgebung extrahiert werden, die Methoden haben damit zunächste einen Geschwindigkeits- und Ressourcenvorteil.

In einer Baumdarstellung wird deutlich, dass die Rekognitionsheuristik einen modularen Bestandteil der TTB-Heuristik bildet (vgl. Abbildung 4.7). Die Kombinierbarkeit von simpleren Heuristik-Bausteinen zu komplexeren Heuristiken ist charakteristisch für das ABC-Programm.

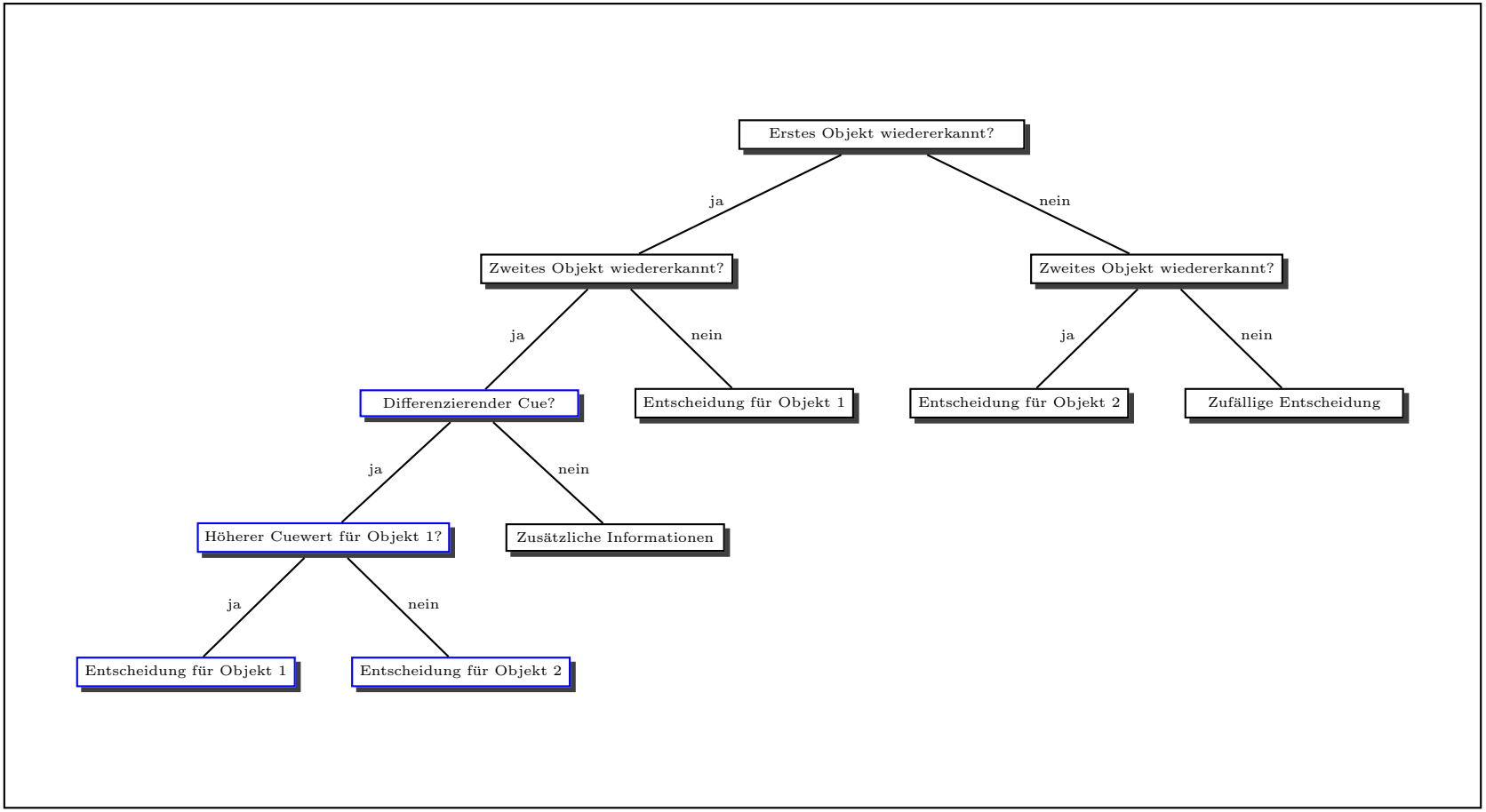


Abbildung 4.7: Darstellungsmöglichkeit des TTB-Algorithmus als Entscheidungsbaum

4.5.1.2 Alternative Bestimmung der Cuereihenfolge

Alternativ diskutieren Martignon und Hoffrage (2002) weitere Möglichkeiten zur Bestimmung der Cuereihenfolge: Eine Reskalierung der Validität liefert die Goodman-Kruskal Rangkorrelation, die bei Fehlen von Objektpaaren mit identischen Kriteriumswerten als

$$\gamma_i = \frac{A_{i,10} + B_{i,01} - (A_{i,01} + B_{i,10})}{T_{i,10} + T_{i,01}} \tag{4.26}$$

definiert ist und sich alternativ als $\gamma_i = 2v_i - 1$ bestimmen lässt. Ein Problem dieser Kriterien wird deutlich, wenn die Abhängigkeit der Gesamtzahl der Paarvergleiche, bei denen ein Cue zwischen zwei Objekten diskriminiert (was dem Nenner der Brüche entspricht) von der Verteilung der Cuewerte betrachtet wird: Haben N Objekte identische Cuewerte, so gibt es keine Diskriminierung bei den möglichen $\frac{N}{2}(N-1)$ Paarvergleichen, weicht lediglich ein Objekt ab, gibt es $N-1$ Paarvergleiche mit Diskriminierung bis hin zu einem Maximum bei gleicher Aufteilung der Cuewerte und $\frac{N^2}{4}$ Paarvergleichen mit Diskriminierung. Die Validitäten werden deshalb aufgrund der Berücksichtigung einer unterschiedlichen Anzahl an Paarvergleichen bestimmt. Eine Alternative bietet ein Erfolgskriterium, das die Anzahl der erfolgreichen Entscheidung bei alleiniger Heranziehung des Cues bestimmt und dabei bei Paarvergleichen ohne Diskriminierung $0,5$ addiert, was der Rateerfolgswahrscheinlichkeit entspricht. Gibt es keine Paare mit identischen Kriteriumswerten, so ergibt sich die Erfolgsrate s_i für den Cue i als

$$s_i = \frac{A_{i,10} + B_{i,01} + 0,5(T_{i,00} + T_{i,11})}{T_{i,00} + T_{i,01} + T_{i,10} + T_{i,11}}. \tag{4.27}$$

Eine weitere Alternative stellt Kendalls τ dar, das als

$$.\tau = \frac{A_{i,10} + B_{i,01} - A_{i,01} - B_{i,10}}{\sqrt{(T_{i,00} + T_{i,01} + T_{i,10} + T_{i,11})(T_{i,01} + T_{i,10})}} \tag{4.28}$$

angegeben werden kann. Lee und Cummins (2004) schlagen zudem die *Bayesianische Cuevalidität* vor, die als

$$\hat{v_i} = \frac{A_{i,10} + B_{i,01} + 1}{T_{i,10} + T_{i,01} + 2} \tag{4.29}$$

bestimmt wird. Erst bei einer höheren Entscheidungsanzahl nähern sich ökologische und Bayesianische Cuevalidität einander an, die Validitätsschätzung ist sensitiv für die Stichprobengröße.

4.5.1.3 Variationen von Take The Best: Take The Last und Minimalist

Durch Variationen der Bestandteile der TTB-Heuristik können verwandte Heuristiken konstruiert werden. Zwei Heuristiken dieser Art werden von Gigerenzer und Goldstein (1999) diskutiert, die sich im Verhältnis zu TTB durch einen niedrigeren Grad an Komplexität auszeichnen. Die *Minimalist-Heuristik* entspricht der TTB-Heuristik mit Ausnahme der Suchregel. Diese wird derart verändert, dass die Reihenfolge der Cues nicht durch die Validität festgelegt wird, sondern in jedem Schritt zufällig ein Cue ausgewählt wird. Aus diesem Grund wird keine Validitätsinformation bezüglich der Cues benötigt mit Ausnahme der Richtung des Zusammenhangs.

Eine Beschleunigung des Auswahlverfahrens soll die *Take The Last-Heuristik* (TTL) erzielen. In der folgenden Beschreibung soll eine Liste als Datenstruktur hinzugefügt werden. Diese Heuristik durchläuft den gleichen Algorithmus wie die Minimalist-Heuristik, bis für das erste Objektpaar eine Entscheidung getroffen werden konnte, bei der ein differenzierender Cue genutzt wurde. In diesem Fall wird eine Liste angelegt, genauer gesagt, ein Stapel mit einem „First In Last Out“-Prinzip. Wird ein differenzierender Cue gefunden, wird eine Referenz auf den Cue auf den Stapel gelegt. Ist mindestens ein Objekt im Stapel enthalten, wird zunächst der Stapel durchsucht, beginnend mit der obersten Referenz. Wird dabei ein differenzierender Cue gefunden, werden die Referenzen auf den Stapel zurückgelegt, wobei die relevante Referenz als letzte zurückgelegt wird. Wird bei der Suche kein differenzierender Cue gefunden, werden die verbleibenden Cues in einer zufälligen Reihenfolge abgefragt, bis entweder ein differenzierender Cue gefunden wurde, dessen Referenz dementsprechend auf den Stapel gelegt wird, oder bis alle Cues überprüft worden sind und eine zufällige Entscheidung zwischen den Objekten getroffen wird. Auf diese Weise wird

Tabelle 4.1: Durchschnittliche Ergebnisse der verglichenen Algorithmen über 20 Datensätze nach Czerlinski et al. (1999)

		Genauigkeit	
Algorithmus	Abgefragte Cues	Fitting	Generalisierung
Minimalist	2,2	0,69	0,65
TTB	2,4	0,75	0,71
DR	7,7	0,73	0,69
Multiple Regression	7,7	0,77	0,68

sichergestellt, dass der Reihe nach zunächst die Cues abgefragt werden, die bei vergangenen Vergleichen erfolgreich differenzieren konnten, um dann gegebenenfalls weitere Cues heranziehen zu müssen. Dabei bleibt jeder Cue, der mindestens einmal erfolgreich differenziert hat, in der privilegierten Vergleichsliste enthalten.

Eine Kombination des TTL-Prinzips mit den Eigenschaften der TTB-Heuristik führt zu der Spezifizierung einer hier neu vorgestellten Heuristik, der Bubble-Heuristik (BBH). Bei dieser Heuristik wird der Stapel mit den Cues in Reihenfolge ihrer Validität so gefüllt, dass der Cue mit der höchsten Validität oben liegt. Die Cues werden also in aufsteigender Validitätsreihenolge auf den Stapel gelegt. Nach dieser Initialisierung verläuft die Heuristik wie TTL. Der Name erklärt sich aus der schwachen Parallele zum Bubble-Sort-Algorithmus, bei dem Objekte, die in der Reihenfolge vor anderen liegen, im Sortiervorgang wie Blasen schrittweise ans Ende der Reihe aufsteigen. Die Bubble-Heuristik hat im Gegensatz zur TTL-Heuristik den Vorteil, dass Cues mit niedriger Validität nur in Fällen nach oben kommen sollten, in denen keine Differenzierung durch Cues mit höherer Validität gelingt. Als Nachteil muss man festhalten, dass die Initialisierung alle Informationen erfordert, die TTB zur Verfügung stehen, während diese Informationen nach der Anfangsinitialisierung nicht mehr gespeichert werden müssen und im Allgemeinen weniger Cues abgefragt werden sollten für einen Objektvergleich, da Cues mit niedriger Differenzierungsfähigkeit nach unten wandern. Komplexere Heuristiken würden ein differenziertes Gedächtnismodell benötigen, aufgrund dessen bspw. Positionswechsel nur bei eine bestimmten Zahl von Differenzierungen erfolgen würde.

4.5.2 Simulationsstudien zur TTB-Heuristik INTRO

4.5.2.1 Testvergleich in zwanzig Umgebungen

Eine umfangreiche Simulation zur Untersuchung der Leistungsfähigkeit von TTB im Vergleich zur Regel von Dawes (DR)[22] und zu multipler Regression zum einen und im Vergleich zur Minimalist-Heuristik zum anderen führen Czerlinski, Gigerenzer und Goldstein (1999) durch. Hierzu werden 20 Datensätze herangezogen, die von den Autoren als *Environments* bezeichnet werden. Alle genannten vier Algorithmen werden für jeden Datensatz auf die Leistungsfähigkeit für Paarvergleiche hinsichtlich des Kriteriums geprüft. In einer ersten Bedingung entspricht der Testdatensatz dem Trainingsdatensatz (*Data Fitting*), in einer zweiten Bedingung wurde auf Generalisierung überprüft, indem der Datensatz in 1000 Trials durch zufällige Halbierung in Test- und Trainingsdatensatz aufgespalten wurde[23].

Die Datensätze stammen aus den Gebieten der Psychologie (Attraktivitätseinschätzungen von Männern und Frauen), Soziologie (Obdachlosigkeit und Schulabbruch), Demographie (Sterblichkeit und Stadtbevölkerung), Ökonomie (Hauspreise, Landmiete und Professorengehälter), Verkehrsplanung (Autounfälle und Treibstoffverbrauch), Gesundheit (Fettleibigkeit und Körperfett), Biologie (Fruchtbarkeit von Fischen, Schlafgewohnheiten von Säugetieren und Kuhdung) sowie Geografie (Biodiversität, Niederschlag, Oxidant und Ozon). Die Fallzahlen der Datensätze schwanken zwischen 14 und 395 ohne fehlende Werte bei ebenfalls variierender Anzahl von Prädiktoren. Das im Datensatz gegebene Kriterium stellte dabei den

[22] Diese Regel entspricht einem Regressionsmodell mit einheitlichen Gewichten.

[23] Dabei wird der Split in Trainings- und Testset als Kreuzvalidierung bezeichnet, Freitas (2002) zum Beispiel bezeichnet die Trennungsprozedur als „hold out“-Methode und konstrastiert sie mit der Prozedur der Kreuzvalidierung, bei der für jeden Fall eine Vorhersage erzeugt wird, für deren Erstellung der Fall nicht beteiligt wird. Dementsprechend müssen eigentlich zwischen zwei und $n-1$ Modelle geschätzt werden, um eine einzige Kreuzvalidierung zu erreichen. Die Halbierung des Datensatzes und Berechnung von zwei Modellen ist dabei nur eine Variante der Kreuzvalidierung. In jedem Fall lässt sich der zweite Teil der Simulation aber als Überprüfung der Generalisierungsleistung betrachten.

Vergleichsmaßstab, an dem die Korrektheit der Wahlen überprüft wird. Die Ergebnisse der Simulation sind in Tabelle 4.1 zusammengefasst. Betrachtet man zunächst die Fitting-Ergebnisse, so wird deutlich, dass die multiple Regression hier im Durchschnitt einen Vorteil gegenüber den einfacheren Heuristiken verbuchen kann - die Größe dieses Vorteils schwankt zwischen den Datensätzen. Die *Minimalist*-Heuristik hingegen scheint sich die lobenswerte Einfachheit durch einen deutlichen Verlust in der Genauigkeit zu erkaufen, und auch die Regel von Dawes wird von TTB übertroffen, obwohl deutlich mehr Informationen verarbeitet werden. Bei der Analyse der Generalisierungsergebnisse fällt eine Veränderung in der Ergebnisreihenfolge der Algorithmen auf. Auch wenn die Minimalist-Heuristik weiterhin das schlechteste Ergebnis aufweist, können TTB und DR nun bessere Ergebnisse vorweisen als die multiple Regression, so dass TTB von Czerlinski et al. (1999) als Gewinner ausgerufen werden kann. Das Ergebnis ist gleichbedeutend mit einer größeren Differenz der Leistung von multipler Regression für Fitting und Generalisierung als für die simpleren Strategien, die sich als robuster erwiesen haben. Als Erklärung wird die mögliche Überanpassung an Daten angeführt, ein *Overfitting*, das durch die Anfälligkeit der Regression für zufällige Korrelationen und somit für die Anpassung an stichprobenspezifische Fehlervarianz hervorgerufen werden kann. Ein ähnlicher Effekt wird auch für DR vermutet.

Die Diskrepanz zwischen Fitting und Generalisierung sollte hinsichtlich politischer Implikationen beachtet werden. Am Beispiel der Ergebnisse für Schulabbrecher zeigen die Autoren, dass eine Konzentration auf die Ergebnisse der Regression (mit höherer Fitting-Genauigkeit in der Stichprobe) zu der Empfehlung führen würde, Englisch als zweite Sprache und die Integration von Minoritäten zu unterstützen, während eine Konzentration auf die Ergebnisse von TTB (mit höherer Generalisierungsgenauigkeit) die Empfehlung beinhalten würde, die Anwesenheit von Studenten im Unterricht zu steigern und eine gründlichere Vermittlung von Basisinhalten zu erreichen[24].

In einer Literaturanalyse zum Vergleich zwischen DR und Regression kommen die Autoren zu dem Schluss, dass eine Verwendung der Regression nur in jenen Fällen einen Vorteil gegenüber DR bietet, wenn eine hohe lineare Vorhersagbarkeit des Kriteriums gegeben ist, viele Objekte pro Cue gegeben sind und die Cues nicht hoch korreliert sind und ansonsten DR eine höhere Genauigkeit aufweist. Bei einem Vergleich von TTB und Regression über die 20 Datensätze stellt sich heraus, dass der Vorteil von TTB bei weniger Objekten pro Cue ausgeprägter ist, was auf die relative Instabilität der Parameterschätzungen zurückgeführt werden kann und durch die weitere Beobachtung gestützt wird, dass bei einem kleineren Verhältnis von Objekten zu Cues die Differenz der Leistung zwischen Fitting und Generalisierung für die multiple Regression höher ausfällt. Bezüglich der linearen Vorhersagbarkeit, gemessen durch das multiple R^2, oder der durchschnittlichen Interkorrelation der Cues, wurde hinsichtlich der Umgebungen kein Effekt auf die Leistungsdifferenz gefunden.

Als mögliche Erklärung kann hingegen ausgeschlossen werden, dass hochvalide Rekognitionsinformation für den Vorteil von TTB verantwortlich ist, da alle Objekte als bekannt behandelt werden. Gravierender ist vielleicht der Einwand, dass eine künstliche Dichotomisierung der Cues das Potenzial der Regression nicht ausschöpft. Bei einer Wiederholung der Simulation mit kontinuierlichen Cues erreichen sowohl TTB als auch die multiple Regression eine Genauigkeit von $0,76$, wobei aufgrund der höheren Wahrscheinlichkeit einer Differenz im Cuewert zwischen Objekten durchschnittlich eine geringere Anzahl von Cues von TTB benötigt wurde. Allerdings wird beim Vergleich der Verfahren ebenfalls deutlich, dass sich TTB im Gegensatz zur multiplen Regression nicht zu einer quantitativen Vorhersage des Kriteriums nutzen lässt, die sich nicht aus den Objektpaarvergleichen rekonstruieren lässt, was als Spezialisierung von TTB für bestimmte Aufgabenklassen bewertet wird.

Die Leistung von DR konnte gesteigert werden, indem Cues mit einer Validität kleiner als $0,70$ aus der Berechnung ausgeschlossen wurden, damit stieg die Genauigkeit von DR auf $0,71$, also auf das gleiche Niveau wie TTB. Eine Reduzierung der Datenbasis, auf deren Grundlage TTB die Reihenfolge der Cues ermittelt, auf zehn zufällig ausgewählte Objekte führt zu einer Verschlechterung der Genauigkeit auf $0,66$, also fast auf das Niveau der Minimalist-Heuristik. Czerlinski et al. (1999) kommentieren dieses Ergebnis allerdings als:

> It seemed that even with very few observations, Take The Best could still make reasonably accurate predictions. (S. 113)

[24] Allerdings sollte generell zwischen der deskriptiven Modellierung von Zusammenhängen und der Ableitung von Interventionsmöglichkeiten unterschieden werden, da allein schon im durch die Variablen der Regressionsgleichung aufgespannten Raum Randbedingungen und Interdependenzen gelten, die nicht im Modell enthalten sind.

Tabelle 4.2: Ergebnisse verschiedener Varianten von TTB nach Martignon und Hoffrage (2002) für die deutsche Stadtumgebung: Fittinggenauigkeit, Einschätzung des komputationalen Aufwandes und Reihenfolge der Generalisierungsergebnisse

Methode	Genauigkeit (Fitting)	Aufwand	Generalisisierung (Reihenfolge)
optimale Reihenfolge	0,758	++	2.
konditionale Validitität	0,756	+	3.
ökologische Validität	0,742	o	1.
Kendalls τ	0,735	o	4.
Erfolgsquote	0,727	o	5.
durchschn. Reihenfolge	0,700	-	6.
ungünstigste Reihenfolge	$\approx 0,620$	++	-

Negativ könnte formuliert werden: Der Vorteil der einfachen Heuristik gegenüber der multiplen Regression geht durch das bloße Ausschöpfen gegebener Informationen in den Datensätzen verloren. Der Vorteil gegenüber DR geht bei einer Veränderung von DR nahezu verloren, die nicht mehr Informationen nutzt, als TTB zur Verfügung stehen. Die Abhängigkeit von TTB von einer ausreichenden Datenbasis für die Erstellung der Cuereihenfolge muss auch als Einschränkung des Anwendungsbereiches betrachtet werden. Zudem können die gewählten 20 Datensätze nicht als repräsentative Auswahl für alle möglichen Anwendungsfälle für die Algorithmen gewertet werden, so dass keine Entscheidung zwischen den Algorithmen getroffen ist. Trotzdem demonstriert die Studie deutlich, dass TTB als nichtkompensatorisches Verfahren, das eine kurze Berechnungszeit aufweist und relativ wenige Informationen nutzt, also als *Fast and Frugal Heuristic* zählen kann, einem komplexeren Algorithmus mit vollständigen Informationen nicht unterlegen sein muss. Oder in den Worten der Autoren:

> Our results indicate ...that mental strategies need not be like multiple regression to make accurate inferences about their environments. (S. 118)

Simulationsstudien wie die beschriebene können auch mit eigenen Datensätzen unter Verwendung der Software GIER$_{[1.0]}$.simple_heuristics repliziert werden, die eine Vielzahl von Algorithmen für eine selbst zusammengestellte Konkurrenz zur Verfügung stellt, dabei stehen einfache und komplexere Bedienoberflächen zur Verfügung [25]

4.5.2.2 Kriterien zur Bestimmung der Cuereihenfolge für TTB

Martignon und Hoffrage (2002) untersuchen die Leistungsfähigkeit verschiedener Methoden der Reihenfolgebestimmung. Als Vergleichsrahmen werden zunächst die Ergebnisse aller möglichen ($9! = 362880$) Reihenfolgen in ihrer Verteilung betrachtet. Da die Cues am Kriterium ausgerichtet werden, gibt es in der klassischen Städteaufgabe keine Reihenfolge mit einer Genauigkeit unter $0,62$. Die optimale Reihenfolge erbringt eine Genauigkeit von $0,758$. Die Leistungen der verschiedenen Methoden zur Bestimmung der Cuereihenfolge sind in Tabelle 4.2 angegeben.

Während die vorgeschlagenen Kriterien zur Bestimmung der Cuereihenfolge (ökologische Validität v, Erfolgsrate s und Kendalls τ) alle zu besseren Ergebnissen führen als die zufällige Sortierung, die durch die durchschnittliche Genauigkeit aller möglichen Reihenfolgen geschätzt werden kann, gibt es Unterschiede zwischen den Kriterien. Lediglich die konditionale Validität kommt an das Ergebnis der optimalen Reihenfolge heran, die ökologische Validität, die von TTB verwendet wird, übertrifft aber noch die beiden anderen Kriterien. Insgesamt ist das Ergebnis nur für 1,8% der Reihenfolgen besser als das für die ökologische Validität. Dabei wurde die Robustheit der Ergebnise für TTB von den Autoren in zweierlei Hinsicht überprüft. Zum einen ergibt sich eine relative Stabilität der Ergebnisse bei Veränderung der Reihenfolge durch Verschieben von Elementen. Bei Verschiebungen im Ausmaß von fünf Schritten sinkt die Genauigkeit durchschnittlich nur um 0,6%. Um das Niveau der *Minimalist*-Strategie zu erreichen, sind 18 Schritte notwendig (die Verschiebung eines Cues um zwei Positionen zählt dabei als zwei Schritte).

Zum anderen werden die Strategien hinsichtlich ihrer Leistung in einer Generalisierungsaufgabe verglichen. Hierzu wurde die Städtestichprobe in 100 Trials jeweils zufällig in zwei Hälften unterteilt, wobei die

[25] Eine Variante der Software wurde dabei auf der *Summer School* des Max-Planck-Instituts für Bildungsforschung zum Thema *Bounded Rationality* im Sommer 2006 als Werkzeug zur Exploration der ökologischen Rationalität von Heuristiken eingesetzt.

Reihenfolge nach den genannten Kriterien an der ersten Hälfte bestimmt wurde, so dass die Leistung der resultierenden Modelle in der ersten Hälfte als Fitting und in der verbleibenden Hälfte als Generalisierung gemessen werden kann. Dabei wird im Fitting die für die Gesamtstichprobe ermittelte Reihenfolge der Leistung repliziert, in der Generalisierungsaufgabe verschiebt sich die über die Trials gemittelte Genauigkeit aber wie in Tabelle 4.2 angegeben. Obwohl *Minimalist* in der Studie nicht zum Einsatz kam, wird hier aus Plausibilitätsgründen eine Leistung erwartet, die der durchschnittlichen Leistung in der Gesamtstichprobe entspricht, so dass der durchschnittlichen Validität der letzte Platz zugewiesen wird, weil die Generalisierungsgenauigeit der anderen Algorithmen zwischen ca. 71% und ca. 73% liegt. Es fällt auf, dass das Ergebnis von TTB in der Generalisierungsaufgabe besser ist als das der lexikographischen Strategien mit konditionaler und optimaler Validität.

Die „optimale“ Reihenfolge ist also nur für die Fitting-Aufgabe optimal. Berücksichtigt man zusätzlich den differenziellen Aufwand für die einzelnen Strategien, so setzt sich bezüglich des Städtedatensatzes die ökologische Validität gegen die „optimale“ Reihenfolge und die konditionale Validität durch, wenn nicht die Anpassungsleistung an bekannte Daten, sondern die Generalisierbarkeit auf unvertraute Daten von Interesse ist.

4.5.3 Empirische Studien zu Take The Best

Rieskamp und Hoffrage (1999) beginnen mit der Intuition, dass die in Simulationen beobachtete Genauigkeit der Fast and Frugal-Heuristiken zusammen mit ihrem geringen Bedarf an kognitiven Ressourcen ihren tatsächlichen Einsatz plausibel machen sollte. Unter der Annahme des Vorhandenseins einer adaptiven Werkzeugkiste mit Strategien, stellen die schnellen und frugalen eine Teilmenge dar, auf die insbesondere zurückgegriffen werden sollte, wenn Wissen und Zeit begrenzt sind. Als Beispiel für die Verwendung einer TTB-Form im Tierreich nennen Gigerenzer und Goldstein (1999) das Phänomen der Partnerwahl bei Guppy-Fischen. Weibliches Guppy-Fische reagieren auf körperliche Cues (grelle Körperfarbe in Orange) und soziale Cues (Männchen wird bei Paarung mit anderen Weibchen beobachtet), wobei der soziale Cue vom Farbcue dominiert wird. Allerdings gibt es einen gewissen Spielraum, und ein geringer Unterschied in der Farbe kann bei differenzierendem sozialen Cue kompensiert werden, bei einem größeren Farbunterschied (ca. 40%) wird die Suche abgebrochen. An diesem Beispiel lässt sich die Implementierung einer einfachen Abbruchregel demonstrieren, nicht aber die Implementierung einer nicht-kompensatorischen Strategie. Diese Aufgabe bei Menschen zu leisten, stellt sich als anspruchsvoller heraus: Rieskamp und Hoffrage (1999) nennen zwei wesentliche Ansätze, mit denen Strategien identifiziert werden können, die von Entscheidern angewandt werden. Die erste Gruppe von Verfahren ist prozessorientiert (vgl. Payne, Bettman & Johnson, 1988): Der zur Entscheidung führende Prozess wird beobachtet, die Abfrage, die Reihenfolge der Abfrage von Informationen und die Beschäftigungsdauer mit Informationen wird protokolliert, zuweilen werden auch lautes Denken, Protokollanalysen oder retrospektive Strategieerklärungen eingesetzt. Die zweite Gruppe von Ansätzen versucht eine Rekonstruktion von Strategien über die Entscheidungsergebnisse. Die tatsächlichen Entscheidungen werden verglichen mit Entscheidungen, die aus einer idealen Anwendung verschiedener Strategien resultieren würden, der Grad der Übereinstimmung oder auch die Art der Abweichungen wird als Unterstützung für die Annahme der Anwendung einer Strategie betrachtet.

Keiner der beiden Ansätze ist ohne inhärente Probleme. Informationsabfrage und Informationsberücksichtigung können divergieren, retrospektive Analysen können aufgrund von Ergebnissen verzerrt sein, parallele Analysen können den Prozess verändern oder aber nur zugängliche, aber nicht entscheidende Prozesskomponenten aufdecken. Im ergebnisorientierten Ansatz stellt sich zum einen das Problem, dass die Ergebnisse verschiedener Verfahren bei vielen Aufgaben identisch sind und auch nicht berücksichtigte Strategien mit den Ergebnissen einer Person kongruent sein können. Idealerweise bietet eine Kombination beider Ansätze bei zusätzlicher theoriegeleiteter Manipulation der Experimentalbedingungen den größtmöglichen Erkenntniswert.

Bröder (2002) entwickelt eine Methode zur ergebnisorientierten Bestimmung der Strategieverwendung. Er diskutiert neben dem Diskriminierungsproblem und der Notwendigkeit der Betrachtung von Daten auf der Individualebene auch das Problem der fehlenden Spezifizierung der Genauigkeit und Fehleranfälligkeit von TTB, weshalb er ein explizites Fehlermodell für eine unverzichtbare Komponente der Methode hält. Die Methode basiert auf einer vollständigen Vorgabe aller Kombinationen von Cuewerten und Vergleichen zwischen uneingeschränkten und eingeschränkten linearen Regressionsmodellen. Bei

fehlerfreier Anwendung von TTB kann das Entscheidungsverhalten durch ein Regressionsmodell mit nicht-kompensatorischen Gewichten beschrieben werden. Für k nicht-kompensatorische Gewichte mit $w_{i+1} > w_i$ gilt[26]:

$$\sum_{j=i+1}^{k} w_j < w_i. \tag{4.30}$$

Eine Anwendung von *Dawes's Rule* kann hingegen durch ein Modell mit Gleichgewichtung beschrieben werden. Also kann die Modellverschlechterung bei Verwendung jeweils einer der beiden Einschränkungen auf Signifikanz geprüft werden. Eine technische Diskussion der erforderlichen Schritte liefert Bröder (2005).

4.5.3.1 Kompensatorische vs. nicht-kompensatorische Strategien

In einer Beispielstudie zur Untersuchung der Frage, ob Personen eher kompensatorische oder nicht-kompensatorische Entscheidungsstrategien einsetzen, untersuchen Rieskamp und Hoffrage (1999) die Entscheidungsaufgabe, aus vier Firmen diejenige mit dem höchsten jährlichen Profit auszuwählen. Dazu wurden vier Cues mit unterschiedlicher Validität (zwischen 0,37 und 0,78) potenziell zur Verfügung gestellt. Die Cuevaliditäten wurden den Teilnehmern als Konzept erklärt und angegeben. Nach einer Trainingsphase mit Feedback (7 Trials) zur Vertrautmachung mit dem Aufgabenformat mussten die Teilnehmer 15 Wahlen unter niedrigem (50s) und 15 Wahlen unter hohem Zeitdruck (20s) treffen. Um Cuewerte abzufragen, mussten Boxen durch Mausklick geöffnet werden, wobei zur Abfrage eines neuen Wertes die letzte Box durch einen erneuten Klick geschlossen werden musste, so dass immer nur eine Information zugleich abgefragt werden konnte. Korrekte Antworten wurden mit einem zusätzlichen Geldbetrag vergütet.

Die Anzahl abgefragter Informationen liegt durchschnittlich bei 33 unter geringem Zeitdruck und bei 19 unter hohem Zeitdruck. Da nur 24 Cuewerte zur Verfügung standen, wurden Boxen unter geringem Zeitdruck im Schnitt mehr als einmal geöffnet. Gleichzeitig standen bei hohem Zeitdruck nicht für alle Teilnehmer alle Informationen zur Verfügung. Die relative Zeit, die auf die Information bezüglich des validesten Cues verwandt wird, steigt bei hohem Zeitdruck, die auf die Information des am wenigsten validen Cues verwandte relative Zeit sinkt. Die Korrelation zwischen Zeit pro Cue und Cuevalidität steigt von $r = 0,69$ auf $r = 0,82$ unter Zeitdruck, so dass die Daten auf einen größeren Effekt der Cuevalidität auf die Informationsabfrage hinweisen und auf einen kleineren des Zeitdrucks. Möglicherweise ändern die Teilnehmer unter großem Zeitdruck ihre Strategie zugunsten eines nicht-kompensatorischen Vorgehens. Da Muster der Informationsabfrage weist eher auf eine nach Cues geordnete Abfrage als auf eine nach Alternativen geordnete Abfrage hin, was eher für eine nicht-kompensatorische Strategie spricht. Für eine Binnendifferenzierung zwischen verschiedenen möglichen nicht-kompensatorischen Strategien reichen die Informationen allerdings nicht aus.

Fehlende Separabilität ist auch ein Thema für die ergebnisorientierte Analyse: ungefähr 92% der Paarvergleiche führten bei Anwendung zweier Strategien zum gleichen Ergebnis. Somit entsteht ein Konflikt zwischen Repräsentativität und Separierbarkeit: Werden Items mit dem Ziel der Steigerung der Separierbarkeit ausgewählt, sinkt die Generalisierbarkeit der Ergebnisse, bei steigender Repräsentativität sinkt die Separierbarkeit. Deshalb wurde 50 Teilnehmern ein ausgewählter Itemsatz gegeben mit geringerer Überlappung, 25 Teilnehmern ein repräsentativer Datensatz. Rieskamp und Hoffrage (1999) betrachten die Leistung von acht Strategien, darunter *Dawes's Rule*, *Franklin's Rule*, drei Varianten lexikographischer Strategien und *Elimination by Aspects* (Tversky, 1972). Die Leistung aller Strategien ist im repräsentativen Set besser, die Varianz der Leistungen ist in diesem Set ebenfalls geringer, die Leistungen der lexkographischen Strategien sind mit denen von Franklin's Rule vergleichbar. Bei der Aufteilung der Teilnehmer mit dem ausgewählten Set auf die Strategien, abhängig von der besten Passung, zeigt sich ein Einfluss der Zeitdruckbedingung. Bei geringem Zeitdruck werden mehr Personen als Benutzer kompensatorischer

[26] Folgende Visualisierung kann die Eigenschaften nicht-kompensatorischer Gewichte veranschaulichen. Angenommen, es sind Bügel jeweils in Länge der Gewichte vorhanden. Hängt man den größten Bügel an eine Stange und die Bügel in der Reihenfolge ihrer Größe jeweils an ein freies Ende des zuletzt angehängten Bügels, dann kann bei nicht-kompensatorischen Gewichten ein Bügel niemals die Linie zwischen Boden und Mittelpunkt eines beliebigen zuvor aufgehängten Bügels erreichen, egal wie die Bügel verteilt werden. Diese Möglichkeit der Darstellung erwächst aus der Eigenschaft, dass Nicht-Kompensierbarkeit auch bei einer linearen Skalierung von Gewichten erhalten bleibt.

Strategien identifiziert, bei hohem Zeitdruck mehr Personen als Benutzer nicht-kompensatorischer Strategien.

Insgesamt kann so gezeigt werden, dass eine Generalisierung der TTB-Strategie das Entscheidungsverhalten einer Teilmenge von Versuchspersonen am besten beschreiben kann, deren relativer Anteil bei Zeitdruck steigt. Als mögliche weitere Einflussvariablen werden Aufgabentyp (Auswahl vs. Einschätzung), Aufgabenkomplexität, sozialer Kontext und Verantwortung diskutiert.

4.5.3.2 Die empirischen Studien von Bröder

Bröder (2000) sieht einen Spalt zwischen Theorie und Daten und untersucht in vier Experimenten vor allem den Grad an Universalität, der der Anwendung von TTB zugeschrieben werden kann. In einer Lernphase werden 50 Versuchspersonen zunächst mit fünf verschiedenen Cues und ihrer Beziehung zu einer Zielvariablen vertraut gemacht. Die Aufgabe wird eingerahmt als extraterrestrisches Forschungsprojekt, als Kriterium wird der Fortschritt einer außerirdischen Kultur gewählt, die fünf Cues stellen Kulturtechniken mit fiktiven Namen dar und haben Validitäten zwischen 0,60 und 0,92. Die Probanden erhalten Informationen über jeweils einen Cuewert von zwei Objekten, wobei ein Objekt einen positiven, ein anderes einen negativen Wert aufweist. Die Versuchspersonen entscheiden und erhalten direktes Feedback sowie ein Histogramm über den summierten Erfolg und Misserfolg bezüglich des Cues. Dabei werden insgesamt 250 Lerntrials durchgeführt, im Anschluss 24 kritische Trials, bei denen für jeweils zwei Objekte fünf konfligierende Cuewerte vorgegeben werden. Die Verteilung wird so gewählt, dass bei Anwendung der TTB-Heuristik das Objekt mit weniger positiven Cuewerten gewählt würde. Für richtige Wahlen wird ein Geldbetrag ausbezahlt. Im Anschluss sollen die Versuchspersonen die Validität der Cues einschätzen. Bröder findet in den Einschätzungen der Cuevaliditäten durch die Teilnehmer Belege für den Erfolg der Lernphase, die Cuereihenfolge und die ungefähre Höhe der Validitäten können reproduziert werden. Zeitdruck und die Richtung irrelevanter (da geteilter) Cuewerte haben keinen nennenswerten Effekt auf die Ergebnisse, allerdings wird die durch die TTB-Heuristik indizierte Alternative seltener gewählt, wenn die Anzahl der sie unterstützenden Cues geringer ist und zum anderen, wenn ihre Validität geringer ist (und dementsprechend mehr bzw. validere Cues die nicht indizierte Alternative unterstützen). Bröder (2000) bewertet die Ergebnisse als Beleg gegen die starke Annahme, dass alle Versuchspersonen wenn möglich, immer die Take The Best-Heuristik anwenden.

In einem zweiten Experiment versucht Bröder die Anzahl der TTB-Anwender zu identifizieren durch Verwendung der oben skizzierten Klassifikationsmethode. Den Versuchspersonen werden in zwei Gruppen im gleichen Setting wie im Experiment oben in 200 Lerntrials die Cuevaliditäten von vier Cues vermittelt, wobei in einer Gruppe die Cuevaliditäten gering streuen (zwischen 0,66 und 0,90), in der anderen Gruppe die Validitäten stark streuen (zwischen 0,54 und 0,96). Anschließend werden den Probanden alle 120 möglichen Aufgaben aus vier Cues ($(2^4 \cdot (2^4 - 1) = 120)$ präsentiert. Bei Anwendung der regressionsbasierten Klassifikationsmethode konnte für 11 von 40 Probanden die TTB-Hypothese nicht abgelehnt werden, und für alle Probanden konnte die Hypothese einer einheitlichen Gewichtung verworfen werden. Als Tendenz findet Bröder eine größere Anzahl von klassifizierten TTB-Anwendern in der Bedingung mit höherer Streuung der Validitätswerte.

Um den Effekt einer sukzessiven versus einer simultanen Präsentation von Cues zu untersuchten, führt Bröder (2000) ein drittes Experiment durch, in dem Probanden Cuewerte der Reihe nach abfragen müssen. Bei der Entscheidung für den Kauf einer Aktie werden vier binäre Cues mit Validitäten zwischen 0,69 und 0,80 zur Verfügung gestellt, die in einer Bedingung für einen günstigen, in einer anderen Bedingung für einen hohen Preis in der Reihenfolge ihrer Validität gekauft werden können. Die Preise werden so gewählt, dass in der Niedrigpreis-Bedingung eine kompensatorische Nutzung aller Cues gegenüber der TTB-Anwendung einen finanziellen Vorteil verspricht, in der Negativbedingung einen Nachteil. In den Ergebnissen zeigt sich ein Effekt dieser Manipulation: In der Niedrigpreisbedingung werden 8 von 20 Personen als TTB-Benutzer klassifiziert, in der Hochpreisbedingung sind es 13 von 20. Die Ergebnisse zeigen erneut ein mit den Vorhersagen der TTB-Heuristik konformes Verhalten durch nicht alle, aber einen substanziellen Anteil der Versuchspersonen. Dabei wird die Heuristik von einigen Teilnehmern auch in der Bedingung angewandt, in der eine alternative Strategie zu besseren Ergebnissen führen würde:

> This suggests that strategy selection is not perfectly adaptive but is also determined by idiosyncratic preferences for strategies. (Bröder, 2000, S. 1342).

Um ein weitere Aufklärung der Faktoren zu ermöglichen, die ein verstärktes Auftreten von TTB-kompatiblen Wahlen bedingen, führt Bröder ein viertes Experiment durch, in dem drei Faktoren in vier Bedingungen variiert wurden. In einer Bedingung werden die Cues gleichzeitig präsentiert, es wird kein Feedback über die Gewinnwahrscheinlichkeit der gewählten Option gegeben und die Cues müssen logischerweise nicht gekauft werden. Die zweite Bedingung entspricht der ersten bis auf die zusätzliche Einführung von Feedbackinformationen. In der dritten Bedingung werden zudem Cuewerte sukzessive durch Anklicken präsentiert, ohne dass dafür ein Betrag zu investieren ist, während die Informationen in der vierten Bedingung zusätzlich Geld kosten. In den ersten drei Bedingungen wird ein kleiner Teil der Versuchspersonen als TTB-Anwender klassifiziert (15%), ein kleinerer Anteil als Benutzer einer Gleichgewichtungsstrategie. In der vierten Bedingung, also der Investitionsbedingung, steigt dieser Anteil drastisch an (auf 65%). Demenstprechend scheinen die Informationskosten den entscheidenden Faktor für den Wechsel einer Strategie bei Entscheidern zu sein, die zusätzliche Mühe des Anklickens der Cueinformationen hat dabei scheinbar keinen Effekt.

Insgesamt zeigen die Ergebnisse der Experimente eine Grundlage für die Ablehnung der starken TTB-Hypothese aber Unterstützung für die Annahme interindividueller Unterschiede und Kontextabhängigkeit. Dementsprechend scheint das Aggregieren individueller Daten für Studien der beschriebenen Art kein gangbarer Weg zu sein. Zudem können die Kosten der Informationssuche als entscheidender Anreiz für die Verwendung der TTB-Heuristik betrachtet werden. Auch Newell und Shanks (2003) finden in einem Experiment keinen Widerspruch der Ergebnisse mit der Position der ABC-Gruppe, merken aber an, dass ihr Versuch, eine Situation mit universeller Anwendung von TTB künstlich zu erzeugen, gescheitert ist.

(Bröder, 2003) geht experimentell der Frage nach, inwiefern ein Wechsel der Struktur der Umgebung zu einer adaptiven Anpassung der Strategiewahl führt. Dabei macht er die Beobachtung, dass der Anteil der TTB-Anwender in einer Umgebung mit nicht-kompensatorischer Auszahlungsstruktur höher ist als in einer Umgebung mit kompensatorischer Struktur. Dabei kann er keinen Effekt der Gedächtnisspanne oder Belastung auf die Strategiewahl nachweisen, allerdings einen schwachen, aber robusten positiven Einfluss von Intelligenz auf die Adaptivität der Strategiewahl.

4.5.3.3 Weitere Studien

Newell, Weston und Shanks (2003) untersuchen das Entscheidungsverhalten von Studienteilnehmern in einer Forced-Choice-Aufgabe bei der Wahl zwischen Aktien mit unterschiedlich vielen Cues. Trotz einer Erhöhung der Cuezahl auf sechs Cues gegenüber vier finden sie eine konsistente Anwendung der TTB-Strategie bei nur 8 von 24 Teilnehmern, von denen ein substanzieller Anteil nichtfrugale Strategien anzuwenden scheint. Sie zweifeln dabei die Nützlichkeit der Strategie nicht an, stellen aber die unkritische Übernahme der Idee in Frage, TTB stünde empirischen Phänomenen näher als andere Strategien, da TTB in den Augen der Autoren zu restriktiv sei, um die Variabilität beobachteten Verhaltens abdecken zu können. Sie argumentieren gegen eine Vermeidung von freien Parametern von Heuristiken durch Vermehrung von Heuristiken und damit Schaffung von freien Parametern in der Heuristikwahl.

Newell, Rakow, Weston und Shanks (2004) untersuchen das Kriterium, das für die Reihenfolgebildung der Cues herangezogen wird, mit einer Forced-Choice-Aufgabe, bei der eine von zwei Aktien aufgrund von Cues mit unterschiedlicher Validität und Diskriminationsrate ausgewählt werden muss. Dabei finden sie, dass neben der Validität auch die Diskriminationsrate berücksichtigt wird, und dass beide Werte sensitive Erfolgskriterium das Verhalten der Versuchspersonen besser beschreibt. Newell (2005) kritisiert die TTB-Heuristik in dem Sinne, dass sie nur in seriellen Architekturen und nicht in parallelen Architekturen einen Suchvorteil aufweist. Ähnlich wie Bröder kritisiert er ebenfalls das Fehlen eines expliziten Fehlermodells:

> ... although the models are clearly specified, there is no indication of the degree of empirical deviation permissible from the deterministic search, stopping and decision rules. Any deviations are merely ignored as noise, allowing proponents to claim 64% of choices consistent with TTB to be an 'excellent result' whereas others might regard the failure to fit 36% as a considerable problem [...]. (S. 12)

Die empirische Bewährung der Heuristik sieht er dementsprechend als fragwürdig, was ihn über das Forschungsprogramm der Adaptive Toolbox urteilen lässt:

> [...] with largely equivocal evidence for the only heuristic to undergo detailed empirical scrutiny, claims for a 'toolbox' of ecologically rational heuristics that are constructed and triggered into action by the properties of particular environments have been weakened. (S. 14)

Lee und Cummins (2004) versuchen eine Vereinigung von TTB mit einem Modell, das sie als *Rational Model* bezeichnen, welches auch den Namen Naïve Bayes trägt und eine Vereinfachung komplexerer Bayesianischer Modelle darstellt (vgl. Martignon & Schmitt, 1999; Martignon, 2001; Katsikopolous & Martignon, 2006), indem Abhängigkeiten der Cues untereinander ignoriert werden. Damit ergeben sich für zwei Objekte A und B und den Cuewerten $\{a_1, a_2, \ldots, a_k\}$ bzw. $\{b_1, b_2, \ldots, b_k\}$ die Log-Odds als

$$L_{AB} = \sum_{i=1}^{k} ln \frac{p(a_i, b_i | A > B)}{p(a_i, b_i | A < B)}, \tag{4.31}$$

was sich bei der Aufteilung in die Menge der Cues F_A, in der Cues zugunsten von Objekt A diskriminieren, die Menge F_B, in der Cues zugunsten von Objekt B diskriminieren und bei Ignorieren der Menge F_U, in der Cues nicht diskriminieren, umformen lässt zu:

$$L_{AB} = \sum_{i \in F_A} ln \frac{\hat{v}_i}{1 - \hat{v}_i} - \sum_{i \in F_B} ln \frac{\hat{v}_i}{1 - \hat{v}_i}. \tag{4.32}$$

Bei positivem L_{AB} wird eine Entscheidung für Objekt A gefällt, bei negativem L_{AB} eine Entscheidung für Objekt B. Das Experiment von Lee und Cummins (2004) mit 40 Versuchspersonen nutzt einen natürlichen Datensatz, dessen Werte für die Erstellung einer künstlichen Aufgabe genutzt werden: Jeder Stimulus wird als Gas in einem Kanister repräsentiert, die Cues entsprachen animierten Gasmolekülen verschiedener Farbe, und die Entscheidungsvariable ist die Giftigkeit des Gases. In einer Trainingsphase wurden 16 Stimuli mit 6 Cues in insgesamt 119 Paaren vorgegeben, bei denen sowohl TTB als auch das Rational Model die gleichen Vorhersagen machten, die in 86% der Vergleiche korrekt waren. Im Anschluss wurden fünf neue Paarvergleiche in der Testphase präsentiert, bei denen TTB und das Rational Model divergierende Vorhersagen machten.

Von 40 Teilnehmern wählten bei den fünf Vergleichen jeweils 11, 18, 14, 17 und 12 Teilnehmer die von TTB indizierte Alternative. Dabei waren die Wahlen von 14 Teilnehmer in allen Vergleichen konsistent mit dem *Rational Model*, die Wahlen von 5 Teilnehmern in allen Wahlen konsistent mit TTB. Die Autoren untersuchen zudem, ob die Verwendung eines individuellen Thresholds für den Erwerb von neuer entscheidungsrelevanter Information zu einer sinnvollen Vereinigung der beiden Modelle führen könnte (TTB-Anwender brauchen weniger Information), da die intraivindividuellen Unterschiede als kleiner eingestuft werden als die interindividuellen. Die Gruppenaufteilung in zwei Gruppen mit eigenem Threshold zeigt eine gute Anpassung an die Daten, wird von den Autoren allerdings nicht kreuzvalidiert.

Diese Darstellung soll die Darstellung der beiden Heuristiken (TTB und *Recognition Heuristik*) beschließen. Im folgenden Kapitel soll eine mit TTB verwandte Klasse von Heuristiken vorgestellt werden, die direkt auf die in dieser Arbeit behandelte Problemstellung angewandt werden können und sich aufgrund der berichteten Befunde zu TTB als attraktive Alternative präsentieren.

Bon, disais-je en moi-même: les arbres tardifs sont ceux qui portent les meilleurs fruits. On grave sur le marbre bien plus malaisément que sur le sable; mais les choses y sont conservées bien plus longtemps; et cette lenteur à comprendre, cette pesanteur d'imagination, est la marque d'un bon jugement à venir.

Der Arzt Monsieur Diafoirus in Molière, Le Malade imaginaire, Akt II Szene 5 (1673)

5 Regelinduktion mit Fast and Frugal Trees

Die Diskussion der Klassifikationsmethode der *Fast and Frugal Trees* (Martignon, Vitouch, Takezawa & Forster, 2003) soll zunächst mit einer Definition der Form dieser Bäume beginnen. In Kapitel 6 werden diese Algorithmen einem umfangreichen empirischen Test unterzogen. Zunächst soll ein Überblick über die Standardform der Bäume und Spezialformen gegeben werden, im Anschluss sollen ihre Eigenschaften als Entscheidungsmodell diskutiert werden. Nach einem Überblick über empirische Studien zu *Fast and Frugal Trees* sollen Konstruktionsregeln und einige analytische Ergebnisse vorgestellt werden.

5.1 Definition von Fast and Frugal Trees

Ein *Fast and Frugal Tree* ist ein binärer Baum mit genau einem Ausgang auf jeder Ebene mit Ausnahme der Wurzel und der letzten Ebene. In der letzten Ebene gibt es genau zwei Ausgänge, so dass der Baum bei k Ebenen nach der Wurzel aus $2k + 1$ Knoten und $2k$ Kanten besteht. Von den Knoten sind $k + 1$ Blätter bzw. Endknoten, d.h. mit jeder Ebene kommen zwei Knoten hinzu. Damit steigt die Anzahl der Knoten mit der Anzahl der Ebenen linear und nicht exponentiell an wie bei vollständigen binären Bäumen (Gigerenzer, 2004a). Ein *Fast and Frugal Tree* (FaF Tree) ist ein Klassifikationsbaum mit folgenden Konventionen (es wird von einer Darstellung mit der Wurzel oben und einer Baumentwicklung von oben nach unten ausgegangen):

- Jeder Knoten, der kein Blatt ist (auch die Wurzel), entspricht der Abfrage eines binären Cues und ist mit genau einem Cue assoziiert.
- Jede Kante, die von einem Knoten oder der Wurzel ausgeht, ist mit einem Wert des mit dem Elternknoten assoziierten Cues verbunden. Konventionell ist die linke Kante mit dem positiven, die rechte Kante mit dem negativen Wert verbunden.
- Jeder Knoten, der ein Blatt ist, ist mit einem von zwei möglichen Klassifikationsergebnissen assoziiert.

Abbildung 5.1 demonstriert ein Beispiel für einen FaF Tree sowie drei Variationen der Struktur, die Bäume repräsentieren, die keine FaF Trees sind. Diese Konventionen führen zu Parsimonie in der Strukturzusammenfassung: Das einzige, was auf einer Ebene variieren kann, sind der Name des Cues und die Seite, auf der sich das Blatt befindet. Im Rahmen dieser Arbeit soll die folgende, vereinfachte Repräsentation für FaF Trees Verwendung finden: Ein Baum mit $e - 1$ Ebenen, die der Wurzel folgen, soll als geordnetes $(e - 1)$-Tupel von ganzen Zahlen dargestellt werden, wobei jedem Cue zuvor eine der Zahlen zugeordnet werden muss. Der Betrag der Zahlen entspricht dem auf einer Ebene verwendeten Cue in der Reihenfolge der Ebenen. Die letzte Zahl ist immer positiv, die übrigen Zahlen sind positiv, wenn sich das Blatt der folgenden Ebene auf der linken Seite des Elternknotens befindet und negativ, wenn sich das Blatt auf der rechten Seite des Elternknotens befindet. Der in Abbildung 5.1 dargestellte FaF Tree A

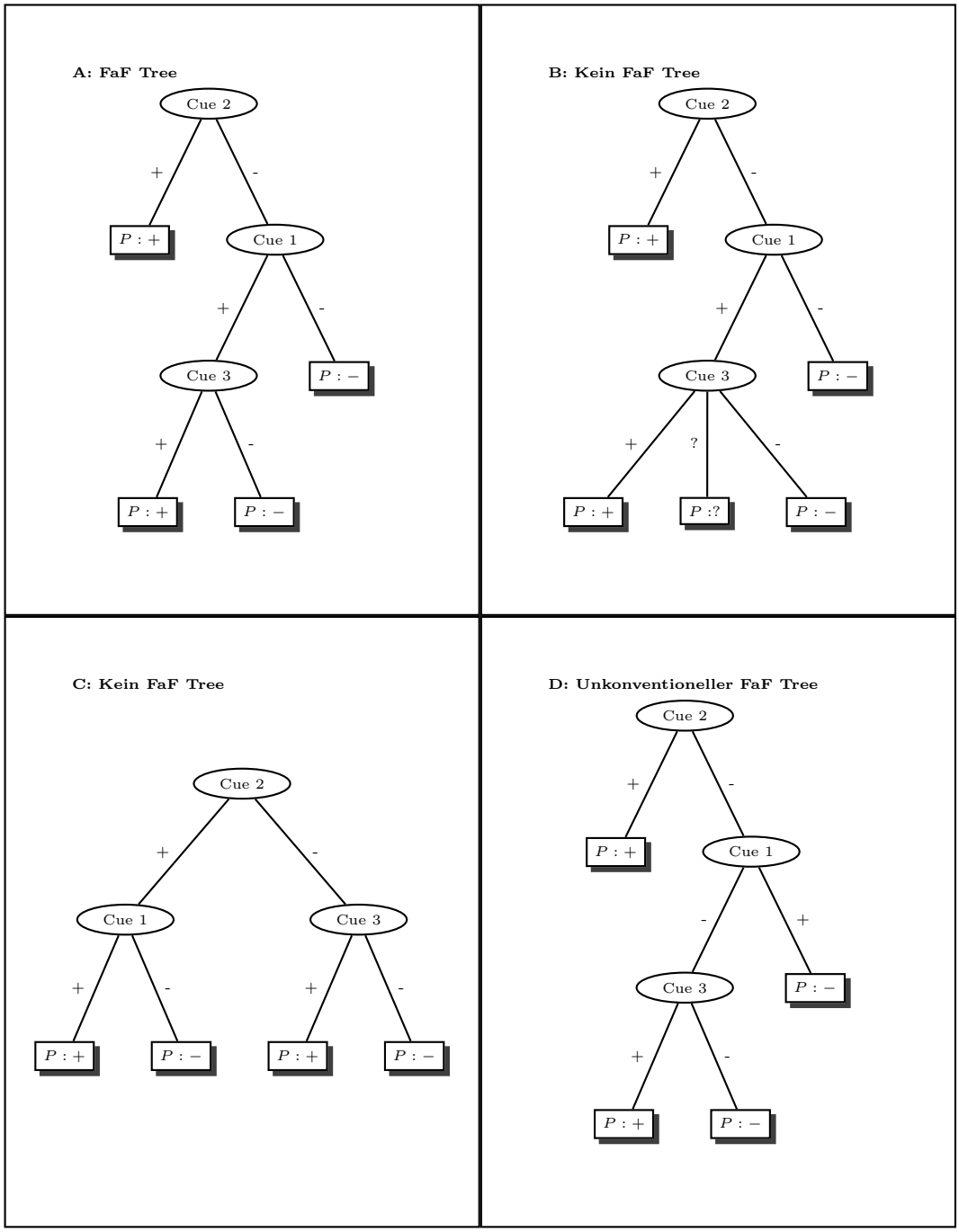


Abbildung 5.1: Abgrenzung von FaF Trees gegenüber anderen Bäumen: Dargestellt ist ein Beispiel für einen FaF Tree (A) und drei Variationen, die keine FaF Trees sind: Der Baum B ist kein binärer Baum, der Baum C hat keinen Ausgang auf der ersten Ebene, der Baum D hat falsche Beschriftungen an den Kanten zwischen den mittleren Ebenen, eine Umpolung von Cue 1 führt zur Form A.

hat bei Verwendung dieser Notation die Form $(2; -1; 3)$. Dieser Darstellungsform erfordert lediglich eine eindeutige Nummerierung der Cues.

Zwei Sonderformen von FaF Trees sollen hier Erwähnung finden. Liegen alle Blätter vor der letzten Ebene zu einer Seite des Baums, sind dementsprechend alle Zahlen im Tupel positiv oder alle bis auf die letzte Zahl negativ, heißt der Baum auch *Pectinate* oder *Rake*. Lässt sich eine Stelle vor der letzten Stelle im Tupel finden, vor der alle Zahlen das gleiche Vorzeichen haben, wobei alle Zahlen ab dieser Stelle das entgegengesetzte Vorzeichen haben (bei positivem Vorzeichen der letzten Zahl), so heißt der FaF Tree auch *L-shaped Tree* (vgl. Katsikopoulos, Martignon & Woike, i. Vorb.). Wie der Buchstabe „L“ hat dieser Baum genau eine Richtungsänderung bei den Terminalknoten vor der letzten Position. Daraus folgt direkt, dass alle FaF Trees mit drei Cues entweder *Pectinates* oder *L-shaped Trees* sind, Baum A in der Abbildung ist bspw. *L-shaped* mit einem Richtungswechsel ab der zweiten Position. Abhängig von der Richtung des ersten Blattes gibt es dementsprechend zwei Typen von Pectinates und zwei Typen von L-Shaped Trees (s. Abbildung 5.2).

5.2 Eigenschaften von Fast and Frugal Trees

Eine hervorstechende Eigenschaft von FaF Trees ist die Möglichkeit, dass Entscheidungen schnell und simpel getroffen werden können. Sie sind zum einen Beispiele für *One-Reason Decision Making*, da jeweils nur ein Cuewert auf jeder Ebene berücksichtigt werden muss, zum anderen müssen im Allgemeinen nicht alle Cuewerte abgefragt werden, so dass Geschwindigkeit neben der Einfachheit der Verrechnung durch eine einfache Stoppregel ermöglicht wird. Mit diesen Eigenschaften kann auch die Namensgebung für die Baumklasse erklärt werden.

Für die vorliegende Arbeit sind sie zudem aufgrund der Tatsache interessant, dass das ihnen zugrunde liegende Entscheidungsprinzip in der Diagnostik auch präskriptiv diskutiert wird: Amelang und Zielinski (2002) diskutieren Vorauswahl- und Vorentscheidungsstrategien, die auf ein *One-Reason Decision Making* hinauslaufen: Aufgrund eines einzelnen Testwerts kann damit eine Entscheidung ohne Berücksichtigung aller anderen Testwerte getroffen werden. Jede mehrstufige Entscheidungsprozedur, die sich in Folgephasen ignorant gegenüber den Anfangsphasen verhält, entspricht in ihrer Logik einem FaF Tree: Ein Teil von Bewerbern auf eine Position wird beispielsweise aufgrund einer Dokumentenanalyse abgelehnt, während die eingeladenen Kandidaten weiteren Tests unterzogen werden (Woike, 2003b). Die Sonderformen von FaF Trees entsprechen in besonderer Weise der Vorgehensweise bei diagnostischen Screening-Untersuchungen: Nur bei einem positiven Wert der ersten Testergebnisse werden aufwändigere weitere Verfahren durchgeführt, nur bei konsistent positiven Ergebnissen wird bei medizinischen Anwendungen eine Krankheit diagnostiziert, bzw. bei Anwendung in der Personalauswahl eine Einladung ausgesprochen oder ein Platz vergeben (vgl. Woike, 2003a).

Der Vorteil liegt in der Reduktion des Erhebungsaufwands sowohl bei Diagnostikern als auch bei Bewerbern und damit der Möglichkeit einer frühzeitigen Information über eine Absage. Die eingesparte Zeit kann für die intensivere Konzentration auf Personen verwendet werden, die das Screening erfolgreich durchlaufen haben (vgl. Wottawa et al., 2005). Zudem könnten auch Screening-Regeln für eine schnelle Zusage entwickelt werden, wenn frühzeitig deutlich wird, dass ein Bewerber für das Unternehmen gewünscht ist und verhindert werden soll, dass er andere Angebote annimmt.

Auch wenn über die Sinnhaftigkeit von spezifischen Screening-Prozeduren diskutiert werden muss (Gigerenzer, 2002), sind FaF Trees unabhängig von der Beantwortung dieser Frage in der Lage, die Logik des Screening-Prozesses abzubilden. Insbesondere bei einer hohen Bedeutung testökonomischer Aspekte (vgl. Woike, 2003c) sollte der Ansatz an Attraktivität gewinnen, wobei ähnlich wie bei nicht-kompensatorischen Strategien bei binären Paarvergleichen nicht notwendigerweise ein hoher Verlust in der Entscheidungsgenauigkeit riskiert werden muss. *Pectinates* entsprechen im Ergebnis einer Konjunktionsregel von Bedingungen, in diesem Sinne ist die Reihenfolgeveränderung von Cues für das Ergebnis bedeutungslos, auch wenn sich die im Prozess abgerufenen Cuezahlen ändern können. Bei L-Shaped Trees gilt diese Beziehung für die beiden Teilmengen von Cues in den beiden Teilen des Baumes.

Katsikopoulos und Fasolo (2006) bewerten Fast and Frugal Trees als nützliche Entscheidungswerkzeuge und unterstreichen die Eigenschaften der Genauigkeit und Robustheit:

> Overall, there are good mathematical reasons why those trees are, under some conditions, accurate, robust, and stable. For example, they are robust and stable because they do not

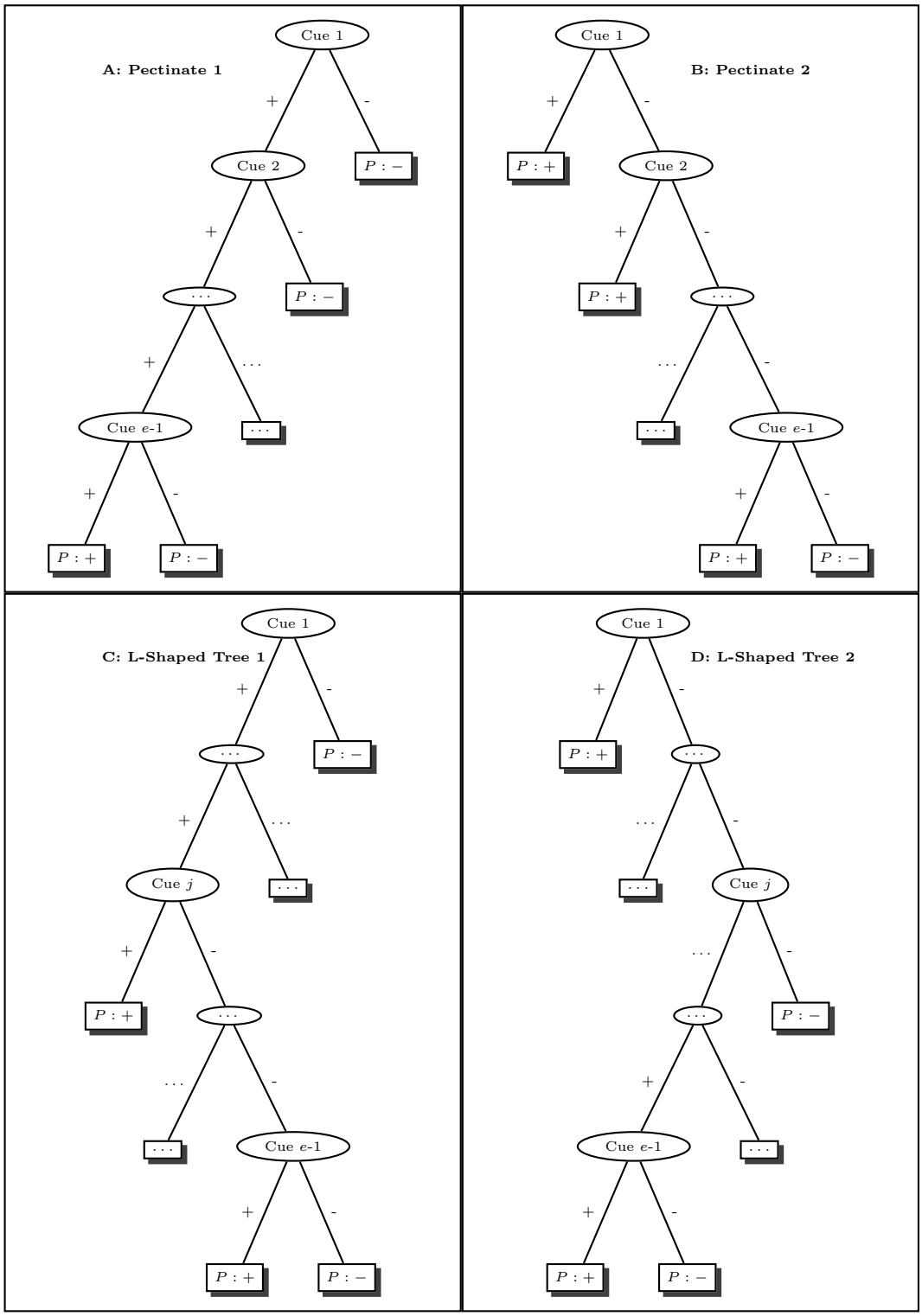


Abbildung 5.2: Pectinates und L-Shaped Trees: Die Bäume A und B repräsentieren die beiden Varianten von Pectinates, die Bäume C und D die Varianten von L-Shaped Trees.

> attempt to model in detail the interdependencies between cues. Because they are frugal, they might [...] plausibly enhance the quality of information elicited through decreased fatigue and increased attention, and they might be more requisite. (S. 965)

Die Anwendung von Fast and Frugal Trees hat den Vorteil einer leichten Vermittelbarkeit und einer Unabhängigkeit der Entscheidungsverfahrens von der Verfügbarkeit des ursprünglichen Datensatzes und seiner Eigenschaften:

> [...] fast and frugal models provide a transparent, non-mathematical description of judgement behaviour. They are simple to apply and can be applied without knowledge of the properties of the data set on which they were developed. For example, if the modes says prescribe if male and the case is male, then we do not need to know the natural variance of gender in the environment. (Dhami & Ayton, 2001, S. 22)

Die in dieser Arbeit betrachteten hierarchischen Regelsysteme weisen substanzielle Ähnlichkeiten mit FaF Trees auf: Betrachtet man auch komplexere Antezedenzbedingungen als legitime Cues, so kann die Entscheidungsprozedur bei jeder Regel abgebrochen werden. Gleichzeitig werden die Regeln in festgelegter Reihenfolge abgearbeitet. Aufgrund dieser Analogie sind die Ergebnisse von FaF Trees sehr relevant für diese Arbeit, da sich die Struktur eines FaF Trees wiederum sehr einfach in ein hierarchisches Regelsystem überführen lässt.

5.3 Empirische Studien zur Anwendung von Fast and Frugal Trees

Es ist eine offene Frage, in welchen Situationen menschliche Entscheider ohne Anweisungen und Hilfsmittel auf Strategien zurückgreifen, die dem Entscheidungsmodell von FaF Trees entsprechen, und ob diese Bäume zu einem Modell für menschliche Kognition werden können (s. Katsikopoulos, Woike & Brighton, in Vorb.).

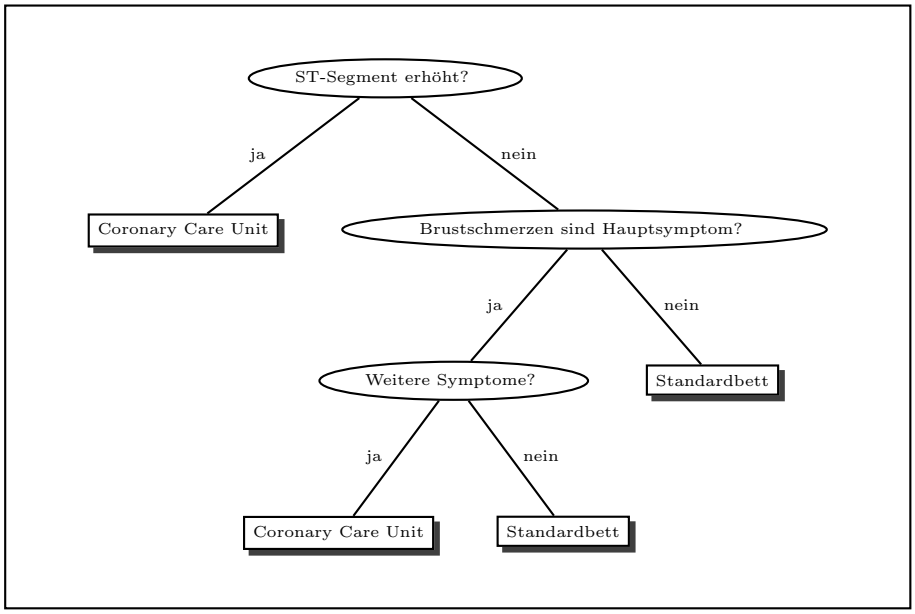


Abbildung 5.3: Entscheidungsbaum für die Einweisung von Patienten mit Verdacht auf Herzkrankheit nach Green und Mehr (1997).

Der Bereich in dem FaF Trees ursprünglich entwickelt wurden, ist der Bereich des Medical Decision Making. Green und Mehr (1997) beobachteten zunächst eine problematische Entscheidungspraxis, wenn es darum ging, bei Patienten mit Verdacht auf ischemische Herzkrankheit die Entscheidung zu treffen, sie entweder in die speziell dafür eingerichtete Abteilung („coronary care unit“) einzuweisen oder sie in ein reguläres Krankenhausbett zu verlegen. Aus Angst vor Nachlässigkeit und rechtlichen Folgen wurden zu viele Patienten in die Spezialabteilung verlegt, 75% der eingewiesenen Patienten hatten keinen myokardialen Infarkt und wären in einem regulären Krankenhausbett besser aufgehoben gewesen, da das Risiko

von Sekundärinfektionen sowie eine Überbelegung der Abteilung die Folge waren. Das Training mit einer Entscheidungshilfe, dem „Heart Disease Predictive Instrument“ (Pozen, D'Agnostino, Selker, Sytkowski & Hood, 1984) wurde nicht bereitwillig angenommen: Ärzte hatten eine Tabelle mit 50 Wahrscheinlichkeiten zur Verfügung, wobei die aufgrund von sieben beobachteten Symptomen ermittelten Tabellenwerte mit Hilfe eines Taschenrechners zur Wahrscheinlichkeit für eine Herzkrankheit verrechnet werden mussten. Bei der Analyse des unterstützten und nicht unterstützten Entscheidungsverhaltens stellten Green und Mehr fest, dass das Instrument zu einer Verbesserung der ärztlichen Entscheidungspraxis führte, die aber unerwarteterweise stabil blieb, nachdem das Instrument zu Kontrollzwecken wieder entfernt worden war.

Die Ärzte hatten im Umgang mit dem Instrument eine Heuristik entwickelt, die das Nachschlagen und Verrechnen von Wahrscheinlichkeiten nicht mehr notwendig machte. Nach dem Vorbild der TTB-Heuristik rekonstruierten Green und Mehr (1997) das Entscheidungsverhalten in einer Form, die sich als Entscheidungsbaum darstellen lässt (s. Abbildung 5.3). Patienten, deren EKG ein erhöhtes ST-Segment zeigt, werden vom Modell ohne Konsultation weiterer Symptome direkt in die Spezialeinheit eingewiesen. Patienten ohne erhöhtes Segment, bei denen Brustschmerzen nicht das Hauptsymptom sind, werden direkt in einem regulären Bett untergebracht. Bei Patienten ohne erhöhtes Segment, aber mit Brustschmerzen als Hauptsymptom, wird überprüft, ob mindestens eines von fünf weiteren Symptomen beobachtet werden kann. Dieses Modell zeigt eine überraschend gute Klassifikationsleistung:

> How did the fast and frugal tree perform? The counterintuitive result is that the fast and frugal tree was more accurate in classifying actual heart attack patients than both the physicians´intuition and the HDPI. It correctly assigned the largest proportion of patients who subsequently had a myocardial infarction into the coronary care unit. At the same time, it had a comparatively low false-alarm rate. Note that the expert system had more information than the smart heuristic and could make use of sophisticated statistical calculations. Nevertheless, in this complex situation, less is more. (Gigerenzer & Kurzenhäuser, 2005, S. 10)

Die Analyse dieses Entscheidungsmodells stellt den Ausgangspunkt für die Entwicklung der Faf Trees bei Martignon et al. (2003) dar, so dass sich der Kreislauf von Modell zur Praxis zurück zum Modell hiermit schließt. Kenneth L. Noller - ein Herausgeber - verweist in Reaktion auf (Gigerenzer & Kurzenhäuser, 2005) auf den Ursprung heuristischer Praktiken in der Behandlung von verwundeten Soldaten, bei denen eine Entscheidung über Leben und Tod buchstäblich in Sekunden zu treffen war. Gebräuchlicher als den Ausdruck „heuristic“ schätzt er in amerikanischen medizinischen Kreisen allerdings den Begriff „triage“ ein. In weitere Anwendungen von FaF Trees im Bereich medizinischer Entscheidungen werden etwa Entscheidungsregeln für das Verschreiben von Medikamenten generiert (Fischer, Steiner, Zucol, Berger & al., 2002).

Einige Hinweise auf den Erfolg dieser Bäume als deskriptives Modell für juristische Entscheidungen liefern Dhami und Ayton (2001). Sie untersuchen das Verhalten von 81 Beamten an 44 Gerichten in England, die in ca. 40 hypothetische Fällen von Anträgen auf Kaution entscheiden sollten, wobei ihnen als Entscheidungsgrundlage unter anderem Angaben zu Geschlecht, Alter, ethnischer Herkunft gegeben wurden, weiterhin die Entscheidung der Polizei, Forderung der Anklage und Informationen über frühere Entscheidungen. Bei dem Versuch, die Entscheidungsstrategie der Beamten zu modellieren, wurde neben kompensatorischen Verfahren wie der Gleichgewichtungsstrategie oder gewichteten Cuekombinationen auch das von den Autoren als *Matching Heuristic* bezeichnete Verfahren geprüft. Dabei wird eine parameterbestimmte Anzahl von Cues überprüft, wobei ein positiver Wert eines Cues zu einer Entscheidung gegen eine Kaution führt und nur bei negativen Werten aller Cues die Kaution gewährt wird. Dieses Entscheidungsprinzip kann durch einen FaF Tree modelliert werden, genauer gesagt, einem Baum in *Pectinate*-Form. In der Auswertung können folgende Beobachtungen angestellt werden: Der Parameter der *Matching Heuristic* wird bei fast allen Beamten empirisch so bestimmt, dass lediglich ein oder zwei Cues in den Baum aufgenommen wurden. Die Ergebnisse der Matching Heuristic übertreffen die Ergebnisse der kompensatorischen Strategien. Geprüft wird dabei neben der Fitting-Leistung auch die Vorhersageleistung. Die *Matching Heuristic* klassifiziert dabei 73,98% (Fitting) bzw. 65,61% (Vorhersage) der Fälle korrekt. Die aufgenommenen Cues divergieren dabei zwischen Beamten. Dabei bewerten die Autoren ihre Ergebnisse durchaus kritisch:

> Gigerenzer and his colleagues have evaluated the prescriptive utility of fast and frugal models solely in terms of their accuracy (this is often global accuracy). In many applied

> domains, such as the legal domain, however, accuracy is not the primary concern [...]. Thus, as we gave argued, although fast and frugal trees are descriptively and predictively valid, they may lack prescriptive utility. (Dhami & Ayton, 2001, S. 162)

Die Entscheidungen der Beamten haben gravierende Auswirkungen auf das Leben der Antragssteller, nachweislich steigen die Chancen auf familiäre Probleme, Reputationsverlust, Wohnungsprobleme bei Ablehnung einer Kaution, weiterhin wird implizit eine Annahme von Schuld getroffen. Die Ergebnisse stehen zudem im Widerspruch zur Strategiebeschreibung der Beamten. Wichtiger als dies würde eine normative Anwendung der Bäume zu einer Verletzung der Grundlagen des Rechtssystems führen:

> The lack of due process observed in magistrates´ bail decision making raises doubts about their performance on more complicated and less frequently made decisions such as the decision to convict. Fortunately, our conceptions of crime and order are socially constructed and so the concerns expressed by different groups in society regarding the appropriateness of some of the legal decisions made have often led to specific changes in legal rules and procedures. (Dhami & Ayton, 2001, S. 164)

Die *Matching Heuristic* wird von Smith und Gilhooly (2006) einem Vergleich mit der logistischen Regression unterzogen, bei dem das Kriterium die Vorhersage von Entscheidungen für die Verschreibung von Medikamenten gegen Depression ist:

> The MH model achieved almost as good a fit as the logistic regression model with a more economical explanation in terms of amount of information used [...]. (Smith & Gilhooly, 2006, S. 272)

Zwischen der deskriptiven und der präskriptiven Tauglichkeit von FaF Trees in Bezug auf praktische Entscheidungsprozesse muss in vielen Fällen allerdings strikt unterschieden werden, wobei Smith und Gilhooly (2006) eine Integration der Prozedur in Werkzeuge zur Entscheidungsunterstützung für erwägenswert halten. In anderen Fällen kann die Explikation impliziter Entscheidungsregeln gewünscht sein, nicht mit der Intention einer Kritik der bestehenden Praxis, sondern mit der Absicht, eine Entscheidungsstrategie kommunizierbar zu machen. Damit kann die Modellrepräsentation der Erfahrungssicherung dienen (Wottawa & Hossiep, 1987), gleichzeitig ermöglicht die Explikation aber auch die Bereitstellung einer Grundlage für die Modifikation der Praxis. Ein weiterer Vorteil ist die Möglichkeit eines Verzichts auf redundante Information:

> In den meisten Anwendungsfällen zeigt es sich, daß für die Routinefälle mit wesentlich weniger Information die Entscheidung getroffen werden kann, als üblicherweise erhoben wird (meist reichten weniger als 25% der routinemäßig erhobenen Daten für die Entscheidung aus). Dies ermöglicht es, die Diagnostik als solche „sequentiell“ zu gestalten, indem zunächst bei den Probanden die im Regelfall für die Entscheidung benötigten Informationen gesammelt werden, und nur für jene Teilgruppe, für die auf der Basis der vorliegenden Information noch keine modellmäßige Entscheidung getroffen werden kann, die Datenerhebung selbst zu intensivieren. (Wottawa & Hossiep, 1987, S. 69)

Diese Betrachtung läuft der Intuition einiger Praktiker zuwider. Auch wenn Wottawa und Hossiep (1987) von einer positiven Aufnahme der explizierten Modelle durch Mediziner berichten, konstatiert Gigerenzer (2005b):

> ...the state of the art in medicine and other fields is still shaped by the misleading ideal of full information. Collecting all pieces of information and combining them in an „optimal“ way, such as in logistic regression, are assumed to always be the best thing to do. The wisdom in simplicity is news to many. (S. 50)

Wottawa und Echterhoff (1982) geben ein Beispiel für einen Entscheidungsbaum, der mit Hilfe von HYPAG/Structure (Wottawa, 1987a) in der Analyse von Umschulungsentscheidungen in Arbeitsämtern erstellt wurde (s. Abbildung 5.4). Dabei wird auf der Grundlage der Ergebnisse von sechs psychologischen Testverfahren (Leistungstests zu Rechenleistung, Intelligenz und Wortschatz) eine Entscheidung

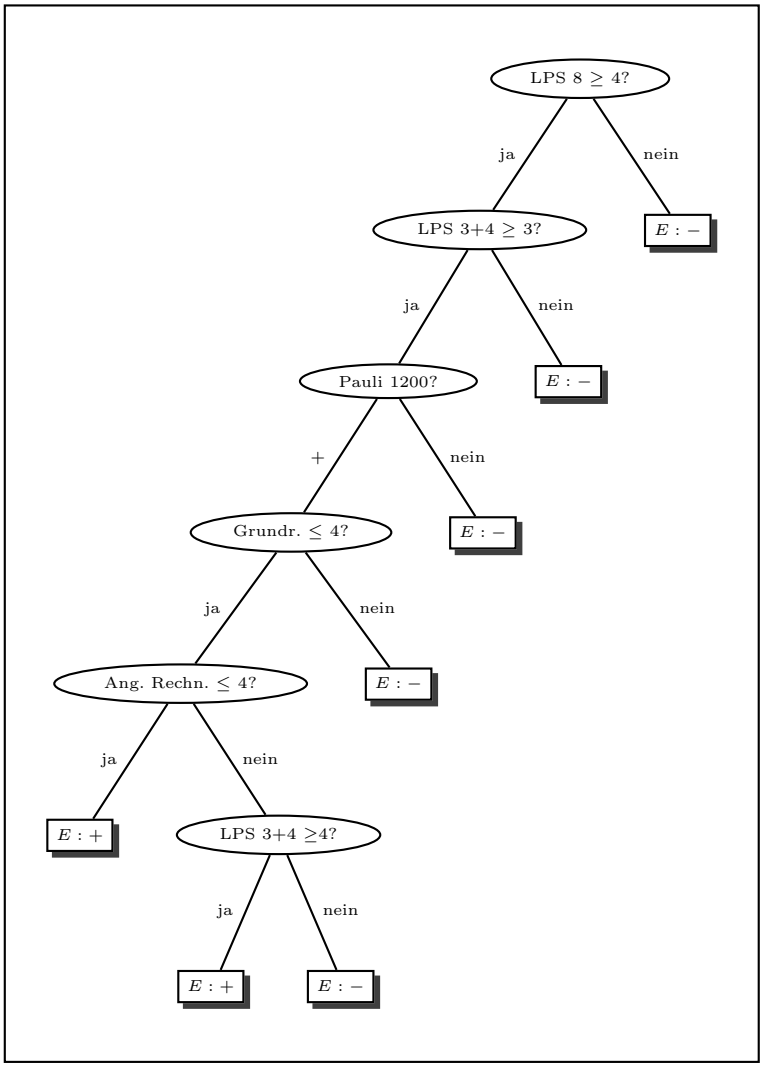


Abbildung 5.4: Entscheidungsbaum für Umschulungsentscheidungen nach Wottawa und Echterhoff (1982): Augrund von sechs Testergebnissen wird eine Entscheidung für oder gegen eine Umschulung getroffen.

für oder gegen eine Maßnahme getroffen. In der Vorhersage weist das Modell eine prädiktive positive Validität von 93,3% und eine prädiktive negative Validität von 100% auf. Lässt man zu, dass ein Cue aus der Zusammenfassung von zwei Testergebnissen konstruiert werden kann und dass weiterhin durch Vergleich mit zwei verschiedenen Grenzen verschiedene Cues konstruiert werden können, dann hat der Entscheidungsbaum die Struktur eines *L-Shaped* Trees, wobei die ersten vier Cues bei einem negativen Wert zu einer sofortigen Negativentscheidung führen, der fünfte Cue bei einem positiven Wert zu einer sofortigen Positiventscheidung. Sind in einem Fall also vier Ergebnisse positiv, dann erfolgt eine negative Entscheidung nur bei zwei Gegenargumenten.

Wottawa und Oenning (2002) geben das Ergebnis einer Auswertung mit AnswerTree in Regellistenform an, das sich in den in Abbildung 5.5 dargestellten Entscheidungsbaum überführen lässt, der erneut nahezu die Struktur eines Fast and Frugal Tree aufweist. Hier werden die Bewertungen der beruflichen Leistung durch Vorgesetzte durch einen Leistungstest (LT) und vier Persönlichkeitsskalen (PT1 bis PT4) vorhergesagt. Lediglich in der letzten Ebene gibt es eine undefinierte Entscheidung, so dass der Baum nicht als vollständiger FaF Tree klassifiziert werden kann. Auch eine Analyse der CART-Ergebnisse von Montel (2005) zeigt, dass die beiden geprunten CART-Bäume eine Form aufweisen, die sich als Fast and Frugal Tree mit drei Variablen darstellen ließe. Damit kann festgehalten werden, dass FaF Trees zum einen in der Entscheidungspraxis gefunden werden können und sich zum anderen auch zur Modellierung von Entscheidungen empfehlen.

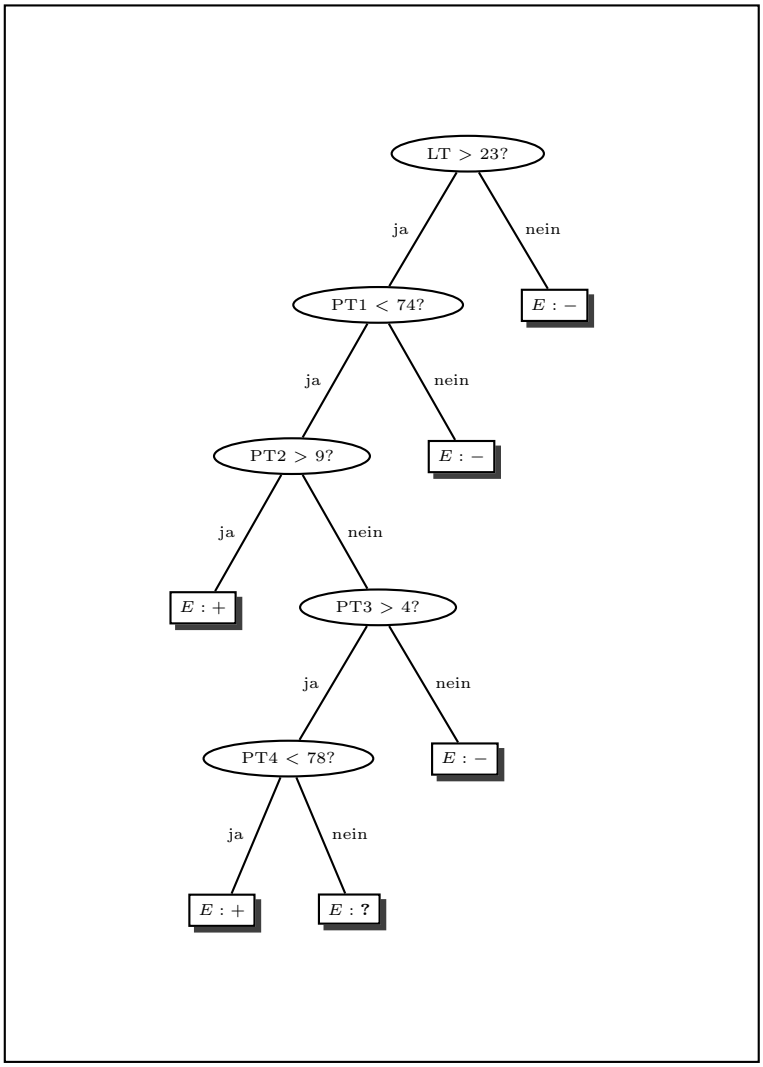


Abbildung 5.5: Entscheidungsbaum für die Vorgesetztenbeurteilung nach Wottawa und Oenning (2002): Aufgrund von fünf Testergebnissen wird die Beurteilung des Vorgesetzten (positiv oder negativ) vorhergesagt. Der Baum ist Fast and Frugal mit Ausnahme einer unbestimmten Entscheidung in der letzten Ebene.

5.4 Konstruktionsregeln für Fast and Frugal Trees

Die bisherige Diskussion orientierte sich an einem vollständig aufgebauten FaF Tree. Wenngleich durchaus die Möglichkeit bestünde, dass konstruierte Bäume in Entscheidungssituationen angetroffen werden könnten, muss an dieser Stelle auf ein zentrales und offenes Probleme des Ansatzes hingewiesen werden:

Angenommen, Cue-Profile und Kriteriumswerte (Entscheidungen bei einer Rekonstruktion, Klassenzugehörigkeiten bei einer Klassifikation) für eine Stichprobe aus Objekten sind bekannt. Die Frage ist nun, wie auf der Grundlage dieser Informationen ein FaF Tree konstruiert werden kann, der in seinen Vorhersagen möglichst gut den gegebenen Kriteriumswerten entspricht. Genauer gesagt muss zum einen die Reihenfolge der Cues bestimmt werden, zum anderen die Richtung der Blätter für alle bis auf den letzten Cue.

5.4.1 Das Konstruktionsproblem

Eine Möglichkeit bestünde sicherlich darin, alle möglichen Bäume zu konstruieren, und ihre Fitting-Genauigkeit zu überprüfen. Die Anzahl der möglichen Bäume ergibt sich als Anzahl der möglichen Permutationen der Cues multipliziert mit der Anzahl der möglichen Strukturen für jede Permutation, also für k Cues als

$$n_k = k! \cdot 2^{k-1}. \tag{5.1}$$

Tabelle 5.1: Anzahl möglicher FaF Trees bei verschiedenen Cuezahlen

Anzahl Cues	Permutationen	Strukturen	Anzahl Bäume
1	1	1	1
2	2	2	2
3	6	4	24
4	24	8	192
5	120	16	1 920
6	720	32	23 040
7	5 040	64	322 560
8	40 320	128	5 160 960
9	362 880	256	92 897 280
10	3 628 800	512	1 857 945 600
20	2 432 902 008 176 640 000	524 288	1 275 541 328 062 914 232 320 000

Tabelle 5.1 zeigt die Anzahl möglicher Bäume, Permutationen und Strukturen abhängig von der Anzahl der Cues. Ist die Vorgehensweise der Berücksichtigung aller Cues bei kleinen Cuezahlen prinzipiell noch möglich, überfordert sie bei höheren Zahlen auch zeitgenössische Computer. Der Suchraum muss also begrenzt werden, um in vertretbarer Zeit zu einer Lösung zu gelangen. Selbst für kleine Cuezahlen ist das Vorgehen nur für die *Fitting*-Aufgabe ideal: die Studien zu TTB zeigen, dass die im *Fitting* optimale Struktur in der Vorhersageleistung schlechtere Werte erreicht als die TTB-Strategie.

5.4.2 Fünf Vorschläge für Konstruktionsregeln

Tabelle 5.2: Beispielergebnisse für vier Cues in einer fiktiven Aufgabestellung

	Cue + / Krit. +	Cue - / Krit. +	Cue + / Krit. -	Cue - / Krit. -
Cue 1	10	30	15	45
Cue 2	25	15	5	55
Cue 3	40	0	30	30
Cue 4	20	20	20	40

Tabelle 5.3: Gütekriterien für vier Cues aus Tabelle 5.2

	Sensitivität	Spezifizität	Positive Validität	Negative Validität	Genauigkeit
Cue 1	*0,250*	**0,750**	*0,400*	**0,600**	0,550
Cue 2	*0,625*	**0,917**	**0,833**	*0,786*	0,800
Cue 3	**1,000**	*0,500*	*0,571*	**1,000**	0,700
Cue 4	*0,500*	**0,667**	*0,500*	**0,667**	0,600

Martignon et al. (2003) schlagen ein ähnliches Vorgehen wie in der TTB-Heuristik für FaF Trees vor. Es stehen mehrere mögliche Kennwerte zur Auswahl, die die Rolle der Validität für TTB im Fall der Bäume übernehmen könnten: Eine Klassifikation ist im Unterschied zum Paarvergleich keine symmetrische Entscheidung, d.h. es gibt zwei unterschiedliche Klassen von korrekten Ergebnissen und zwei unterschiedliche Arten von Fehlern. Aus diesem Grund sollen einige der für die Bewertung von Klassifikationen genannten Kennwerte berücksichtigt werden, namentlich Sensitivität, Spezifität, negative und positive Validität sowie die Klassifikationsgenauigkeit einzelner Cues. Diese Werte werden berechnet, indem angenommen wird, dass die Klassifikationsentscheidung mit dem Wert des Cues übereinstimmt (ein positiver Cuewert bedeutet eine positive Kriteriumsvorhersage und umgekehrt), so dass alle Werte für jeden einzelnen Cue berechnet werden können. Tabelle 5.9 zeigt eine Beispielverteilung von Ergebnissen für vier Cues, Tabelle 5.3 zeigt die aus der Verteilung berechneten Gütekriterien. Es sollen hier fünf Verfahren zur Konstruktion von FaF Trees vorgeschlagen werden: zwei Verfahren (FaF ZigVal und FaF ZigSens) entsprechen Verfahren in Martignon et al. (2003), zwei Verfahren (FAF MaxVal und FAF MaxSens) stellen leichte Modifikationen von Verfahren in Martignon et al. (2003) dar, ein Verfahren (FAF MaxAcc) wird als Vervollständigung der Analogie zu TTB eingeführt.

Die Verfahren sind im Einzelnen:

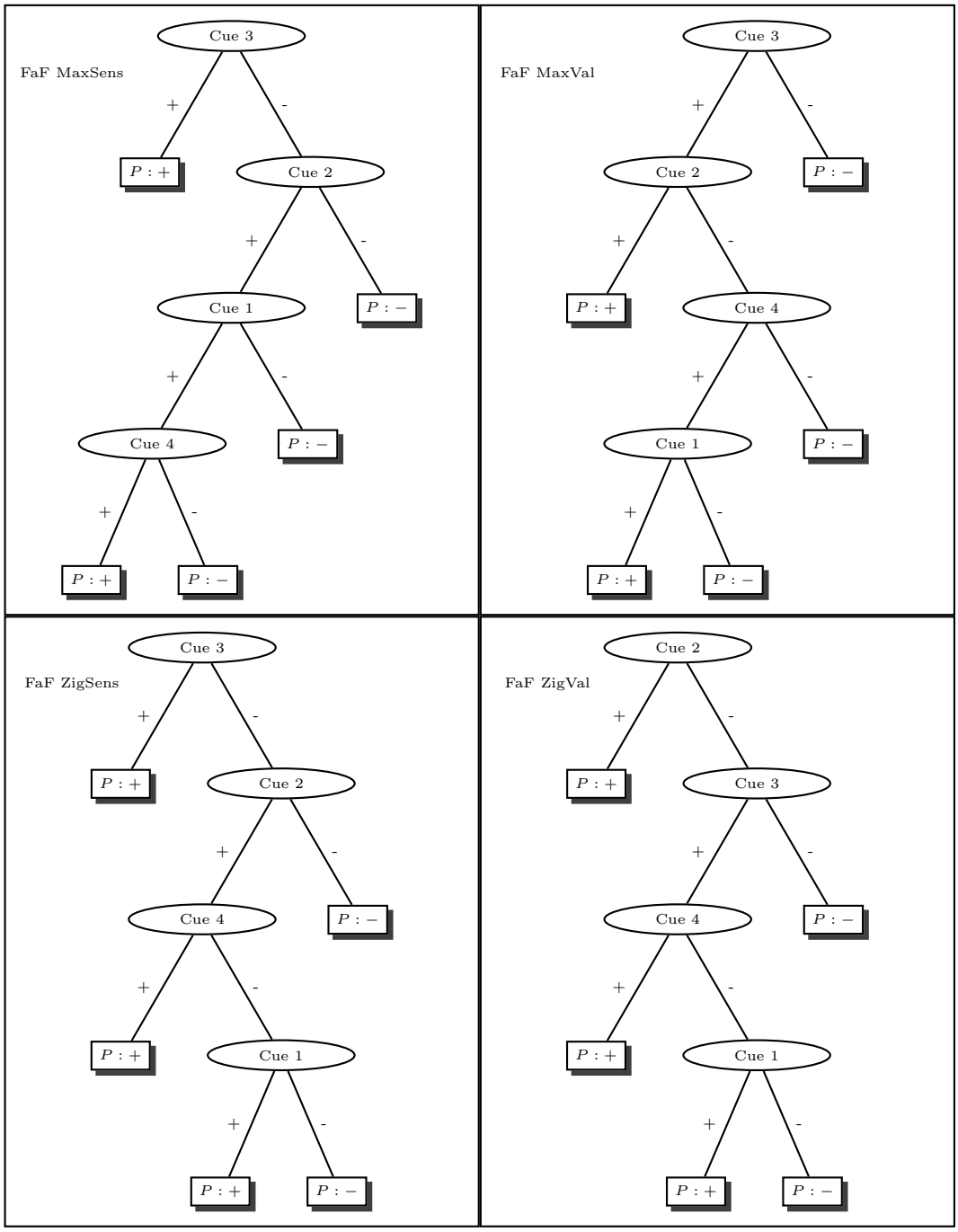


Abbildung 5.6: Anwendung der Konstruktionsregeln Faf MaxSens, Faf MaxVal, FaF ZigSens und FaF ZigVal auf die fiktiven Cues aus Tabelle 5.2

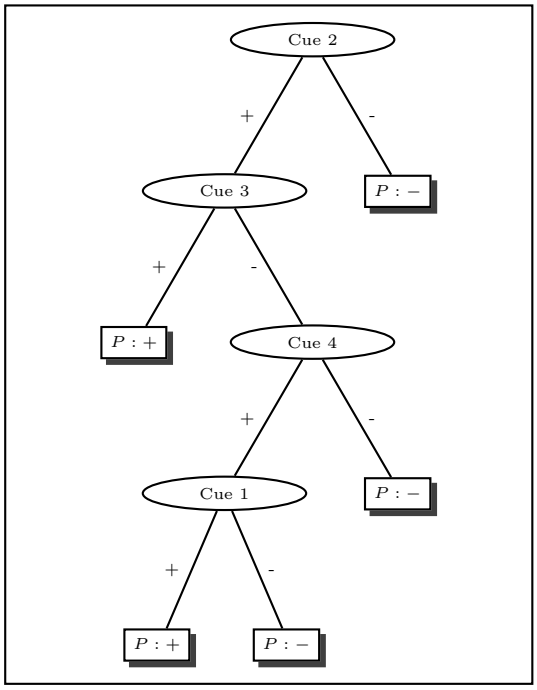


Abbildung 5.7: Anwendung der Konstruktionsregel Faf MaxAcc auf die fiktiven Cues aus Tabelle 5.2

1. **FaF MaxVal**: Die Cues werden in der Reihenfolge der Maxima aus positiver und negativer Validität des Cues sortiert. Ist die positive Validität der einer Ebene zugeordneten Variablen größer als die negative Validität, wird der Baum auf dieser Ebene mit einem Entscheidungsknoten für die positive Kategorie verlassen, ansonsten mit einem Entscheidungsknoten für die negative Kategorie. Das Verfahren entspricht dem *Max(val+, val-)*-Verfahren mit dem Unterschied, dass bei FaF MaxVal auch auf der ersten Ebene die Richtung aufgrund des Validitätenvergleichs bestimmt wird und nicht als positiv gesetzt wird.

2. **FaF MaxSens**: Die Cues werden in der Reihenfolge der Maxima aus Sensitivität und Spezifität eines Cues sortiert. Ist die Sensitivität der einer Ebene zugeordneten Variablen größer als die Spezifität, wird der Baum auf dieser Ebene mit einem Entscheidungsknoten für die positive Kategorie verlassen, ansonsten mit einem Entscheidungsknoten für die negative Kategorie. Das Verfahren entspricht dem *Max(sens, spec)*-Verfahren mit dem Unterschied, dass bei FaF MaxSens auch auf der ersten Ebene die Richtung aufgrund des Vergleichs bestimmt wird und nicht als positiv gesetzt wird.

3. **FaF ZigVal**: Auf der ersten Ebene wird der Baum mit einem positiven Entscheidungsknoten verlassen, auf den folgenden Ebenen wird zwischen negativen und positiven Entscheidungsknoten alterniert. Der verwendete Cue wird auf jeder Ebene durch den noch nicht eingesetzten Cue gestellt mit dem Maximum an positiver Validität bei positiven Entscheidungsknoten und dem Maximum an negativer Validität bei negativen Entscheidungsknoten. Das Verfahren entspricht dem *Zigzag(val+,val-)-Verfahren.*

4. **FaF ZigSens**: Auf der ersten Ebene wird der Baum mit einem positiven Entscheidungsknoten verlassen, auf den folgenden Ebenen wird zwischen negativen und positiven Entscheidungsknoten alterniert. Der verwendete Cue wird auf jeder Ebene durch den noch nicht eingesetzten Cue gestellt mit dem Maximum an Sensitivität bei positiven Entscheidungsknoten und dem Maximum an Spezifität bei negativen Entscheidungsknoten. Das Verfahren entspricht dem *Zigzag(sens,spec)-Verfahren.*

5. **FaF MaxAcc**: Die Cues werden in der Reihenfolge der Klassifikationsgenauigkeit sortiert. Ist die Anzahl korrekt zugeordneter Fälle der positiven Kategorie größer als die Anzahl der korrekt zugeordneten Fälle der negativen Kategorie, wird der Baum auf dieser Ebene mit einem Entscheidungsknoten für die positive Kategorie verlassen, ansonsten mit einem Entscheidungsknoten für die

negative Kategorie.

Abbildung 5.6 zeigt die aus einer Anwendung der ersten vier Regeln auf die Cues aus Tabelle 5.2 resultierenden vier Bäume, Abbildung 5.7 das Ergebnis der Anwendung von FaF MaxAcc. Die Struktur von alternierenden Bäumen ist dabei mit Ausnahme der Cuereihenfolge durch den Algorithmus (ohne Betrachtung der Daten) vorgegeben, Struktur und Cuereihenfolge bei den anderen Verfahren können theoretisch frei variieren.

Tabelle 5.4: Drei Mögliche Verteilungen mit den Ergebnissen aus Tabelle 5.2

Cue 1	Cue 2	Cue 3	Cue 4	Krit.	Verteilung 1	Verteilung 2	Verteilung 3
-	-	-	-	-	17	26	11
-	-	-	-	+	0	0	0
-	-	-	+	-	5	2	9
-	-	-	+	+	0	0	0
-	-	+	-	-	12	11	25
-	-	+	-	+	5	5	0
-	-	+	+	-	8	6	0
-	-	+	+	+	3	0	10
-	+	-	-	-	1	0	0
-	+	-	-	+	0	0	0
-	+	-	+	-	0	0	0
-	+	-	+	+	0	0	0
-	+	+	-	-	1	0	0
-	+	+	-	+	1	5	0
-	+	+	+	-	1	0	0
-	+	+	+	+	1	0	0
+	-	-	-	-	4	0	3
+	-	-	-	+	0	0	0
+	-	-	+	-	2	2	7
+	-	-	+	+	0	0	0
+	-	+	-	-	4	2	0
+	-	+	-	+	6	10	5
+	-	+	+	-	3	6	0
+	-	+	+	+	1	0	0
+	+	-	-	-	0	0	0
+	+	-	-	+	0	0	0
+	+	-	+	-	1	0	0
+	+	-	+	+	0	0	0
+	+	+	-	-	1	1	1
+	+	+	-	+	8	0	15
+	+	+	+	-	0	4	4
+	+	+	+	+	15	20	10

Die Klassifikationsgüte der Bäume lässt sich nicht aufgrund der bisher aufgeführten Daten beurteilen. Hierzu ist die Kenntnis über die Verteilung von Kriteriumswerten auf die vorhandenen Cueprofile notwendig. Die genannten Kennwerte schränken allerdings die Verteilungen, die sie generieren können, in bestimmter Hinsicht ein: Aus der Tabelle lässt sich entnehmen, dass für Cue 1 insgesamt 25 Fälle einen positiven Cuewert besitzen, für Cue 2 sind es 30 Fälle, 70 und 40 sind die Fallzahlen mit positivem Cuewert für Cue 3 und Cue 4 respektive. Die Anzahl der positiven und negativen Kriteriumswerte muss für alle Cues die gleiche sein[1]. Drei hypothetische Verteilungen, die diesen Einschränkungen gerecht werden, sind in Tabelle 5.4 aufgeführt. Für jede Cuepermutation kann eine Verteilung auch hierarchisch präsentiert werden. Für die Cureihenfolge $(1, 2, 3, 4)$ ergibt sich das in Abbildung 5.8 präsentierte Bild: Auf der ersten

[1] Weitere Einschränkungen ergeben sich aus Abhängigkeiten zwischen den Zahlen. Da es kein Objekte geben kann, das einen negativen Cuewert für Cue 3 aufweist und einen positiven Kriteriumswert, gibt es ebenso kein Objekt mit einem positiven Kriteriumswert und einem Cueprofil, das einen negativen Cuewert für Cue 3 enthält.

Ebene wird die Menge der Objekte in eine Teilmenge mit positivem Cuewert für Cue 1 und eine Teilmenge mit negativem Cuewert aufgeteilt. Die Objekte mit unterschiedlichem Kriteriumswert teilen sich dabei auf die Teilmengen auf wie in Tabelle 5.2 angezeigt. In den folgenden Baumebenen werden die Teilmengen feiner differenziert, bis auf der letzten Ebene die Verteilung der Kriteriumswerte für vollständige Cueprofile abgelesen werden kann. Auf dieser Ebene kann auch die obere Schranke für die Klassifikationsgenauigkeit eines Verfahrens abgelesen werden, das seine Entscheidung deterministisch auf Cueprofile stützt: Wann immer ein Cueprofil von Objekten demonstriert wird, die unterschiedliche Kriteriumswerte aufweisen, kann eine der beiden Teilgruppen nicht korrekt klassifiziert werden, die obere Grenze ergibt sich als Anteil jener Objekte an der Stichprobe, die sich auf der unteren Ebene in der größeren Gruppe befinden an allen Objekten. Im Beispiel ist dieser Anteil als $\frac{15+8+1+3+6+2+4+1+1+1+8+12+5+17}{100} = 0,84$ bestimmbar, mit anderen Worten: 16 Objekte müssen mindestens von jedem Verfahren falsch klassifiziert werden, das deterministisch aufgrund von Cueprofilen urteilt. Das optimale Ergebnis für die Fitting-Aufgabe wäre also bei Gleichbewertung der Fehlerarten mit einer Klassifikationsgenauigkeit von $0,84$ erreicht.

Die hierarchische Darstellungsform ist ebenfalls geeignet für eine Analyse der Klassifikationsgenauigkeit der konstruierten FaF Trees. Dies soll am Beispiel für den FaF MaxSens-Baum hinsichtlich der ersten Verteilung gezeigt werden. Hierzu wird die Permutation gewählt, die der Baum aus den Daten ermittelt hat, also $(3, 2, 1, 4)$. Die Darstellung in Abbildung 5.9 zeigt deshalb zunächst die Kriteriumswertverteilung bei hierarchischer Cueaufteilung der Objekte. Weiterhin sind die Entscheidungsschritte des Baums in folgender Weise markiert: Jeder Knoten, der dem Cueprofil in einem Entscheidungsknoten des Baumes entspricht, ist blau markiert, die Teilmengen von Fällen, für die in den einzelnen Baumstufen noch eine Entscheidung zu treffen ist, sind in den stark schwarz umrandeten Knoten enthalten. Im ersten Schritt werden alle Objekte mit positivem Cuewert für Cue 3 positiv klassifiziert, darunter sind 30 Objekte mit negativem Kriteriumswert und 40 Objekte mit positivem Kriteriumswert. Es verbleiben 30 Fälle mit negativem Kriteriumswert im Modell. Die Entscheidung für die erstgenannten 70 Fälle wird also irrespektive der Cuewerte für die anderen drei Cues gefällt, acht Cueprofile sind im ersten Schritt positiv klassifiziert worden. Im zweiten Schritt werden 28 Objekte mit negativem Kriteriumswert negativ klassifiziert, was vier Cueprofile beinhaltet. Es verbleiben zwei Objekte, von denen ein Objekt im dritten und das letzte im vierten Schritt klassifiziert wird. Wird die hierarchische Darstellung an die von einem FaF Tree verwendete Permutation der Cues in beschriebener Weise angepasst, lässt sich zwischen den Profilen in der letzten Ebene immer eine Grenze bestimmen, die zwei Teilgruppen mit identischer Entscheidung voneinander trennt. Das Profil mit positiver Entscheidung, das sich am weitesten rechts befindet, heißt auch *Splitting Profile* und kann zur Charakterisierung eines FaF Trees verwendet werden (vgl. Abschnitt 5.5).

5.4.3 Beispielergebnisse für die Konstruktionsverfahren

Tabelle 5.5: Ergebnisse der fünf Bäume für Verteilung 1 aus Tabelle 5.4

	FaF MaxSens	FaF MaxVal	FaF ZigSens	FaF ZigVal	FaF MaxAcc
Schritt 1 korrekt:falsch	40:30	30:0	40:30	25:5	55:15
Schritt 2 korrekt:falsch	28:0	25:3	28:0	28:0	25:3
Schritt 3 korrekt:falsch	1:0	16:11	0:1	4:11	1:0
Schritt 4 korrekt:falsch	0:1	9:6	1:0	18:9	0:1
Krit. + /Präd. +	40	26	40	35	25
Krit. + /Präd. -	0	14	0	5	15
Krit. - /Präd. +	31	6	31	20	4
Krit. - /Präd. -	29	54	29	40	56
Klassifikationsgenauigkeit	0,690	0,800	0,690	0,750	0,810
Sensitivität	1,000	0,650	1,000	0,875	0,625
Spezifität	0,483	0,900	0,483	0,667	0,933
Δ_{UAPV}-Score	0,483	0,550	0,483	0,542	0,558
Ø Cues genutzt	1,330	2,270	1,330	1,330	1,300

Die Ergebnisse aller fünf Verfahren für die erste Verteilung sind in Tabelle 5.5 zusammengefasst. Die Tabelle zeigt die in jedem Schritt korrekt und falsch klassifizierten Fälle, die Klassifikationsgenauigkeit und Δ_{UAPV}-Score des Gesamtbaums, sowie Sensitivität und Spezifität. Weiterhin ist die durchschnittliche Anzahl von Cues angegeben, deren Werte für eine Entscheidung abgefragt wurden. Für den ersten Baum erweist sich der FaF MaxAcc-Algorithmus als der beste *Fitting*-Algorithmus, gefolgt von FaF MaxVal und FaF ZigVal. Die beiden übrigen Algorithmen zeigen allerdings eine höhere Sensitivität, so dass kein Verfahren von einem anderen dominiert wird. Die Relationen zwischen den Verfahren sind auch im ROCS-

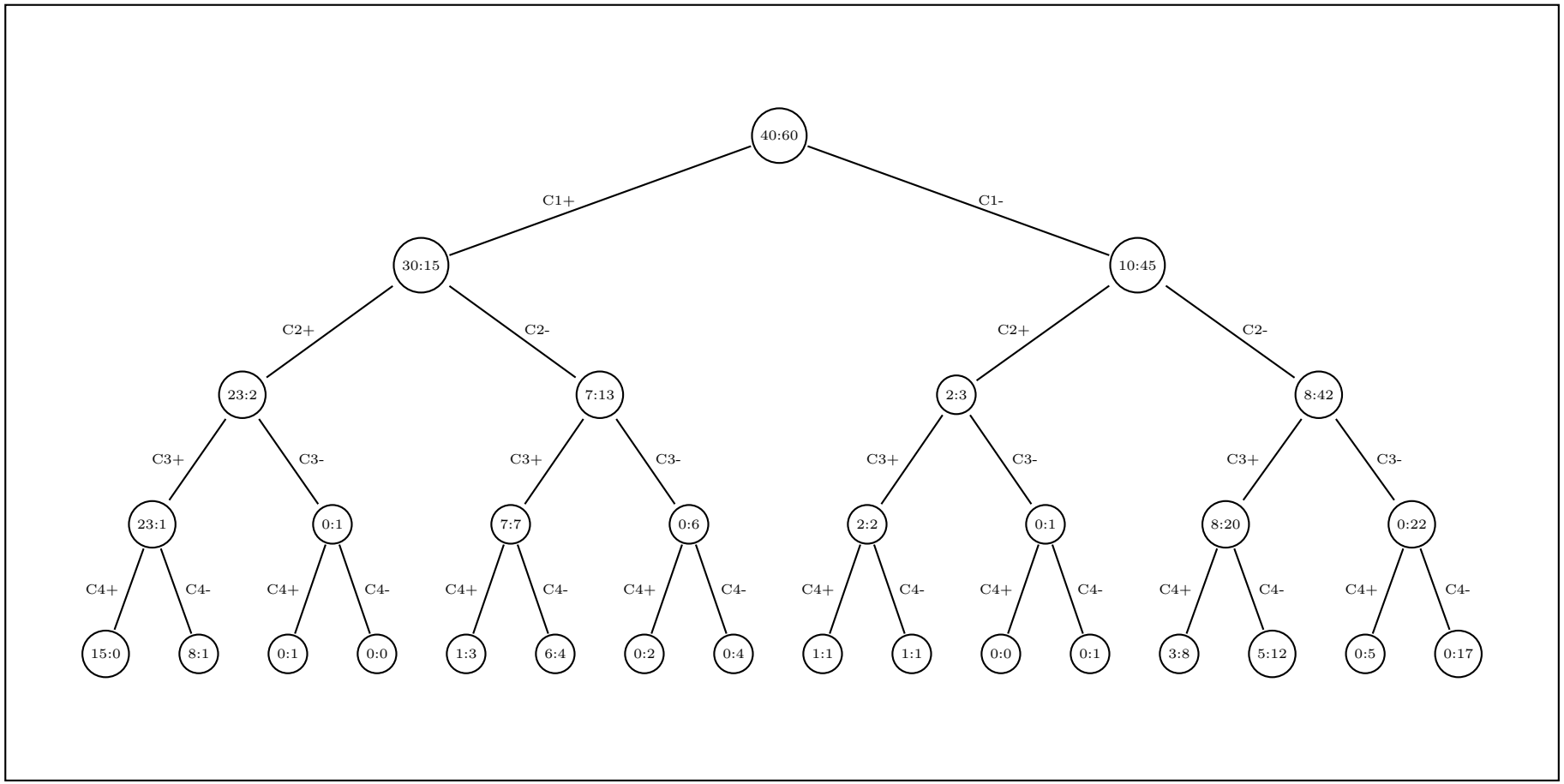


Abbildung 5.8: Hierarchischer Cueverteilungsbaum für Verteilung 1 und die Cuereihenfolge (1,2,3,4): Der erste Wert gibt die Anzahl von Objekten mit negativem Kriteriumswert, der zweite Wert die Anzahl von Objekten mit positivem Kriteriumswert an.

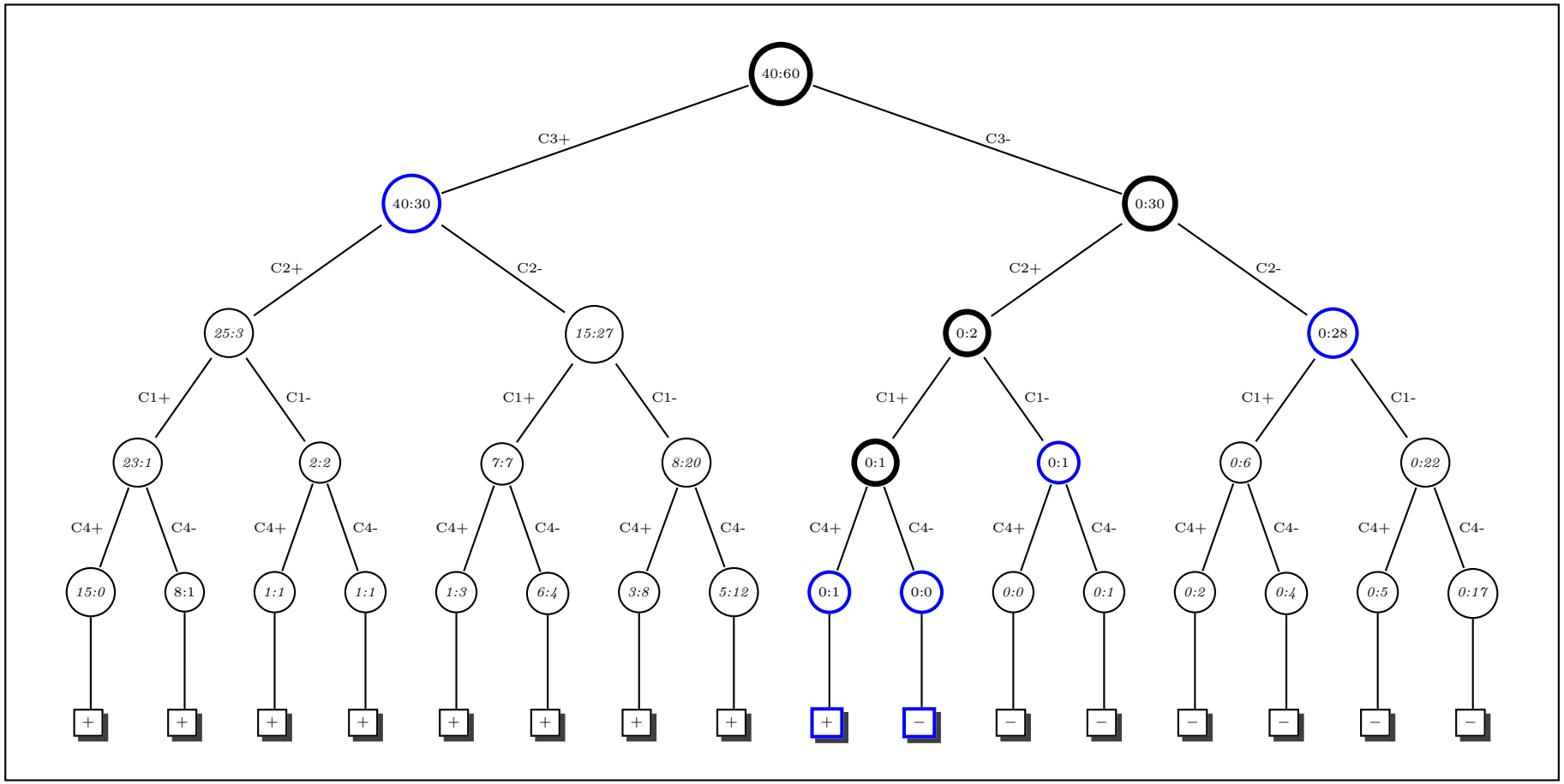


Abbildung 5.9: Hierarchischer Cueverteilungsbaum mit Profilentscheidungen für Verteilung 1 und FaF MaxSens mit der Cuereihenfolge (3,2,1,4): Der erste Wert gibt die Anzahl von Objekten mit negativem Kriteriumswert, der zweite Wert die Anzahl von Objekten mit positivem Kriteriumswert an. Die blau umrandeten Knoten entsprechen Entscheidungsknoten im FaF MaxSens-Baum, die schwarz umrandeten sind Knoten, die das Splitting Profile charakterisieren.

Diagramm zu verfolgen (s. Abbildung 5.10).

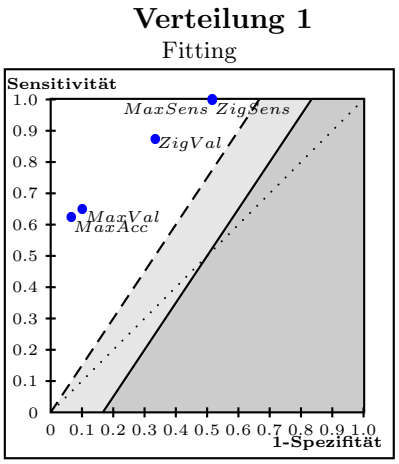


Abbildung 5.10: ROCS-Diagramm für die FaF Trees für Verteilung 1

Die beiden übrigen Verteilungen wurden gewählt, um zu demonstrieren, dass die Rangfolge der Verfahren nicht allein aufgrund der zur Konstruktion verwendeten Informationen prognostiziert werden kann und dementsprechend kein Verfahren a priori aus der Menge der sinnvollen Verfahren ausgeschlossen werden kann: Für Verteilung 2 sind die Ergebnisse in Tabelle 5.6 zusammengestellt und in Abbildung 5.11 visualisiert. Für diese Verteilung zeigt das Ergebnis des FaF ZigVal eine höhere Spezifität bei niedrigerer Sensitivität.

Tabelle 5.6: Ergebnisse der fünf Bäume für Verteilung 2 aus Tabelle 5.4

	FaF MaxSens	FaF MaxVal	FaF ZigSens	FaF ZigVal	FaF MaxAcc
Schritt 1 korrekt:falsch	40:30	30:0	40:30	25:5	55:15
Schritt 2 korrekt:falsch	30:0	25:5	30:0	30:0	25:5
Schritt 3 korrekt:falsch	0:0	13:15	0:0	0:12	0:0
Schritt 4 korrekt:falsch	0:0	6:6	0:0	21:7	0:0
Krit. + /Präd. +	40	25	40	35	25
Krit. + /Präd. -	0	15	0	19	15
Krit. - /Präd. +	30	11	30	5	5
Krit. - /Präd. -	30	49	30	41	55
Klassifikationsgenauigkeit	0,700	0,740	0,700	0,760	0,800
Sensitivität	1,000	0,625	1,000	0,648	0,625
Spezifität	0,500	0,817	0,500	0,891	0,917
Δ_{UPAV}-Score	0,500	0,442	0,500	0,539	0,542
Ø Cues genutzt	1,300	2,220	1,300	2,380	1,300

Zuletzt zeigt Tabelle 5.7 die Ergebnisse für Verteilung 3. Bei dieser Verteilung dominiert das FaF ZigVal-Verfahren die übrigen Konstruktionsstrategien. Eine Analyse der Verteilung zeigt zudem, dass kein Verfahren eine bessere Klassifikationsgenauigkeit aufweisen kann als der konstruierte Baum, da für das Profil „+++-“ 15 Objekte mit positivem und ein Objekt mit negativem Kriteriumswert vorhanden sind und für das Profil „++++“ 10 Objekte mit positivem und vier Objekte mit negativem Kriteriumswert, weshalb die fünf Objekte mit negativem Kriteriumswert von jedem Verfahren, dass allein aufgrund der Cueprofile entscheidet, falsch klassifiziert werden, wenn nicht ein größerer Fehler in Kauf genommen werden soll. Das Beispiel demonstriert damit plastisch, dass die Verwendung von Konstruktionsheuristiken zwar nicht zu optimalen Baumstrukturen führen muss, aber - wie hier für das *Fitting*-Problem gezeigt - durchaus führen kann.

Im ROCS-Diagramm sind die drei Ergebnisse des FaF ZigVal-Verfahrens als Ecken eines Dreiecks eingezeichnet, um die für drei Verteilungen beobachtete Streuung der Ergebniswerte zu markieren. Die Verteilungen sind nicht zufällig gewählt worden, in jenem Fall wären geringere Schwankungen wahrscheinlicher gewesen. Gleichzeitig zeigt der Vergleich der durchschnittlich abgefragten Cuewerte, dass für das FaF ZigVal-Verfahren in den unterschiedlichen Verteilungen zwischen 1,330 und 2,800 Cuewerte abgefragt werden, dementsprechend ist auch die Geschwindigkeit und Frugalität der Bäume im Allgemeinen nicht unabhängig von der Verteilung der Cueprofile. Ein besonderer Augenmerk verdient die beobachtete Konstanz der Ergebnisse von FaF ZigSens und FaF MaxSens. Ein zweiter Blick kann allerdings die Hoffnung nicht stützen, eine generelle Invarianz identifiziert zu haben. Für die gewählte Verteilung der Kennwerte

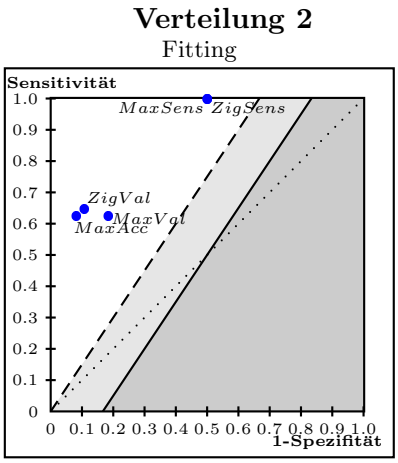


Abbildung 5.11: ROCS-Diagramm für die FaF Trees für Verteilung 2

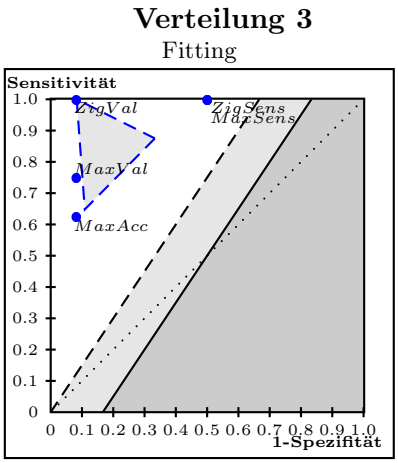


Abbildung 5.12: ROCS-Diagramm für die FaF Trees für Verteilung 3: Die Ecken des Dreiecks markieren die drei Ergebnisse für das FaF ZigVal-Verfahren für die drei Verteilungen.

Tabelle 5.7: Ergebnisse der fünf Bäume für Verteilung 3 aus Tabelle 5.4

	FaF MaxSens	FaF MaxVal	FaF ZigSens	FaF ZigVal	FaF MaxAcc
Schritt 1 korrekt:falsch	40:30	30:0	40:30	25:5	55:15
Schritt 2 korrekt:falsch	30:0	25:5	30:0	30:0	25:5
Schritt 3 korrekt:falsch	0:0	25:0	0:0	10:0	0:0
Schritt 4 korrekt:falsch	0:0	0:10	0:0	30:0	0:0
Krit. + /Präd. +	40	30	40	40	25
Krit. + /Präd. -	0	10	0	0	15
Krit. - /Präd. +	30	5	30	5	5
Krit. - /Präd. -	30	55	30	55	55
Klassifikationsgenauigkeit	0,700	0,850	0,700	0,950	0,800
Sensitivität	1,000	0,750	1,000	1.000	0,625
Spezifität	0,500	0,917	0,500	0,917	0,917
Δ_{UPAV}-Score	0,500	0,667	0,500	0,917	0,542
ØCues genutzt	1,300	2,050	1,300	2,800	1,300

müssen die Ergebnisse allerdings aus folgendem Grund konstant bleiben: Beide Baumprofile beginnen mit dem Teilbaum $(3+; 2-, \ldots)$. Im ersten Schritt werden unabhängig von der Verteilung 70 Objekte klassifiziert, dies ergibt sich aus der durch die Kennwerte bestimmten Invarianz der Cuewertverteilung. Dabei werden alle Objekte mit positivem Kriteriumswert im ersten Schritt klassifiziert, da es kein einziges Objekt mit negativem Kriteriumswert und negativem Cuewert für Cue 3 gibt. Die Kreuztabelle für Cue 2 und das Kriterium zeigt, dass weiterhin maximal fünf Objekte im zweiten Schritt fehlklassifiziert werden könnten. Aus dieser Betrachtung folgt, dass die Ergebnisvarianz der Konstruktionsprinzipien abhängig von der Verteilung der Objekte auf Cueprofile bei Konstanthaltung der Cuewert-Kriteriums-Kreuztabellen als im Allgemeinen unterschiedlich erwartet werden muss. Die formale Analyse der möglichen Verteilungen würde allerdings zunächst keine generellen Aussagen über die Konstruktionsverfahren unabhängig von den Kreuztabellen machen können. Trotzdem scheint diese Betrachtung einen interessanten Analyseweg für die Abschätzung der Robustheit von Lösungen zu bieten.

Tabelle 5.8: Verteilung der Cueprofile auf die Entscheidungsschritte der fünf Bäume: für jedes Cueprofil ist in der mit dem Namen des Konstruktionsverfahrens betitelten Spalte der Schritt angegeben, in dem das Cueprofil von dem erstellten Baum klassifiziert wird sowie das Ergebnis der Klassifikation. Cueprofile, für die durch unterschiedliche Bäume verschiedene Klassifikationen vorgenommen werden, sind durch ein „x“ in der letzten Spalte markiert.

Cue 1	Cue 2	Cue 3	Cue 4	MaxSens	MaxVal	ZigSens	ZigVal	MaxAcc	Konflikte
-	-	-	-	2-	1-	2-	2-	1-	
-	-	-	+	2-	1-	2-	2-	1-	
-	-	+	-	1+	3-	1+	1+	1-	x
-	-	+	+	1+	4-	1+	1+	1-	x
-	+	-	-	3-	1-	4-	4-	3-	
-	+	-	+	3-	1-	3+	3+	4-	x
-	+	+	-	1+	2+	1+	1+	2+	
-	+	+	+	1+	2+	1+	1+	2+	
+	-	-	-	2-	1-	2-	1+	1-	x
+	-	-	+	2-	1-	2-	1+	1-	x
+	-	+	-	1+	3-	1+	2-	1-	x
+	-	+	+	1+	4+	1+	2-	1-	x
+	+	-	-	4-	1-	4+	4+	3-	x
+	+	-	+	4+	1-	3+	3+	4+	x
+	+	+	-	1+	2+	1+	1+	2+	
+	+	+	+	1+	2+	1+	1+	2+	

Zuletzt sind in Tabelle 5.8 die Entscheidungen der fünf Bäume für alle Cueprofile aufgeführt, um Cueprofile zu identifizieren, die differenziell klassifiziert werden. Dabei ist die Entscheidung der Bäume für sieben Cueprofile identisch, bei drei Cueprofilen weicht ein Baum von der Klassifikation der übrigen vier ab, und bei sechs Cueprofilen klassifizieren zwei Bäume anders als die übrigen drei. Für die gegebene Verteilung sind einige dieser Diskrepanzen irrelevant, da verschiedene Cueprofile nicht auftreten. Bei Fällen außerhalb der Stichprobe, die zur Konstruktion genutzt wurde, können diese Cueprofile allerdings auftreten, was plastisch die Gefahr des *Overfitting* demonstriert.

5.4.4 Einige Unterschiede zwischen den Strategien

Zu den vorgeschlagenen Verfahren lässt sich in einer ersten Annäherung Folgendes anmerken: ZigZag-Algorithmen nehmen eine Strukturinvarianz in Kauf, um einer bestimmten Eigenschaft von Daten gerecht werden zu können: der Zusammensetzung von Stichproben aus positiv und negativ zu klassifizierenden Fällen. Sind zahlreiche Cues gegeben und liegt die positive Validität der Cues bei allen Cues höher als die negative, würde ein MaxVal-Baum nur sehr wenige Fälle in die negative Kategorie sortieren. Nur eine einzige Kombination von Cuewerten würde negativ klassifiziert, und es wäre sogar möglich, dass kein Fall in der Stichprobe dieser Kategorie angehört. Der Vorteil der ZigZag-Bäume ist also die Sicherstellung einer gewissen Ausgewogenheit der Klassifikation. Die allein auf Maxima beruhenden Ordnungen versuchen Cues in der für die Gesamtstichprobe optimalen Richtung zu nutzen. Jeder Cue kann entweder positiv oder negativ genutzt werden, deshalb muss eine Entscheidung für eine der beiden Richtungen getroffen werden. ZigZag-Bäume stellen nicht sicher, dass ein Cue in diesem Sinne „optimal“ genutzt wird.

Der Unterschied zwischen der Heranziehung von Validität und Sensitivität/Spezifität liegt in der Ver-

änderung des Bezugrahmens für Treffer und Fehler beider Arten. Validität ist dabei direkt mit der Funktionsweise des Cues im Baum verbunden: Wird ein Cue für eine Positiventscheidung im ersten Schritt verwendet, entspricht die positive Validität dem Anteil korrekt klassifizierter Fälle im ersten Schritt. Wird im ersten Schritt negativ klassifiziert, ist die negative Validität die Bezugsgröße. Für weitere Schritte wird die bereits erfolgte Klassifikation von Fällen für die Bestimmung der Reihenfolge ignoriert, ähnlich wie bei TTB werden unkonditionale Validitäten herangezogen, weshalb über die Klassifikationsleistung im zweiten Schritt nur dann eine Aussage aufgrund der Validität getroffen werden kann, wenn der Baum nicht erneut der gleichen Klasse zuordnet. Ist in jede Klasse bereits einmal klassifiziert worden, kann nur bei Gültigkeit der Unabhängigkeit der Cue-Kriteriums-Verteilungen die Leistung ermittelt werden.

Die Sachlage liegt bei einer Berücksichtigung von Sensitivitäts-/Spezifitätswerten anders: Bezugsrahmen für die Beurteilung ist hier nicht die Teilgruppe der Objekte mit gleichem Cuewert („sichtbar“ für den Entscheidungsbaum), sondern die Teilgruppe der Objekte mit gleicher Klassenzugehörigkeit („unsichtbar“ für den Entscheidungsbaum). Ein Cue wird dann positiv bewertet, wenn er einen großen Teil einer Klasse korrekt zuordnet, unabhängig von der Anzahl der Fehler, die in einem Entscheidungsschritt gemacht werden. Gibt es beispielsweise 5 Objekte mit positiver Klassenzugehörigkeit und 95 Objekte mit negativer, dann würde ein Cue, der alle Objekte in die negative Klasse einordnet, als Cue mit maximaler Spezifität bewertet.

Bereits bei dieser Betrachtung wird deutlich, dass nicht eindeutig bestimmt werden kann, welches das beste Verfahren zur Konstruktion von Bäumen darstellt. Als ein Problemfeld stellt sich beispielsweise die Cuewertverteilung pro Cue und die Verteilung der Kriteriumswerte heraus: Bei einer Verteilung der Cuewerte, die von einer Gleichverteilung substanziell abweicht, wird jeweils einer der beiden Validitätswerte auf der Grundlage einer sehr kleinen, der andere auf der Grundlage einer sehr großen Stichprobe ermittelt. Damit sollten sich die Werte bei einer Anwendung auf neue Stichproben als sehr unterschiedlich robust erweisen, die Streuung der einen Art von Validitätswerten wird höher. Damit steigt die Wahrscheinlichkeit, dass die ersten (bzw. die letzten) Positionen an Cues vergeben werden, die in der Gesamtpopulation weiter hinten (bzw. weiter vorn) liegen. Analog hat eine Ungleichverteilung der Kriteriumswerte differenzielle Konsequenzen für die Robustheit der Sensitivitäts- und Spezifitätsschätzung.

Daraus ergibt sich als eine erste Empfehlung die folgende: Der Anwendungsnutzen der auf Sensitivität und Spezifität beruhenden Verfahren sinkt mit der Abweichung der Kriteriumswertverteilung von der Gleichverteilung, der Anwendungsnutzen der auf Validität basierenden Verfahren sinkt mit der Abweichung der Cuewertverteilung von der Gleichverteilung. Die Variante MaxAcc wurde hinzugefügt, um die Bewertung eines Cues von der gesamten Stichprobe abhängig machen zu können. Der Nachteil der Methode ist die Gleichbehandlung beider Fehlerklassen, obwohl für Cues im Baum nur eine der beiden Fehlerklassen relevant ist. Diese Überlegungen sollen mit Hilfe des folgenden Beispiels vertieft werden.

5.4.5 Herausforderungen für die Strategien

Tabelle 5.9: Beispielergebnisse für vier Cues bei stark unterschiedlichen Klassengrößen

	Cue + / Krit. +	Cue - / Krit. +	Cue + / Krit. -	Cue - / Krit. -
Cue 1	8	2	30	60
Cue 2	6	4	40	50
Cue 3	5	5	10	80
Cue 4	9	1	45	45

Tabelle 5.10: Gütekriterien für vier Cues aus Tabelle 5.9

	Sensitivität	Spezifizität	Positive Validität	Negative Validität	Genauigkeit
Cue 1	**0,800**	*0,667*	*0,211*	**0,968**	0,680
Cue 2	**0,600**	*0,556*	*0,130*	**0,923**	0,560
Cue 3	*0,500*	**0,889**	*0,333*	**0,941**	0,850
Cue 4	**0,900**	*0,500*	*0,167*	**0,978**	0,540

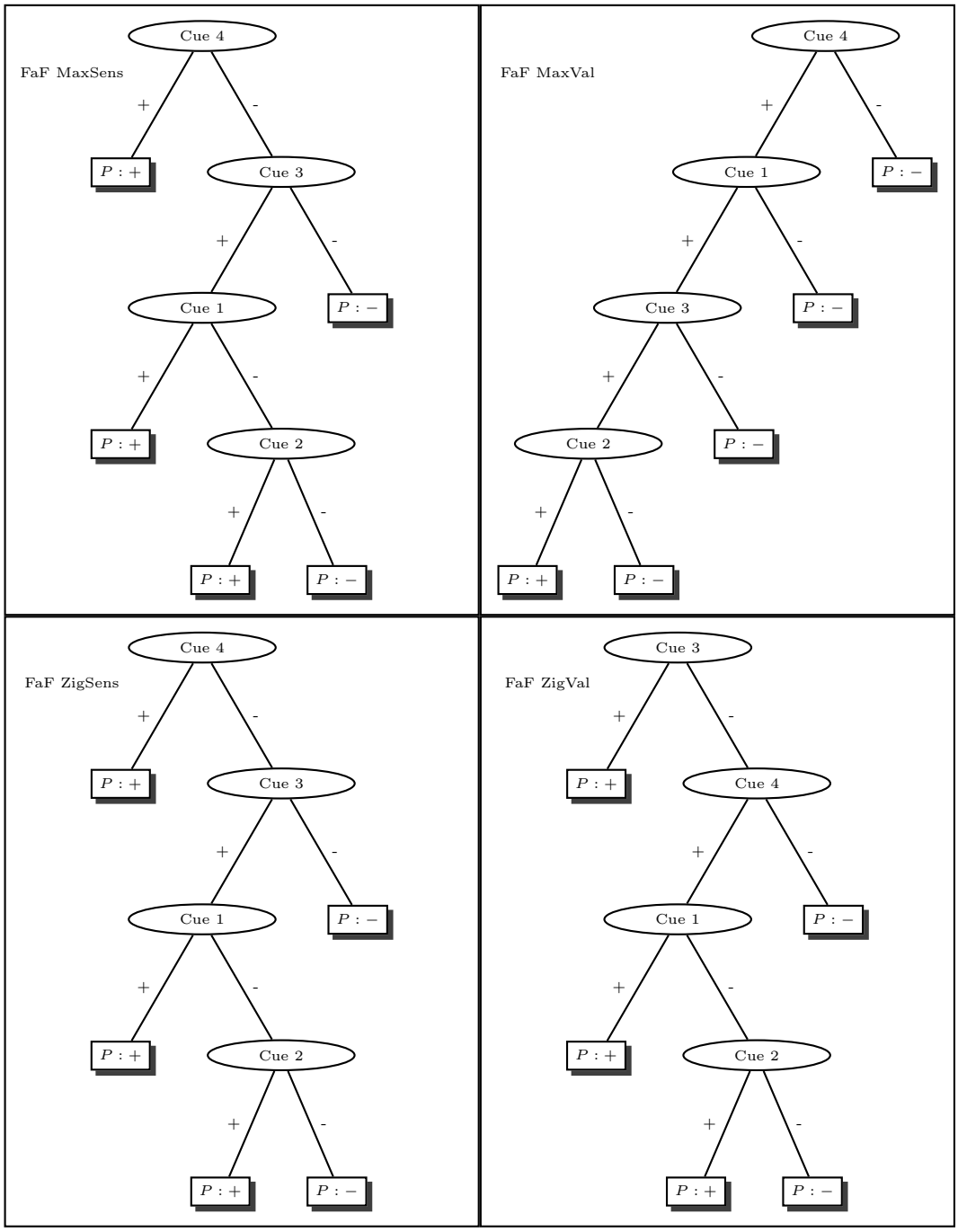


Abbildung 5.13: Anwendung der Konstruktionsregeln Faf MaxSens, Faf MaxVal, FaF ZigSens und FaF ZigVal auf die fiktiven Cues aus Tabelle 5.9

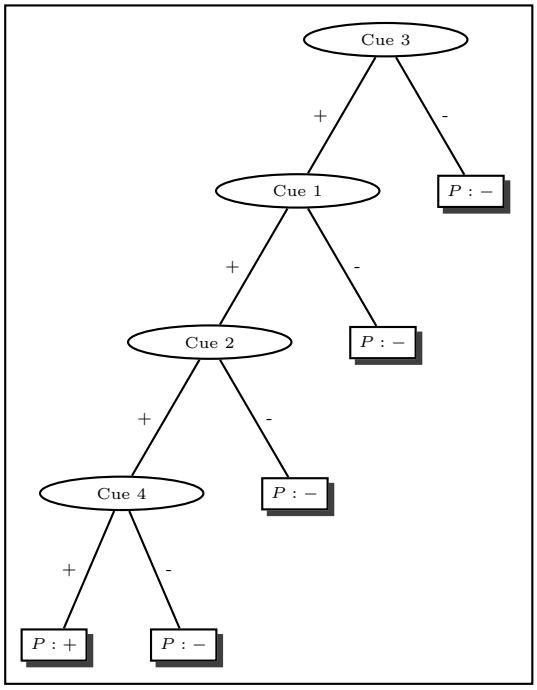


Abbildung 5.14: Anwendung der Konstruktionsregel Faf MaxAcc auf die fiktiven Cues aus Tabelle 5.9

An einem zweiten Beispiel sollen einige weitere Komplikationen erörtert werden. Gegeben sei die in Tabelle 5.2 gezeigte Verteilung von Cuewerten und Kriteriumswerten mit den daraus resultierenden Cuekennwerten in Tabelle 5.10. Der Hauptunterschied zum ersten Beispiel liegt in der starken Ungleichverteilung der beiden Kriteriumsklassen. Die durch die vier auf Validität oder Sensitivität/Spezifität beruhenden Baumstrukturen, die auf der Grundlage dieser Daten konstruiert werden können, sind in Abbildung 5.13 präsentiert, die von FaF MaxAcc erzeugte Struktur in Abbildung 5.14.

Auch für diese Gütezahlen soll an einer Beispielverteilung die Klassifikationsgenauigkeit der erstellten Bäume überprüft werden (s. Tabelle 5.11): Die Ergebnisse der Algorithmen sind in Tabelle 5.12 zusammengestellt, das entsprechende ROCS-Diagramm zeigt Abbildung 5.15. Besonders auffallend ist bei diesem Beispiel das schlechte Ergebnis der auf Sensitivität und Spezifität beruhenden Bäume, vor allem in Bezug auf die Klassifikationsgenauigkeit. Dies liegt begründet in der Diskrepanz zwischen der Art der Güteschätzung für einen Cue und seiner Verwendung im Klassifikationsbaum: Die Sensitivität der Cues wird im Beispiel anhand einer Stichprobe von lediglich zehn Objekten geschätzt, dabei werden Objekte mit negativem Kriteriumswert ignoriert. Der Baum hingegen klassifiziert alle Objekte mit positivem Cuewert positiv, d.h. auch fälschlicherweise eine beträchtliche Anzahl von Objekten mit negativem Kriteriumswert. Diese Anzahl ist für die Berechnung der Sensitivität unsichtbar. Im Gegenzug ist natürlich die Berechnung der Spezifität auf eine stärkere Grundlage gestellt, hier dienen die 90 Objekte mit negativem Kriteriumswert als Grundlage. Aus den angestellten Überlegungen erwachsen zwei Arten von Problemen für die beiden Konstruktionsverfahren:

Zum einen wird bei ungleichen Stichprobengrößen die Hälfte der Bestimmungen von Gütekriterien auf eine sehr kleine Stichprobe gestützt. Die Stabilität der gewonnenen Ergebnisse ist dabei sehr fraglich, wenn der gewonnene Baum zur Klassifikation unbekannter Objekte herangezogen werden soll. Zum anderen ist die Streuung der die kleinere Klasse betreffenden Gütewerte in der Erwartung größer als die Streuung der die kleinere Klasse betreffenden Gütewerte. Dieser Umstand erzeugt einen Bias des Verfahrens: Angenommen es stehen Cues zur Verfügung, die in der Population identische Sensitivität und Spezifität aufweisen. In der Schätzung der Werte aufgrund empirischer Stichproben würden die ersten und die letzten Plätze der Rangfolge durch die auf der Stichprobengröße beruhenden Streuungsdifferenz tendenziell durch Werte der auf die kleinere Teilgruppe bezogenen Gütekriterien gestellt. Im FaF MaxSens-Verfahren hat das zur Konsequenz, dass die ersten Klassifikationen in Richtung der kleineren Klasse vorgenommen werden. Da auf den ersten Ebenen die Mehrheit der Fälle klassifiziert wird, stehen nicht genug Objekte mit entsprechendem Kriteriumswert zur Verfügung, um eine fehlerfreie Klassifikation vornehmen zu können. Dementsprechend wird eine beträchtliche Anzahl von Fehlklassifikationen

Tabelle 5.11: Mögliche Verteilung mit den Ergebnissen aus Tabelle 5.9

Cue 1	Cue 2	Cue 3	Cue 4	Krit.	Verteilung 4
-	-	-	-	-	13
-	-	-	-	+	0
-	-	-	+	-	18
-	-	-	+	+	0
-	-	+	-	-	2
-	-	+	-	+	0
-	-	+	+	-	1
-	-	+	+	+	1
-	+	-	-	-	8
-	+	-	-	+	1
-	+	-	+	-	13
-	+	-	+	+	0
-	+	+	-	-	5
-	+	+	-	+	0
-	+	+	+	-	0
-	+	+	+	+	0
+	-	-	-	-	9
+	-	-	-	+	0
+	-	-	+	-	5
+	-	-	+	+	1
+	-	+	-	-	0
+	-	+	-	+	0
+	-	+	+	-	2
+	-	+	+	+	2
+	+	-	-	-	6
+	+	-	-	+	0
+	+	-	+	-	8
+	+	-	+	+	3
+	+	+	-	-	0
+	+	+	-	+	0
+	+	+	+	-	0
+	+	+	+	+	2

automatisch in Kauf genommen.

Bei der ZigZag-Variante des Algorithmus kann das Problem nicht in gleicher Stärke auftreten, da das Alternieren zu einer ausgewogeneren Klassifikation führen muss. Allerdings wird im ersten Schritt dann ein Fehler in Kauf genommen, wenn die kleinere Klasse aus Objekten mit positivem Kriteriumswert gebildet wird. Das Problem der ungleichen Basis für die Schätzungen der Gütekriterien stellt sich bei diesem Algorithmus allerdings weiterhin. Das positive Ergebnis des Algorithmus FaF MaxVal sollte nicht darüber hinwegtäuschen, dass es nur erreicht werden konnte, indem fast alle Objekte identisch klassifiziert wurden. Sind die Klassen ungleich verteilt, zeigt einer der beiden Validitätstypen tendenziell, bei extremer Ungleichverteilung konsistent höhere Werte als der andere, wenn nicht die Cuewertverteilung der Kriteriumswertverteilung entspricht. Werden die Hälfte der Objekte positiv klassifiziert und sind nur 10% der Objekte aus der positiven Kategorie, stellt 0,2 die obere Grenze für die positive Validität und 0,8 die untere Grenze für die negative Validität dar, unabhängig vom inhaltlichen Zusammenhang zwischen Cue und Kriterium. Bei starker Ungleichverteilung sind Pectinates als Ergebnis von FaF MaxVal deshalb keine Seltenheit. Für die alternierende Variante FaF ZigVal wird verhindert, dass sich die Struktur wie oben beschrieben ausformt. Dies geht allerdings auf Kosten der Klassifikationsgenauigkeit, vor allem weil im ersten Schritt eine positive Entscheidung forciert wird. Ähnlich wie bei FaF ZigSens kann die streng alternierende Struktur nicht jeder Kriteriumswertverteilung gerecht werden.

Auch der Faf MaxAcc-Algorithmus erzielt im Beispiel sehr gute Ergebnisse. Die Stärke des Verfahrens liegt sicherlich daran, dass die gesamte Stichprobe für die Bestimmung der Cuegüte herangezogen wird.

Tabelle 5.12: Ergebnisse der fünf Bäume für Verteilung 4 aus Tabelle 5.11

	FaF MaxSens	FaF MaxVal	FaF ZigSens	FaF ZigVal	FaF MaxAcc
Schritt 1 korrekt:falsch	9:47	47:9	9:47	5:10	80:5
Schritt 2 korrekt:falsch	36:1	28:1	36:1	36:1	8:1
Schritt 3 korrekt:falsch	0:0	15:0	0:0	4:13	2:2
Schritt 4 korrekt:falsch	2:5	0:0	2:5	18:13	2:0
Krit. + /Präd. +	9	0	9	9	2
Krit. + /Präd. -	1	10	1	1	8
Krit. - /Präd. +	52	0	52	36	0
Krit. - /Präd. -	38	90	38	54	90
Klassifikationsgenauigkeit	0,470	0,900	0,470	0,630	0,920
Sensitivität	0,900	0,000	0,900	0,900	0,200
Spezifität	0,422	1,000	0,422	0,600	1,000
Δ_{UPAV}-Score	0,322	0,000	0,322	0,500	0200
øCues genutzt	1,580	1,590	1,580	2,640	1,230

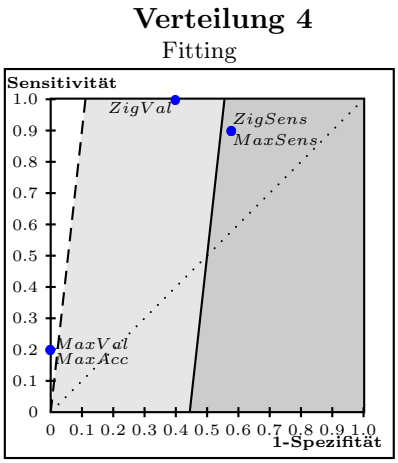


Abbildung 5.15: ROCS-Diagramm für die FaF Trees für Verteilung 4

Ähnlich wie das Zig MaxVal-Verfahren ist das Ergebnis verzerrt in Richtung der Vorhersage der größeren Klasse. Ein weiteres Problem muss zudem berücksichtigt werden: Ein Cue kann prinzipiell nur in einer der beiden Richtungen zum Einsatz kommen, zudem wird zur Bestimmung der Güte des Cues eine Teilmenge von Fällen mit herangezogen, die für die Entscheidung im Baum nicht relevant ist.

5.4.6 Konstruktion neuer Konstruktionsverfahren

Aus den angestellten Überlegungen ergeben sich einige Ideen für modifizierte Strategien. Es sollen sechs Modifikationen von Zigzag-Verfahren vorgestellt werden, die einige der dargestellten Probleme berücksichtigen können. Des weiteren sollen einige weitere Konstruktionsverfahren präsentiert werden, die als Anregungen für weitere Studien verstanden werden können.

Die beiden dargestellten ZigZag-Verfahren lassen sich zunächst auf zwei natürlichen Wegen modifizieren. Zunächst kann vereinbart werden, dass der erste ausgewählte Cue nicht automatisch eine positive Entscheidung treffen muss. Alternativ könnte der Cue und die Entscheidungsrichtung analog zum ersten Schritt der maximierenden Verfahren ausgewählt werden: Es wird der Cue gewählt mit dem Maximum aus beiden Kennzahlen. Basiert die Entscheidung auf positiver Validität bzw. Sensitivität, wird der Cue mit einer positiven Entscheidung verbunden, basiert die Entscheidung auf negativer Validität bzw. Spezifität, wird der Cue mit einer negativen Entscheidung verbunden. Die so veränderten Verfahren sollen *FaF ZigSens Plus* und *FaF ZigVal Plus* heißen.

Um das Größenproblem direkt anzugehen, könnte die Cuerichtung auch automatisch so festgelegt werden, dass die erste Entscheidung für die Klasse mit größerem Anteil an der Stichprobe getroffen wird. Als erster Cue wird dementsprechend der Cue ausgewählt, dessen Gütewert für diese Entscheidungsrichtung maximal ist. Die beiden neuen Verfahren heißen *FaF ZigSens Majority* und *FaF ZigVal Majority*.

Bei extremer Ungleichverteilung zwischen den Klassen reicht ein einzelner Klassifikationsschritt allerdings oft nicht aus, um Ausgewogenheit herzustellen. Für den ersten Schritt kann die Veränderung der Klassengrößen aus Cuewertverteilung und entsprechender Validität berechnet werden, für die folgenden Schritt nur unter Annahme von konditionaler Unabhängigkeit der Cuewertverteilungen und Validitäten.

Für eine exakte Berechnung müssten die konditionalen Validitäten berechnet werden. Eine Annäherung an die notwendige Anzahl von Klassifikationsschritten ergibt sich aus folgender Vereinfachung: Angenommen, die Cuewerte sind ungefähr gleichverteilt und die Entscheidungen sind perfekt. Gesucht ist die Anzahl von Entscheidungen in Richtung der größeren Klasse, die notwendig ist, bevor sich die Mehrheitsverhältnisse ändern. Ist n_k die Anzahl der Objekte in der kleineren Klasse und n_g die Anzahl der Objekte in der größeren Klasse, ergibt sich die Anzahl notwendiger Entscheidungen unter den genannten Annahmen und bei konditionaler Unabhängigkeit als

$$s = \left\lfloor -\log_2 \frac{n_k}{n_k + n_g} \right\rfloor \tag{5.2}$$

Die Grenzen für die Schrittzahl ergeben sich damit als $\frac{1}{2}$ für einen Schritt, $\frac{1}{4}$ für zwei Schritte, $\frac{1}{8}$ für drei Schritt etc. Diese Grenzen lassen sich auch intuitiv begreifen: Angenommen ein Papierstreifen ist in zwei Farben geteilt und wird wiederholt in der Mitte gefaltet. Dabei entspricht die Anzahl von Faltungen, die den kleineren Teilstreifen unberührt lassen, der Anzahl von Schritten, bis sich die Mehrheitsverhältnisse ändern.

Zwei Modifikationen des ZigZag-Algorithmus ergeben sich durch folgende Regel: Der Algorithmus wählt zunächst k Cues, die in Richtung der Mehrheitskategorie klassifizieren anhand der diesbezüglichen Gütekriterien aus und alterniert zwischen den Klassen erst ab dem $k+1$-ten Cue. Mit Referenz auf die Faltanalogie werden die Algorithmen als *FaF ZigSens Fold* und *FaF ZigVal Fold* bezeichnet.

Tabelle 5.13: Ergebnisse und Struktur der sechs neuen Bäume für Verteilung 4 aus Tabelle 5.11

	ZigVal Plus	ZigVal Majority	ZigVal Fold	ZigSens Plus	ZigSens Majority	Zig Sens Fold
Schritt 1 korrekt:falsch	43:1	43:1	43:1	9:47	80:5	80:5
Schritt 2 korrekt:falsch	5:3	5:3	32:1	36:1	5:3	8:1
Schritt 3 korrekt:falsch	31:0	31:0	13:4	0:0	7:0	2:2
Schritt 4 korrekt:falsch	8:9	8:9	4:2	2:5	0:0	2:0
Krit. + /Präd. +	8	8	2	9	5	2
Krit. + /Präd. -	2	2	8	1	5	8
Krit. - /Präd. +	11	11	0	52	3	0
Krit. - /Präd. -	79	79	90	38	87	90
Klassifikationsgenauigkeit	0,870	0,870	0,920	0,470	0,920	0,920
Sensitivität	0,800	0,800	0,200	0,900	0,500	0,200
Spezifität	0,878	0,878	1,000	0,422	0,967	1,000
Δ_{UPAV}-Score	0,678	0,678	0,200	0,322	0,467	0,200
ØCues genutzt	2,210	2,210	1,850	1,580	1,220	1,230
Baumstruktur	(4-,3+,1-,2)	(4-,3+,1-,2)	(4-,1-,3-,2)	(4+,3-,1+,2)	(3-,4+,1-,2)	(3-,1-,2-,4-)

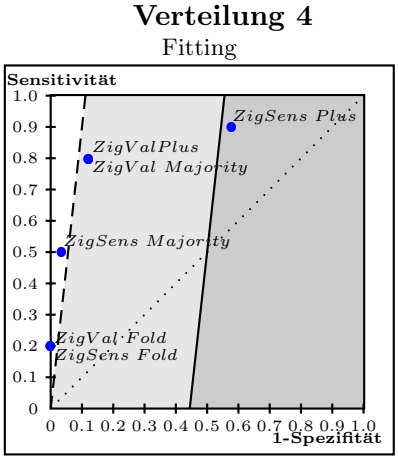


Abbildung 5.16: ROCS-Diagramm für die sechs neuen FaF Trees für Verteilung 4

Tabelle 5.13 zeigt die Ergebnisse und Baumstrukturen der sechs neu konstruierten Klassifikationsbäume, Abbildung 5.16 das entsprechende ROCS-Diagramm.

FaF ZigVal Fold erzeugt den gleichen Baum wie FaF MaxVal, FaF ZigSens Plus den gleichen Baum wie FaF ZigSens und FaF ZigSens Fold den gleichen Baum wie FaF MaxAcc. Bis auf FaF ZigSens Plus zeigen alle Verfahren eine Verbesserung in der Klassifikationsgenauigkeit. FaF ZigVal Plus und FaF ZigVal Plus erzeugen den gleichen Baum. In Kapitel 6 soll ein ausführlicher Vergleich der Verfahren an empirischen Datensätzen erfolgen, der die Grenzen konstruierter Beispiele verlassen kann.

Abschließend sollen einige Anmerkungen zur Erzeugung von FaF Trees mit speziellen Strukturen stehen. Zur Erstellung von L-Shaped Trees und Pectinates können weitere Algorithmen konzipiert werden.

Bei Pectinates müssen lediglich de grundsätzliche Richtung und die Cuereihenfolge bestimmt werden. Letztere hat allerdings keinen Einfluss auf das Klassifikationsergebnis, sondern lediglich einen Einfluss auf die Anzahl der abgefragten Cues. Dementsprechend könnte der Anteil der klassifizierten Objekte als Ordnungskriterium für die Cues gewählt werden. Weiterhin kann überlegt werden, ob die Richtung der Cues nicht im Sinne einer Maximierung der positiven bzw. negativen Validität vorgenommen werden soll und nicht im Sinne einer Maximierung der Klassifikationsgenauigkeit.

Bei der Erstellung von L-Shaped Trees müssen weitere Entscheidungen getroffen werden. Zunächst muss die generelle Form des Baums (positiv dann negativ oder umgekehrt) bestimmt sein, dann muss zudem die Anzahl der Schritte in die Anfangsrichtung festgelegt werden. Je mehr Schritte gewählt werden, desto größer ist der Anteil der Objekte, die in diese Richtung klassifiziert werden. Die Grenzsetzung hängt dabei von der Gewichtung der Entscheidungsfehler ab. Eine Variante des Verfahrens könnte die Anzahl der maximal klassifizierten Objekte in die entsprechende Klasse festlegen oder die minimale Anzahl der korrekt klassifizierten Objekte, entweder exakt oder in verschiedenen Formen der Vereinfachung (konditionale Unabhängigkeit, perfekte Klassifikation etc.). Anschließend müsste die Menge der Cues auf die beiden Teile des Baums aufgeteilt werden. Entweder wird von oben nach unten die relevante Validität maximiert, oder es werden zunächst Teilmengen mit maximaler Gesamtsumme der jeweils relevanten Validität gebildet. Die Reihenfolge innerhalb der Teilmengen hat erneut keinen Einfluss auf das Klassifikationsergebnis und könnte dementsprechend wie bei den Pectinates durch die Entscheidungsgeschwindigkeit bestimmt werden.

Eine letzte hier angedachte Erweiterung könnte sich aus folgender Betrachtung ergeben. Ist die Richtung der Klassifikation zweier aufeinander folgender Cues unterschiedlich, gibt es Bedingungen, unter denen ein Austausch der Cues sinnvoll erscheint: Ist bspw. die positive Validität des zweiten Cues nur unwesentlich geringer als die des ersten, während der erste Cue eine deutliche höhere negative Validität als der zweite aufweist, wäre in dem Fall, in dem der erste Cue für eine positive und der zweite Cue in eine negative Richtung klassifiziert, der Austausch der Cues bei Beibehaltung der Klassifikationsrichtung sinnvoll, wenn genügend Fälle im ersten Schritt unklassifiziert bleiben. Sind die aufeinanderfolgenden Cues die ersten Cues im Baum, kann der Nutzen eines Austausches direkt aus den beiden Cuewert-Kriteriumswert-Verteilungen abgeschätzt werden, indem unter Annahme konditionaler Unabhängigkeit der Validitäten die Klassifikationsgenauigkeit der beiden Ordnungen geschätzt wird. Angenommen, der erste Cue hat eine positive Validität von 0,9 und eine negative Validität von 0,7 und zudem einen Anteil von 0,3 an positiven Werten, der zweite Cue hingegen eine positive Validität von 0,7 bei einer negativen Validität von 0,5 und einem Anteil von 0,4 an positiven Werten. In diesem Fall würde der Baum $(1+, 2-, \ldots)$ bei Annahme konditionaler Unabhängigkeit einen Anteil von $0,3 \cdot 0,9 + 0,7 \cdot 0,6 \cdot 0,5 = 0,48$ korrekt klassifizieren und einen Anteil von $0,3 \cdot 0,1 + 0,7 \cdot 0,6 \cdot 0,5 = 0,24$ falsch. Der Teilbaum $(2+, 1-, \ldots)$ hingegen käme auf ein Ergebnis von $0,4 \cdot 0,7 + 0,6 \cdot 0,7 \cdot 0,8 = 0,616$ korrekt und einen Anteil von $0,4 \cdot 0,3 + 0,6 \cdot 0,7 \cdot 0,2 = 0,204$ falsch. Die Umstellung der Cuereihenfolge scheint in diesem Beispiel angezeigt, auch wenn der zweite Cue eine niedrigere positive Validität hat als der erste. Eine Prüfung von Tauschmöglichkeiten könnte dementsprechend eine sinnvolle Ergänzung des Konstruktionsalgorithmus darstellen.

5.5 Analytische Ergebnisse zu Fast and Frugal Trees

Zum Abschluss der Darstellung von *Fast and Frugal Trees* sollen abschließend einige analytische Ergebnisse präsentiert werden, in denen die Bedeutung des *Splitting Profile* deutlich wird und das Verhältnis zwischen den Bäumen und linearen Klassifikatoren einerseits und TTB andererseits beleuchtet werden kann.

5.5.1 Das Splitting Profile

Tversky, Sattath und Slovic (1988) betonen in der Diskussion lexikografischer Strategien die Attraktivität hinsichtlich der Einfachheit der Anwendung und der Möglichkeit einer überzeugenden Rechtfertigung von Entscheidungen. Ähnlich wie TTB für Paarvergleiche zeichnet FaF Trees die Verwendung einfacher Regelkomponenten aus:

> A fast and frugal tree has the same building blocks as Take The Best: ordered search, one-reason stopping rule, and one-reason decision making. (Gigerenzer, 2004a, S.82)

Das *Splitting Profile* $S(f)$ eines *Fast and Frugal Trees* f kann in folgender Weise relativ zur verwendeten Cuereihenfolge definiert werden: Folgen dem Baum nach der Wurzel k Ebenen, so sei $S(f)_k = 0$. Für alle anderen Ebenen bestimmt die Position des Ausgangsknotens die Stelle des Profils: Führt der Ausgang auf der Ebene i nach der Wurzel zu einer positiven Klassifikation, so sei $S(f)_i = 1$, ansonsten sei $S(f)_i = 0$. Der Baum in Abbildung 5.3 hat damit das *Splitting Profile* $(1, 0, 0)$, der Baum in Abbildung 5.4 das Profil $(0, 0, 0, 0, 1, 0)$. *Pectinates* haben ein *Splitting Profile* von $(0, 0, \ldots, 0)$ oder $(1, 1, \ldots, 1, 0)$. Das Cueprofil eines Objekts kann in ähnlicher Weise kodiert werden. Werden die Cues in der gleichen Reihenfolge geordnet wie sie im Baum von der Wurzel ausgehend verwendet werden, und ist $c_{j,i}$ der Cuewert eines Objekts j für den i-ten Cue, dann kann das Cueprofil eines Objekts o_j als $C(o_j) = (c_{j,1}, c_{j,2}, \ldots, c_{j,k})$ definiert werden. Zur Bestimmung der Klasse, der ein Objekt durch den Entscheidungsbaum zugeteilt wird, reicht ein lexikografischer Vergleich zwischen *Splitting Profile* und Cueprofil. Ein Profil a gilt als lexikografisch größer als ein zweites Profil b, wenn:

$$(\exists i)(c_{a,i} > c_{b,i} \wedge \neg(\exists j)(j < i \wedge c_{a,j} < c_{b,j})) \tag{5.3}$$

Ein Profil a gilt einem zweitens Profil b lexikografisch gleich, wenn:

$$(\forall i)(c_{a,i} = c_{b,i}). \tag{5.4}$$

Diese Definitionen ermöglichen eine einfache Methode der Bestimmung, in welche Klasse ein Objekt durch einen *Fast and Frugal Tree* klassifiziert wird: Ist $S(f)$ das *Splitting Profile* eines *Fast and Frugal Trees* f und ist $C(o_j)$ das Cueprofil eines Objekts o_j, dann wird das Objekt o_j genau dann in die positive Klasse einsortiert, wenn das Cueprofil lexikografisch größer ist als das *Splitting Profile* des Baums. Katsikopoulos et al. (i. Vorb.) beweisen diesbezüglich das folgende Theorem (vgl. Martignon, Katsikopoulos & Woike, i. Vorb.):

Satz 5.1 (*Splitting Profile* eines *Fast and Frugal Tree*) *(i) Jeder Fast and Frugal Tree verhält sich wie ein lexikografischer Klassifikator insofern, als dass es ein eindeutiges Splitting Profile gibt, so dass Objekte nur dann in die positive Klasse kategorisiert werden, wenn das Cueprofil lexikografisch größer ist als das Splitting Profile. (ii) Für jedes Cueprofil gibt es einen vollständigen oder gekürzten Baum, der dieses Profil als sein Splitting Profile aufweist.*

Der zweite Teil des Satzes verweist auf die Möglichkeit, einen *Fast and Frugal Tree* für jedes mögliche Cueprofil zu konstruieren. Da ein Cueprofil im Gegensatz zu einem *Splitting Profile* mit einer 1 an der letzten Stelle enden kann, muss der Zusatz eingeräumt werden, dass dem Cueprofil ein gekürzter Baum entsprechen kann. Ein Cueprofil ist aber genau dann lexikografisch größer als das Profil $(\ldots, 1)$, wenn das um die letzte Stelle reduzierte Cueprofil lexikografisch größer ist als das um die letzte Stelle reduzierte Vergleichsprofil. Ein Profil (c_1, c_2, c_3, c_4) ist bspw. genau dann lexikografisch größer als das Profil $(0, 1, 0, 1)$, wenn das Profil (c_1, c_2, c_3) lexikografisch größer ist als das Profil $(0, 1, 0)$. Die Regel kann auch wiederholt angewendet werden: Ein Profil (c_1, c_2, c_3, c_4) ist genau dann lexikografisch größer als das Profil $(1, 0, 1, 1)$, wenn (c_1, c_2) lexikografisch größer ist als das Profil $(1, 0)$. Ein *Fast and Frugal Tree* hat per Definition zwei Ausgänge auf der letzten Ebene: würde ein *Splitting Profile* auf 1 enden können, dürfte aber kein Objekt aufgrund des letzten Cuewerts in die positive Klasse einsortiert werden. Wäre die Entscheidung auf dieser Ebene bereits gefallen, so ist nur das bis zur letzten 0 im Profil reduzierte Cueprofil das Profil eines legitimen *Fast and Frugal Trees*, die letzten Cues würden ignoriert.

5.5.2 Fast and Frugal Trees und lineare Klassifikatoren

Ein lineares Schwellenwertmodell klassifiziert ein Objekt o_j mit k binären Cuewerten genau dann in die positive Klasse, wenn

$$\sum_{i=1}^{k} c_{j,i} \cdot w_i > G \tag{5.5}$$

Dabei sind $w_1, w_2, \ldots w_k$ den Cuewerten zugeordnete Gewichtungen und G stellt einen Cut-Off dar. Unter bestimmten Bedingungen sind die Entscheidungen dieses linearen Klassifikators mit den Entscheidungen eines *Fast and Frugal Trees* für alle Cueprofile äquivalent. Katsikopoulos et al. (i. Vorb.) beweisen das folgende Theorem:

Satz 5.2 (*Fast and Frugal Tree* und lineares Schwellenwertmodell) *(i) Für jeden Fast and Frugal Tree gibt es ein lineares Schwellenwertmodell mit nicht-kompensatorischen Gewichten, so dass das lineare Modell und der Baum Fälle in identischer Weise klassifizieren. (ii) Für jedes lineare Schwellenwertmodell mit nicht-kompensatorischen Gewichten gibt es einen vollständigen oder gekürzten Fast and Frugal Tree, so dass der Baum und das lineare Modell Fälle in identischer Weise klassifizieren.*

Die Gewichte werden dann als nicht-kompensatorisch bezeichnet, wenn

$$(\forall i)(w_i > \sum_{j>i} w_j). \tag{5.6}$$

Sind die Gewichte eines linearen Modells also nicht-kompensatorisch, kann es in keinem Fall bessere Entscheidungen treffen als ein konstruierbarer *Fast and Frugal Tree*.

5.5.3 Fast and Frugal Trees und Take The Best

Die *Take The Best*-Heuristik ist eine Heuristik für binäre Paarvergleiche. Diese Aufgabe kann allerdings aus einem bestimmten Blickwinkel auch als eine Klassifikationsaufgabe betrachtet werden: Klassifiziert werden nicht Objekte, sondern Objektpaare und die Kategorien sind für das geordnete Tupel (o_a, o_b) die Kategorien $h(o_a) > h(o_b)$ und $h(o_a) < h(o_b)$, wenn $h(o_b)$, wenn $h(o_i)$ den Kriteriumswert für das Objekt o_i darstellt. In diesem Fall können aus den ursprünglichen Cuewerten der Objekte Cuewerte generiert werden, die es ermöglichen, TTB als *Fast and Frugal Tree* darzustellen. Sind ursprünglich k Cuewerte vorhanden, so wird ein neues Cueprofil d für das Objektpaar mit $2k$ Stellen kreiert in der folgenden Form: Ist $c_{a,i} > c_{b,i}$ dann ist $d_{2i-1} = 1$, ist hingegen $c_{a,i} \leq c_{b,i}$, dann ist $d_{2i-1} = 0$. Zudem, ist $c_{b,i} > c_{a,i}$, dann ist $d_{2i} = 0$, ist $c_{b,i} \leq c_{a,i}$, dann ist $d_{2i} = 1$. Das Splittingprofil, mit dem d verglichen werden kann, hat ebenfalls $2k$ Stellen, die abwechselnd mit 0 und 1 besetzt werden, also in der Form $(0, 1, 0, 1, \ldots 0, 1)$. Dabei wird $h(o_a) > h(o_b)$ mit der positiven Klassifikation assoziiert. Die Äquivalenz zwischen diesem Entscheidungsverfahren und TTB ist dann gegeben, wenn es nicht zwei Objekte mit gleichem Cueprofil gibt, da TTB an dieser Stelle eine Zufallsentscheidung treffen würde, die in *Fast and Frugal Trees* nicht vorgesehen ist. Abbildung 5.17 visualisiert einen *Fast and Frugal Tree*, der das Entscheidungsverhalten der TTB-Heuristik imitieren kann.

Zu den Besonderheiten dieser Klassifikationsaufgabe gehören bestimmte Symmetrien: Da die Reihenfolge von Objekten ausgetauscht werden kann, kann jedes Objektpaar zweimal vorgegeben werden. Es gibt damit notwendigerweise ebenso viele falsch positive wie falsch negative Klassifkationen und genauso viele richtige positive wie richtige negative Klassifikationen. Zudem kann niemals der Fall auftreten, dass ein Cuewert an ungerader Stelle den Wert 1 und ein Cuewert an der darauf folgenden Stelle den Wert 0 annimmt, da $c_{a,i}$ nur entweder größer oder kleiner als $c_{b,i}$ sein kann. Diese Bedingungen sind sehr restriktiv. Liegen sie aber in einer Klassifikationsaufgabe vor, so können die Ergebnisse zu Optimalitätsbedingungen für TTB (Katsikopolous & Martignon, 2006) auch auf die entsprechenden *Fast and Frugal Trees* übertragen werden. Die im Allgemeinen auftretenden Asymmetrien verhindern allerdings eine einfache Übertragung der Resultate für beliebige *Fast and Frugal Trees*.

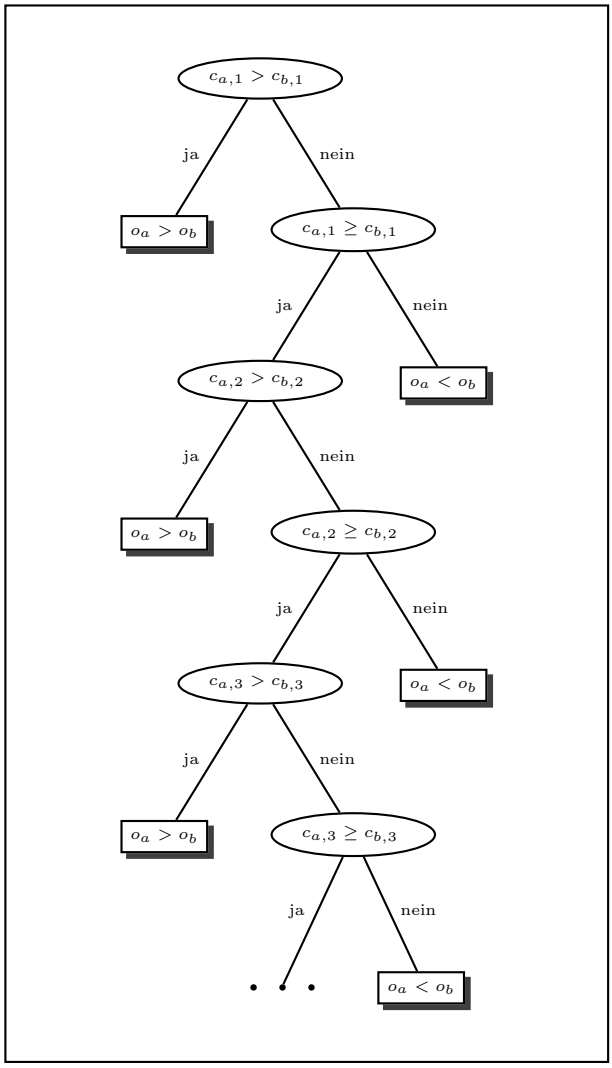


Abbildung 5.17: Darstellung von TTB als *Fast and Frugal Tree*: An der Skizze können die ersten sechs Ebenen des Bams nachvollzogen werden. Verglichen werden zwei Objekte o_a und o_b mit den Cueprofilen c_a und c_b. Sobald sich die Cueprofile der beiden Objekte unterscheiden, wird ein Ausgang des Baums erreicht. Gibt es einen Unterschied an Stelle k, so wird ein Ausgang auf Ebene $2k - 1$ oder $2k$ erreicht.

I have always liked to think that the scientist and the artist belong to the same broad category. Granted, the scientist is more integrated into society than the artist cutting loose in a garret, but the end goal and many aspects of the job are the same. Both are attempting to make sense of their experience of some aspect of reality and in doing this often create imaginary worlds to support their understanding.

(Keane, 1988)

6 Simulationen zur Regelinduktion mit empirischen Daten

In diesem Abschnitt werden die Ergebnisse von drei Simulationsstudien berichtet, die mit ausgewählten Komponenten von GIER$_{[1.0]}$und GIER$_{[1.0]}$.simple_heuristics an insgesamt 33 Datensätzen durchgeführt wurden. Zunächst sollen die Ergebnisse zu GIER$_{[1.0]}$.downhill dargestellt werden. Im Anschluss werden die Ergebnisse der umfangreichsten Untersuchung dargestellt, die für die Varianten der *Fast and Frugal Trees* durchgeführt wurde. Der dritte Teil stellt Ergebnisse für Varianten der Algorithmenklassen und GIER$_{[1.0]}$.ga und GIER$_{[1.0]}$.swarm vor. Abschließend sollen die Ergebnisse der Studien im Überblick diskutiert werden.

6.1 Untersuchung der Klassifikationsleistung des gier.downhill-Algorithmus an drei Datensätzen

Der Algorithmus GIER$_{[1.0]}$.downhill soll kurz vorgestellt werden, bevor einige empirische Ergebnisse berichtet und diskutiert werden.

6.1.1 Darstellung von gier.downhill

Der hier neu vorgestellte Downhill-Algorithmus stellt einen gierigen Algorithmus dar (vgl. Cormen, Leiserson, Rivest & Stein, 2001). Auch wenn er als Hill*climbing*-Algorithmus beschrieben werden kann, orientiert sich die Namensgebung am Prinzip der sinkenden Intervallgrößen, die wie die Höhe des Skifahrers in einem Abfahrtslauf mit jedem Schritt des Algorithmus kleiner werden. Motivation für die Entwicklung des Algorithmus war die Beobachtung von Diagnostikern in realen Entscheidungssituationen. Es soll das Verhalten nachgeahmt werden, an einem Datensatz vorsichtig zu prüfen, welche Extremwerte in Profilen für oder gegen eine Entscheidung sprechen.

6.1.1.1 Grundform des Algorithmus

Der GIER$_{[1.0]}$.downhill-Algorithmus ist ein Klassifikationsverfahren zur sukzessiven Erstellung hierarchisch geordneter Entscheidungsregeln. Nacheinander wird für die Erstellung jeder Regel eine Reihe von Schritten durchlaufen, bis eine Abbruchbedingung erfüllt ist. Zu Beginn jeder Regelerstellung wird eine für alle Personen erfüllte Antezedenz-Bedingung erstellt, die aus einer Konjunktion von Intervallen für jede der im Profil enthaltenen Variablen besteht, wobei die Grenzen der Intervalle auf Minimum und Maximum des Prädiktors in der Stichprobe gesetzt werden. Die Konsequenz der Regel könnte prinzipiell frei gewählt werden, möglich ist das Alternieren zwischen positiven und negativen Entscheidungen, ein wiederholt positives oder ein wiederholt negatives Entscheiden oder auch eine Berücksichtigung der Stichprobenergebnisse zur Festlegung der Entscheidungsrichtung.

In jedem Schritt wird das Profil in der Weise weiter eingeschränkt, dass die Anzahl der durch die Regel abgedeckten Objekte sinkt, sich dabei aber die Qualität der Entscheidungsregel verbessert. Dabei wird der Prozess durch zwei Parameter gesteuert: Zunächst einmal wird festgelegt, um wie viele Personen sich der Kreis der durch die Regel Betroffenen pro Schritt mindestens verringern muss, zum anderen wird festgelegt, wie viele Fälle die Regel am Ende der Prozedur mindestens abdecken muss. Die Einschränkung der Intervalle ist pro Schritt auf die Veränderung eines einzigen Grenzwerts festgelegt, dementsprechend gibt es bei k Variablen $2 \cdot k$ potenziell mögliche Einschränkungen. Ein Zusammenfallen von Minimal- und Maximalwert führt zu einem Intervall, dass von beiden Seiten aus nicht mehr verändert werden kann, während andere Intervalle bei einer Verkürzung zu viele Fälle ausschließen würden, so dass im Einzelfall nicht an jeder Stelle eine Veränderung vorgenommen werden kann.

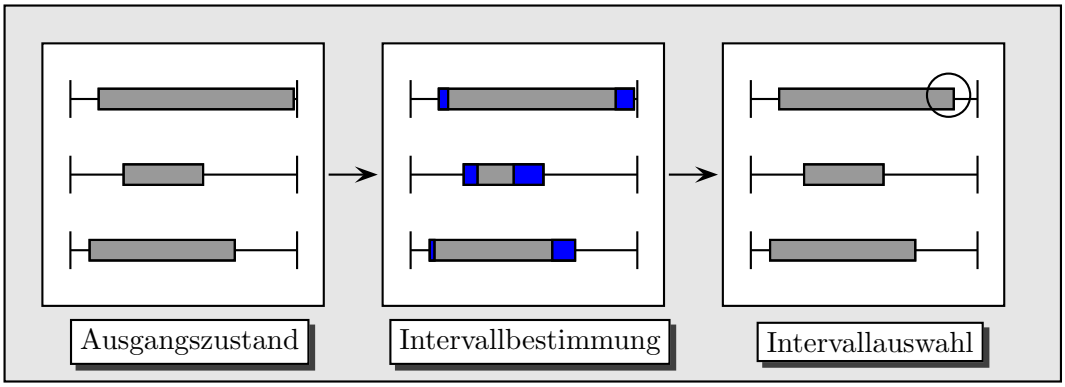


Abbildung 6.1: Iteration im $\text{GIER}_{[1.0]}$.downhill-Algorithmus: In einem ersten Teilschritt werden die Intervallgrößen bestimmt, im zweiten wird ergebnisabhängig ein Intervall ausgewählt und die Regel angepasst.

Jede Iteration besteht aus zwei Teilschritten (vgl. Abbildung 6.1). In einem ersten Schritt wird ermittelt, wie groß die Intervalleinschränkung an jeder möglichen Stelle minimal sein müsste, damit die Anzahl der durch die Regel betroffenen Fälle um den Mindestwert oder einen höheren Wert absinkt. Führt dieser Schritt nicht zu einer Menge möglicher Veränderungen, so wird der Algorithmus für die momentan bearbeitete Regel abgeschlossen. Ansonsten schließt an dieser Stelle der zweite Teilschritt an: Jede der möglichen Veränderungen wird hinsichtlich der Konsequenzen für die Entscheidungsqualität beurteilt. Diese kann entweder durch die Klassifikationsgenauigkeit oder das *KNMK* abgebildet werden. Führt kein Verfahren zu einer positiven Veränderung, wird der Algorithmus für die aktuelle Regel abgebrochen, sonst wird die Veränderung mit der größtmöglichen Verbesserung des Klassifikationsergebnisses vorgenommen. Die neuen Intervallgrenzen bilden wiederum die Ausgangsbedingung für den nächsten Schritt, und das gewählte Prüfintervall muss durch ein angepasstes Prüfintervall ersetzt werden (siehe Abbildung 6.2). Ist nach einem Schritt die Anzahl der durch die Regel abgedeckten Fälle kleiner als die Summe aus Mindestanzahl pro Regel und Mindestanzahl pro Schrittgröße, wird der Algorithmus ebenfalls für die aktuelle Regel abgebrochen. Nach jeder Regelerstellung werden Fälle, die durch diese Regel abgedeckt werden konnten, aus der Stichprobe entfernt, bevor die Intervallgrenzen für die folgende Regel ermittelt werden.

Auf diesem Weg werden sukzessive alle Regeln erstellt, bis die verbleibende Stichprobe kleiner ist als die festgesetzte Mindestgröße für Regeln. In diesem Fall kann wahlweise eine Regel für die verbleibenden Fälle eingefügt werden.

6.1.1.2 Die Bedeutung der Parameter

Die Festsetzung der Parameterwerte beeinflusst das Verhalten des Algorithmus in *Fitting* und Generalisierung: Je kleiner die Mindestgröße für Regeln gewählt wird, desto präziser können Regeln auf dementsprechend kleinere Stichprobenteile zugeschnitten werden. Damit steigt auf der einen Seite die Regelzahl und auf der anderen Seite die Klassifikationsgenauigkeit im Fitting. Im Extremfall kann auf diesem Weg für jedes distinkte Profil eine eigene Vorhersage gemacht werden. Ein Nachteil dieser Vorgehensweise ist eine höhere Wahrscheinlichkeit von *Overfitting*. Die Generalisierungsleistung wird durch eine hohe Regelzahl mit Regeln, die nur aufgrund weniger Beobachtungen gebildet worden sind, im Allgemeinen negativ

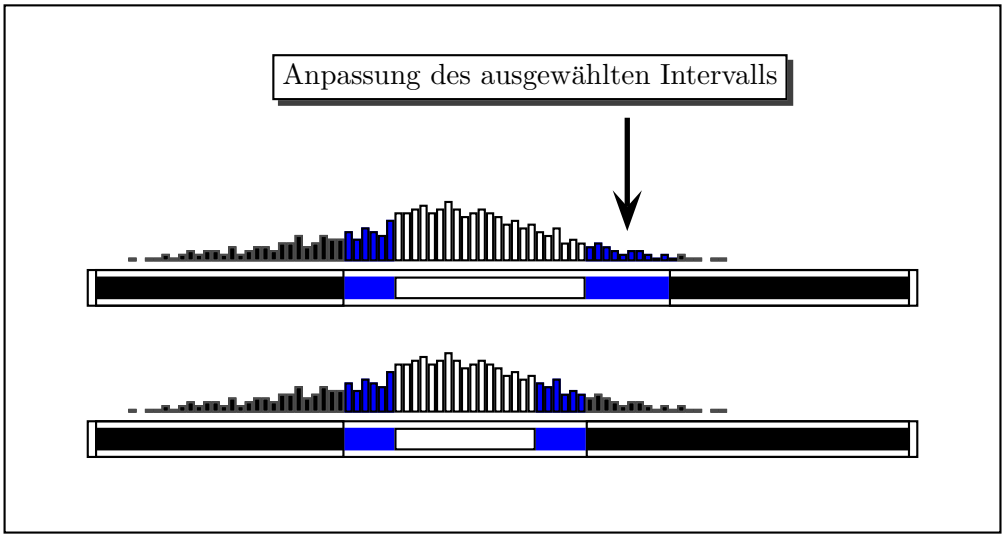


Abbildung 6.2: Intervallanpassung nach einem Downhill-Schritt: Das Histogramm repräsentiert im Beispiel die Häufigkeit der Prädiktorwerte in der unklassifizierten Stichprobe, die blauen Intervalle stellen die Prüfintervalle dar. Nach der Veränderung der ausgewählten Intervallgrenze, muss das Prüfintervall für die betroffene Variable unter Berücksichtigung der Verteilung angepasst werden.

beeinflusst. Eine Ausnahme wären Anwendungen, in denen der Zusammenhang zwischen Profilen und Klassenzugehörigkeiten ausgesprochen stabil ist und Werte mit geringem Fehler erhoben werden können.

Eine ähnliche Rolle spielt der Parameter der minimalen Schrittgröße pro Schritt: Je kleiner der Wert gewählt wird, desto kleiner ist die Schrittgröße. Dies bedeutet auf der einen Seite eine höhere Gefahr, im Suchprozess in lokalen Minima gefangen zu werden, auf der anderen Seite ermöglicht eine klein gewählte Schrittgröße eine flexiblere Anpassung an die vorgefundenen Strukturen. Ein zu hoher Wert kann den Algorithmus im Ausgangszustand einfrieren, genau wie sich ein Algorithmus mit zu kleinem Wert verfangen oder von *Overfitting* betroffen werden kann.

6.1.2 Simulationsergebnisse hinsichtlich der Klassifikationsgenauigkeit

In diesem Abschnitt sollen die Ergebnisse einer Simulationsstudie zur $\text{GIER}_{[1.0]}$.downhill-Heuristik präsentiert werden. Dabei wird der Algorithmus an zwei Datensätzen bei systematischer Variation seiner Parameter eingesetzt, um Einblick in die Funktionsweise und den Anwendungserfolg zu erlangen. Der Downhill-Algorithmus wird dabei ohne Snowboard-Komponente und Variablenreduktion (vgl. Anhang C) mit der Trefferquote als Kriterium verwendet, die Regelentscheidung wird jeweils im Sinne der Maximierung der Trefferquote im Trainingsset gesetzt. In jedem Fall wird eine Regel für Restfälle eingeführt, so dass alle Fälle klassifiziert werden.

Systematisch variiert werden die Parameter für die Fallzahl pro Regel und die Fallzahl pro Schritt. Jede dieser Variablen wird in Schritten von 5 zwischen 5 und 95 variiert, so dass insgesamt $19^2 = 361$ Varianten des $\text{GIER}_{[1.0]}$.downhill-Algorithmus untersucht werden. Die Struktur der Algorithmen ist in Abbildung 6.3 zusammengefasst. Jede dieser Varianten wurde in 500 Trials mit einem Teil der untersuchten Stichprobe als Trainingsset konfrontiert, während die vom Algorithmus erstellten Regeln sowohl an dieser Teilmenge der Objekte (Trainingsset) als auch an der nicht zur Regelkonstruktion verwendeten Teilmenge überprüft werden (Testset). So kann die Leistung in *Fitting* und *Prediction* verglichen werden.

Die untersuchten Datensätze sind die im ersten Kapitel vorgestellten Datensätze *Hochschule* und *Beratungsunternehmen*. Damit für den kleineren *Hochschule*-Datensatz das verwendete Parameterintervall sinnvoll eingesetzt werden kann, wurden in jedem Trial 80% der Stichprobe, also 153 Fälle zufällig in die Trainingsstichprobe aufgenommen, während die 38 verbleibenden Fälle die Teststichprobe darstellen. Der *Beratungsunternehmen*-Datensatz kann aufgrund seiner Größe in zwei gleich große Teilgruppen aufgetrennt werden. Hier wird die Trainingsstichprobe aus 218 Objekten, die Teststichprobe aus 219 Objekten gebildet. Es steht hier ein prozentual kleinerer Anteil an Objekten für die Modellschätzung zur Verfügung (ungefähr 50%). Es werden 180 500 Modellschätzungen pro Datensatz durchgeführt, für jedes Modell werden Klassifikationsgenauigkeit, Sensitivität, Spezifität getrennt für *Fitting* und *Prediction* erhoben, sowie

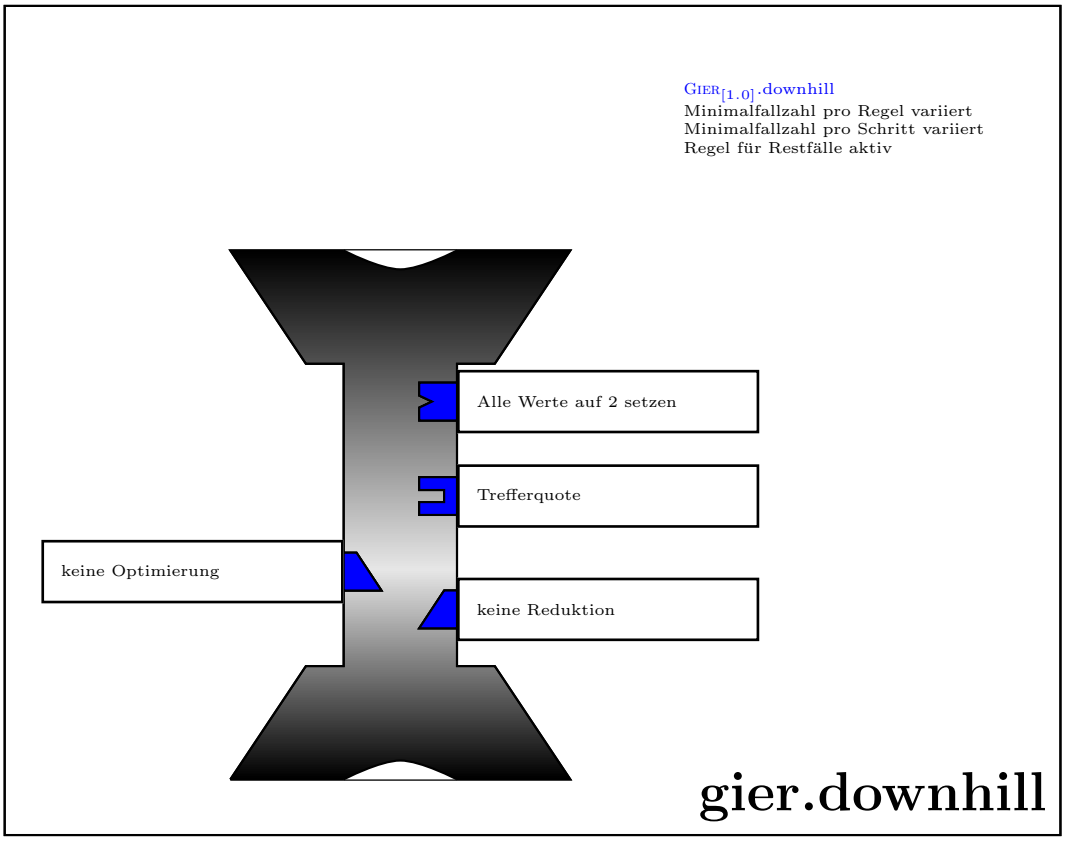


Abbildung 6.3: VBk-Darstellung des Algorithmus GIER$_{[1.0]}$.downhill: Die Abbildung zeigt die Paramter der eingesetzten GIER$_{[1.0]}$.downhill-Varianten.

die Regelzahl, die im *Fitting* festgelegt wird, und dementsprechend auch in der *Prediction*-Bedingung zum Einsatz kommt.

6.1.2.1 Ergebnisse für den Datensatz Hochschule

Zunächst sollen die Ergebnisse für den Datensatz *Hochschule* dargestellt werden. Die Ergebnisse hinsichtlich der Klassifikationsgenauigkeit im *Fitting* sind in Tabelle 6.1, die Ergebnisse für die *Prediction*-Aufgabe in Tabelle 6.2 aufgeführt. Um eine Interpretation der Daten zu erleichtern, soll eine alternative Ergebnisdarstellung gewählt werden: Für die 361 Ergebniswerte werden dazu zunächst Minimum und Maximum ermittelt, so dass jedem einzelnen Wert durch die Transformation $\frac{x-x_{min}}{x_{max}-x_{min}}$ ein Anteil an der Variablenspannweite zugewiesen werden kann, der zwischen 0 und 1 liegen muss. Jeder Anteil wird im Kastendiagramm als Quadrat dargestellt, dessen Flächeninhalt proportional zum Anteil ist, dessen Seitenlänge, also als

$$l = l_{max} \cdot \sqrt{\frac{x - x_{min}}{x_{max} - x_{min}}} \tag{6.1}$$

berechnet werden kann. Die Darstellungsform ermöglicht deshalb einen schnellen Überblick über systematische Variationen.

Das Kastendiagramm für die Klassifikationsgenauigkeit im *Fitting* ist in Abbildung 6.4, das Diagramm für die *Prediction*-Aufgabe in Abbildung 6.5 aufgeführt. Im *Fitting* ist die Klassifikationsgenauigkeit für kleinere Parameterwerte höher als für größere. Die Schrittgröße hat dabei den größten Einfluss. Bei kleiner Schrittgröße und großer Regelmindestfallzahl sind die Ergebnisse besser als bei mittlerer Regelmindestfallzahl. Das farblich markierte Maximum liegt bei einer Setzung von beiden Parametern auf 5. In der *Prediction*-Aufgabe kehren sich diese Ergebnisse nahezu um. Die schlechtesten Ergebnisse werden bei kleiner Regelfallzahl und kleiner Schrittgröße erreicht, ab einer Schrittgröße von 50 wird die Variation kleiner. Das Maximum liegt hingegen bei einer Schrittgröße von 20 und einer Mindestregelfallzahl von 60.

Tabelle 6.1: Klassifikationsgenauigkeit Fitting für die Downhillvarianten (Hochschule)

		Regelgröße																			
		5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	Ø
	5	0,829	0,818	0,814	0,809	0,808	0,804	0,798	0,798	0,801	0,794	0,797	0,791	0,776	0,755	0,728	0,736	0,742	0,752	0,758	0,785
	10	0,821	0,816	0,809	0,805	0,806	0,800	0,794	0,796	0,791	0,792	0,793	0,783	0,763	0,737	0,726	0,729	0,735	0,741	0,749	0,778
	15	0,807	0,801	0,800	0,797	0,795	0,791	0,786	0,780	0,773	0,780	0,779	0,770	0,751	0,723	0,727	0,729	0,734	0,741	0,753	0,769
	20	0,805	0,800	0,795	0,791	0,786	0,782	0,778	0,773	0,768	0,773	0,772	0,771	0,736	0,724	0,727	0,727	0,728	0,737	0,764	0,765
	25	0,790	0,785	0,780	0,779	0,775	0,764	0,749	0,747	0,751	0,759	0,742	0,727	0,724	0,722	0,725	0,731	0,729	0,730	0,734	0,750
	30	0,777	0,771	0,769	0,767	0,767	0,762	0,747	0,752	0,746	0,743	0,742	0,737	0,723	0,722	0,724	0,725	0,725	0,728	0,749	0,746
	35	0,763	0,757	0,756	0,755	0,756	0,755	0,750	0,743	0,731	0,724	0,722	0,723	0,723	0,722	0,723	0,728	0,737	0,737	0,738	0,739
	40	0,749	0,742	0,741	0,740	0,739	0,739	0,738	0,736	0,726	0,723	0,725	0,722	0,723	0,724	0,728	0,727	0,728	0,727	0,728	0,732
	45	0,739	0,731	0,728	0,727	0,727	0,727	0,728	0,727	0,727	0,722	0,722	0,723	0,724	0,725	0,725	0,724	0,725	0,724	0,724	0,726
Schrittgröße	50	0,733	0,726	0,724	0,724	0,724	0,725	0,724	0,724	0,724	0,723	0,724	0,722	0,723	0,721	0,724	0,724	0,723	0,724	0,724	0,724
	55	0,730	0,726	0,722	0,723	0,723	0,724	0,722	0,724	0,723	0,723	0,723	0,722	0,724	0,723	0,721	0,723	0,724	0,723	0,722	0,723
	60	0,728	0,726	0,723	0,723	0,723	0,724	0,723	0,723	0,723	0,723	0,723	0,722	0,722	0,723	0,722	0,722	0,723	0,723	0,723	0,723
	65	0,727	0,723	0,723	0,723	0,723	0,722	0,722	0,721	0,722	0,721	0,724	0,723	0,723	0,722	0,722	0,722	0,722	0,723	0,722	0,723
	70	0,724	0,722	0,723	0,722	0,722	0,723	0,722	0,723	0,722	0,723	0,723	0,723	0,723	0,723	0,722	0,723	0,723	0,723	0,723	0,723
	75	0,723	0,723	0,722	0,723	0,722	0,722	0,723	0,723	0,723	0,723	0,723	0,722	0,722	0,722	0,723	0,723	0,724	0,723	0,723	0,723
	80	0,723	0,723	0,723	0,722	0,724	0,722	0,723	0,722	0,723	0,722	0,723	0,723	0,721	0,724	0,722	0,723	0,723	0,723	0,723	0,723
	85	0,723	0,723	0,723	0,723	0,722	0,723	0,723	0,722	0,723	0,723	0,722	0,723	0,722	0,722	0,722	0,724	0,723	0,722	0,724	0,723
	90	0,722	0,722	0,723	0,723	0,723	0,722	0,724	0,721	0,722	0,723	0,723	0,722	0,721	0,723	0,722	0,723	0,723	0,723	0,722	0,723
	95	0,723	0,723	0,723	0,721	0,723	0,723	0,723	0,721	0,722	0,722	0,723	0,723	0,722	0,723	0,722	0,722	0,722	0,724	0,722	0,722
	Ø	0,754	0,751	0,748	0,747	0,747	0,745	0,742	0,741	0,739	0,739	0,738	0,735	0,730	0,725	0,724	0,725	0,727	0,729	0,733	0,738

Tabelle 6.2: Klassifikationsgenauigkeit Prediction für die Downhillvarianten (Hochschule)

		Regelgröße																			
		5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	Ø
	5	0,637	0,637	0,639	0,641	0,644	0,654	0,653	0,653	0,659	0,661	0,683	0,695	0,699	0,695	0,664	0,642	0,645	0,644	0,661	0,658
	10	0,643	0,639	0,641	0,648	0,656	0,656	0,653	0,657	0,660	0,686	0,710	0,714	0,715	0,698	0,680	0,664	0,653	0,661	0,658	0,668
	15	0,654	0,665	0,664	0,678	0,685	0,683	0,690	0,680	0,685	0,696	0,718	0,741	0,721	0,712	0,685	0,684	0,675	0,682	0,699	0,689
	20	0,661	0,668	0,676	0,682	0,696	0,702	0,689	0,690	0,699	0,723	0,740	0,749	0,731	0,703	0,685	0,683	0,687	0,685	0,729	0,699
	25	0,666	0,673	0,686	0,682	0,691	0,689	0,682	0,676	0,685	0,697	0,712	0,724	0,715	0,721	0,700	0,684	0,694	0,689	0,695	0,693
	30	0,673	0,682	0,688	0,683	0,683	0,692	0,683	0,699	0,708	0,711	0,708	0,717	0,715	0,716	0,709	0,704	0,705	0,709	0,716	0,700
	35	0,679	0,680	0,673	0,678	0,681	0,674	0,680	0,688	0,708	0,719	0,722	0,718	0,721	0,724	0,721	0,705	0,693	0,692	0,690	0,697
	40	0,678	0,673	0,669	0,673	0,670	0,670	0,677	0,680	0,712	0,719	0,714	0,724	0,721	0,716	0,693	0,693	0,692	0,697	0,688	0,693
	45	0,679	0,677	0,684	0,686	0,689	0,688	0,688	0,689	0,683	0,724	0,725	0,715	0,702	0,700	0,698	0,705	0,699	0,700	0,699	0,696
Schrittgröße	50	0,689	0,697	0,704	0,708	0,707	0,705	0,703	0,703	0,703	0,708	0,707	0,714	0,712	0,718	0,707	0,707	0,710	0,706	0,708	0,706
	55	0,692	0,696	0,717	0,714	0,715	0,712	0,714	0,713	0,717	0,714	0,716	0,716	0,714	0,714	0,724	0,714	0,714	0,715	0,723	0,713
	60	0,695	0,699	0,716	0,714	0,715	0,712	0,712	0,715	0,718	0,715	0,714	0,722	0,721	0,719	0,716	0,716	0,717	0,717	0,720	0,714
	65	0,713	0,716	0,721	0,719	0,719	0,722	0,722	0,726	0,723	0,727	0,716	0,720	0,719	0,722	0,724	0,725	0,723	0,721	0,724	0,721
	70	0,724	0,721	0,721	0,725	0,723	0,720	0,726	0,721	0,723	0,721	0,720	0,720	0,722	0,722	0,723	0,722	0,720	0,721	0,720	0,722
	75	0,724	0,720	0,724	0,721	0,723	0,723	0,720	0,722	0,719	0,720	0,719	0,724	0,724	0,724	0,722	0,722	0,718	0,722	0,721	0,722
	80	0,722	0,722	0,721	0,725	0,718	0,724	0,722	0,723	0,722	0,726	0,722	0,722	0,727	0,715	0,724	0,722	0,721	0,722	0,721	0,722
	85	0,721	0,719	0,721	0,720	0,723	0,721	0,722	0,725	0,721	0,721	0,726	0,722	0,724	0,726	0,726	0,718	0,719	0,726	0,718	0,722
	90	0,723	0,724	0,722	0,720	0,720	0,726	0,717	0,728	0,726	0,721	0,720	0,725	0,728	0,721	0,723	0,721	0,721	0,720	0,723	0,723
	95	0,722	0,720	0,720	0,728	0,721	0,722	0,722	0,729	0,723	0,725	0,721	0,722	0,724	0,722	0,725	0,723	0,726	0,715	0,724	0,723
	Ø	0,689	0,691	0,695	0,697	0,699	0,700	0,699	0,701	0,705	0,712	0,717	0,721	0,719	0,715	0,708	0,703	0,702	0,702	0,707	0,704

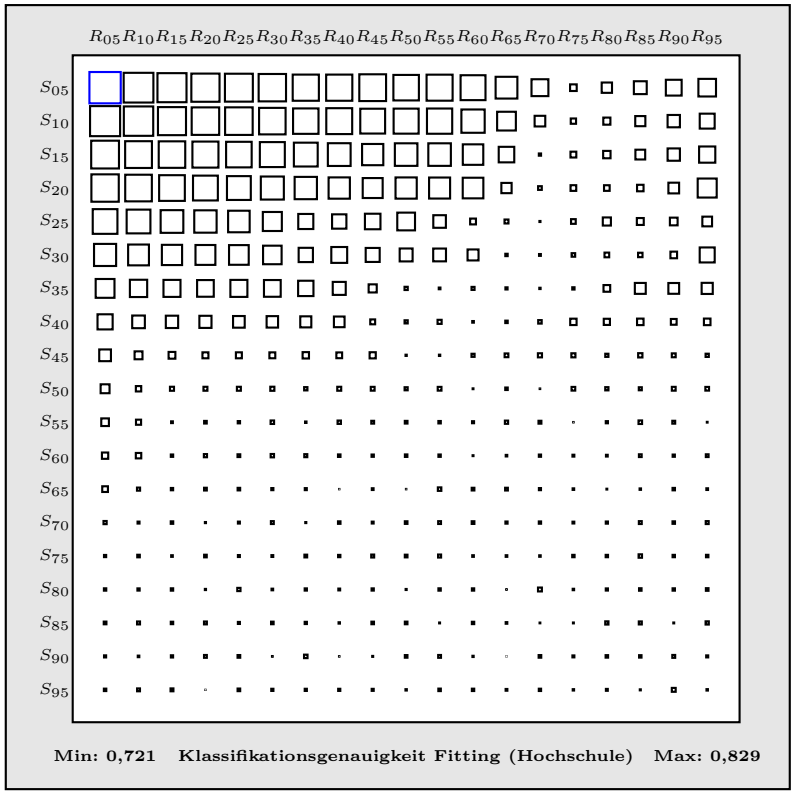


Abbildung 6.4: Kastendiagramm Klassifikationsgenauigkeit Fitting für die Downhillvarianten (Hochschule)

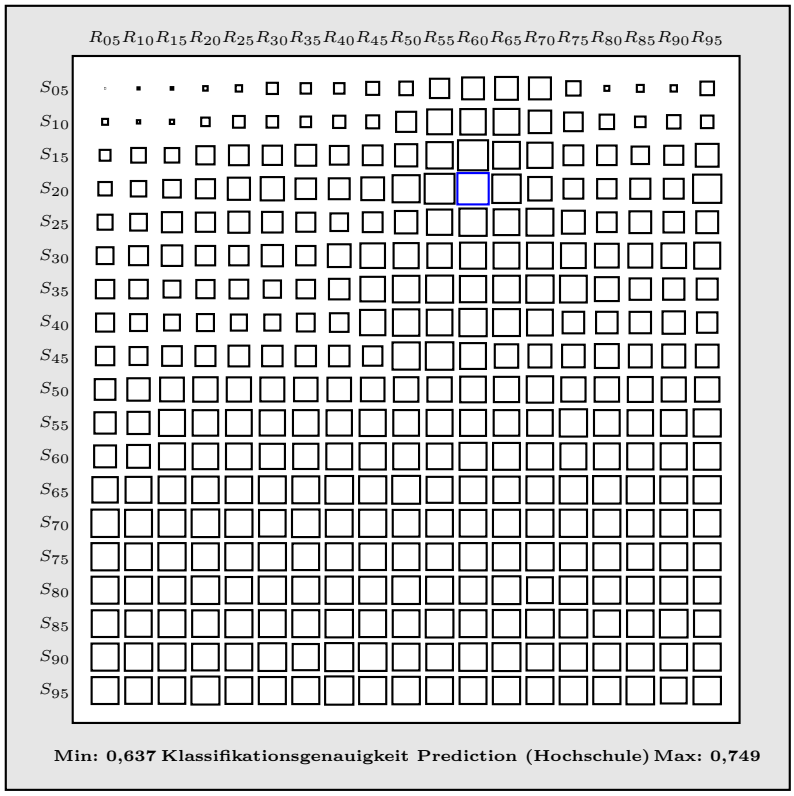


Abbildung 6.5: Kastendiagramm Klassifikationsgenauigkeit Prediction für die Downhillvarianten (Hochschule)

Einen Blick auf die Verschlechterung der Klassifikationsgüte zwischen *Fitting* und *Prediction* ermöglicht die Lokalisierung von Parameterkombinationen, bei denen *Overfitting* in besonderem Ausmaß auftritt. Abbildung 6.6 verdeutlicht, dass sich für den Datensatz *Hochschule* die Diagramme für *Fitting* und *Overfitting* stark ähneln, mit anderen Worten, eine bessere *Fitting*-Leistung geht mit einer stärken Verschlechterung in der *Prediction*-Task einher, was kongruent sein muss zu der beobachteten Reihenfolgeveränderung zwischen *Fitting* und *Prediction*. Ausgenommen ist hier lediglich ein enger Bereich um das Generalisierungsmaximum.

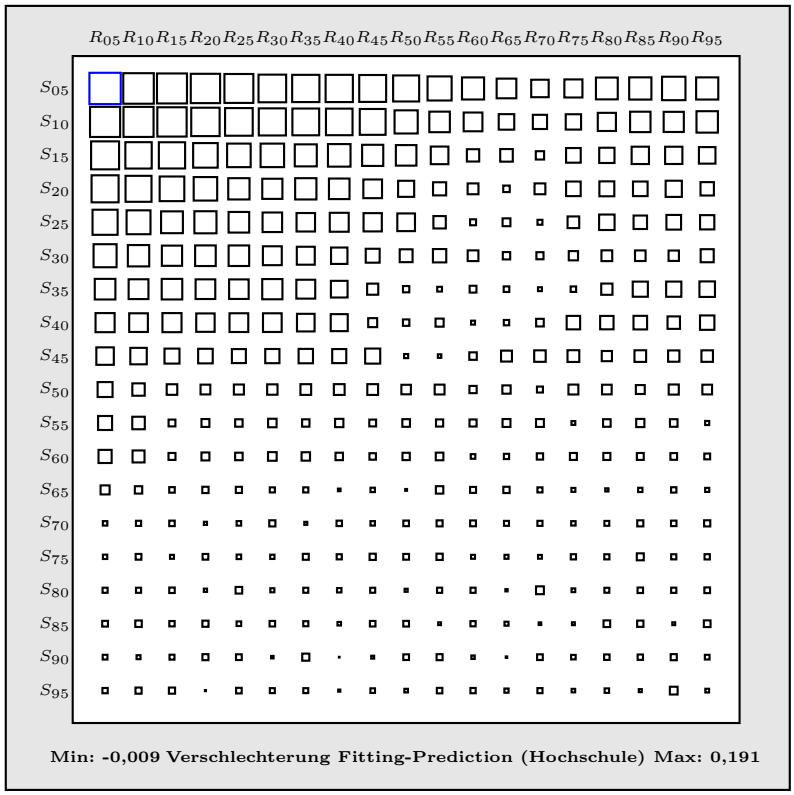


Abbildung 6.6: Kastendiagramm Verschlechterung der Klassifikationsgenauigkeit von Fitting zu Prediction für die Downhillvarianten (Hochschule)

6.1.2.2 Ergebnisse für den Datensatz Beratungsunternehmen

Die Ergebnisse für den Datensatz *Beratungsunternehmen* sind in den Tabellen 6.3 und 6.4 aufgeführt. Die Veranschaulichung in Kastendiagrammen zeigt ein mit den Ergebnissen zum ersten Datensatz gut vergleichbares Ergebnis. In Abbildung 6.7 ist das *Fitting*-Ergebnis veranschaulicht: Die Klippe zwischen guten und schlechten Parameterkombinationen ist weniger stark ausgeprägt als im letzten Beispiel, grob lassen sich etwa vier Regionen mit ähnlicher Ergebnisqualität ausmachen, deren Grenzen diagonal durch das Diagramm verlaufen, d.h. durch eine Kombination von Werten beider Parameter bestimmt werden. Auch bei diesem Datensatz werden die besten *Fitting*-Ergebnisse durch die Wahl kleiner Werte für beide Parameter erzielt. Bei der Analyse der *Prediction*-Ergebnisse in Abbildung 6.8 zeigt sich wiederum eine Umkehrung der Rangfolge der Ergebnisse. Vor allem Kombinationen mit kleinen Werten weisen sehr schlechte *Prediction*-Ergebnisse auf. Im Unterschied zum Datensatz *Hochschule* gibt es allerdings lediglich geringe Unterschiede zwischen Algorithmen ab einer Regelgröße von ungefähr 30 und einer Schrittzahl von ungefähr 30. Das Maximum liegt bei einer Regelgröße von 75 und einer Schrittzahl von 30.

Die Visualisierung der Verschlechterung zwischen Fitting und Prediction in Tabelle 6.9 illustriert eine weitere Systematik: Die Verschlechterung zeigt die bereits vertraute Korrelation mit dem Fittingergebnis, allerdings gibt es einen Bereich um das Prediction-Maximum, in dem die Verschlechterung geringer ist als aufgrund der Fitting-Ergebnisse erwartet, dies gilt ungefähr bei Schrittzahlen zwischen 20 und 60 und Regelgrößen zwischen 60 und 80.

Tabelle 6.3: Klassifikationsgenauigkeit Fitting für die Downhillvarianten (Beratungsunternehmen)

		Regelgröße																			
		5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	Ø
	5	0,799	0,800	0,803	0,799	0,804	0,806	0,805	0,810	0,808	0,797	0,792	0,795	0,797	0,786	0,790	0,791	0,788	0,784	0,775	0,796
	10	0,807	0,805	0,809	0,808	0,804	0,805	0,805	0,808	0,799	0,793	0,792	0,792	0,786	0,782	0,785	0,787	0,781	0,774	0,759	0,794
	15	0,807	0,807	0,808	0,806	0,801	0,801	0,799	0,797	0,790	0,786	0,784	0,778	0,772	0,775	0,777	0,775	0,772	0,762	0,750	0,787
	20	0,806	0,803	0,801	0,800	0,794	0,794	0,791	0,786	0,783	0,779	0,777	0,766	0,765	0,768	0,770	0,769	0,763	0,754	0,739	0,779
	25	0,802	0,798	0,795	0,792	0,791	0,789	0,783	0,778	0,778	0,774	0,769	0,761	0,764	0,766	0,765	0,765	0,756	0,743	0,732	0,774
	30	0,796	0,791	0,787	0,783	0,782	0,782	0,773	0,772	0,769	0,763	0,758	0,758	0,761	0,762	0,764	0,755	0,742	0,735	0,729	0,766
	35	0,788	0,784	0,780	0,778	0,776	0,772	0,772	0,769	0,765	0,755	0,751	0,754	0,758	0,761	0,758	0,751	0,741	0,731	0,728	0,762
	40	0,781	0,779	0,776	0,773	0,774	0,772	0,770	0,769	0,760	0,753	0,758	0,761	0,759	0,758	0,754	0,751	0,741	0,729	0,729	0,760
	45	0,775	0,770	0,768	0,767	0,767	0,766	0,765	0,763	0,762	0,752	0,750	0,756	0,751	0,747	0,736	0,733	0,727	0,729	0,729	0,753
Schrittgröße	50	0,766	0,763	0,758	0,757	0,757	0,757	0,756	0,754	0,755	0,751	0,747	0,748	0,735	0,730	0,730	0,728	0,728	0,730	0,730	0,746
	55	0,762	0,756	0,752	0,749	0,750	0,749	0,749	0,748	0,749	0,748	0,749	0,745	0,736	0,731	0,729	0,727	0,729	0,729	0,732	0,743
	60	0,754	0,748	0,748	0,744	0,741	0,741	0,742	0,741	0,743	0,740	0,743	0,742	0,735	0,730	0,727	0,728	0,729	0,731	0,735	0,739
	65	0,748	0,743	0,739	0,736	0,735	0,736	0,735	0,735	0,736	0,735	0,734	0,736	0,733	0,731	0,729	0,729	0,730	0,731	0,732	0,735
	70	0,745	0,741	0,735	0,733	0,732	0,733	0,733	0,733	0,732	0,732	0,731	0,732	0,730	0,730	0,731	0,731	0,730	0,729	0,730	0,733
	75	0,741	0,733	0,732	0,731	0,730	0,732	0,732	0,732	0,731	0,731	0,731	0,731	0,730	0,730	0,729	0,730	0,729	0,729	0,729	0,731
	80	0,739	0,735	0,731	0,729	0,729	0,727	0,730	0,729	0,730	0,729	0,730	0,729	0,728	0,729	0,728	0,729	0,729	0,729	0,729	0,730
	85	0,739	0,732	0,731	0,731	0,729	0,728	0,728	0,727	0,726	0,729	0,727	0,727	0,727	0,728	0,727	0,729	0,727	0,727	0,728	0,729
	90	0,737	0,731	0,729	0,728	0,728	0,729	0,725	0,727	0,726	0,728	0,727	0,727	0,728	0,728	0,727	0,728	0,728	0,727	0,729	0,728
	95	0,735	0,731	0,729	0,727	0,728	0,726	0,727	0,729	0,728	0,729	0,727	0,728	0,728	0,728	0,727	0,728	0,728	0,727	0,729	0,728
	Ø	0,770	0,766	0,764	0,762	0,761	0,760	0,759	0,758	0,756	0,753	0,751	0,751	0,749	0,747	0,746	0,745	0,742	0,738	0,735	0,753

Tabelle 6.4: Klassifikationsgenauigkeit Prediction für die Downhillvarianten (Beratungsunternehmen)

		Regelgröße																			
		5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	Ø
	5	0,625	0,626	0,631	0,631	0,639	0,648	0,652	0,660	0,669	0,661	0,665	0,673	0,677	0,678	0,688	0,696	0,706	0,708	0,713	0,796
	10	0,637	0,640	0,647	0,650	0,656	0,662	0,674	0,683	0,682	0,679	0,679	0,688	0,688	0,698	0,710	0,717	0,724	0,724	0,724	0,794
	15	0,654	0,654	0,662	0,668	0,671	0,682	0,686	0,698	0,693	0,692	0,697	0,702	0,697	0,711	0,722	0,725	0,726	0,723	0,723	0,787
	20	0,669	0,670	0,679	0,681	0,691	0,694	0,700	0,703	0,702	0,707	0,703	0,704	0,712	0,720	0,728	0,731	0,731	0,726	0,725	0,779
	25	0,681	0,686	0,690	0,696	0,699	0,703	0,707	0,707	0,712	0,714	0,707	0,706	0,717	0,725	0,736	0,739	0,736	0,727	0,724	0,774
	30	0,694	0,697	0,704	0,704	0,711	0,712	0,718	0,720	0,721	0,714	0,715	0,722	0,732	0,739	0,741	0,736	0,729	0,727	0,723	0,766
	35	0,705	0,704	0,710	0,711	0,714	0,722	0,725	0,723	0,722	0,719	0,716	0,729	0,738	0,740	0,735	0,728	0,726	0,724	0,723	0,762
	40	0,707	0,707	0,708	0,715	0,711	0,713	0,715	0,722	0,720	0,726	0,732	0,735	0,736	0,736	0,736	0,733	0,725	0,722	0,714	0,760
	45	0,709	0,712	0,712	0,714	0,711	0,713	0,714	0,715	0,722	0,720	0,727	0,727	0,728	0,727	0,730	0,726	0,724	0,722	0,721	0,753
Schrittgröße	50	0,711	0,710	0,708	0,708	0,708	0,707	0,709	0,710	0,711	0,714	0,719	0,718	0,724	0,725	0,725	0,728	0,726	0,725	0,725	0,746
	55	0,708	0,704	0,704	0,707	0,704	0,705	0,706	0,708	0,707	0,711	0,708	0,714	0,720	0,724	0,725	0,728	0,726	0,726	0,724	0,743
	60	0,706	0,706	0,705	0,705	0,705	0,707	0,709	0,709	0,708	0,709	0,705	0,707	0,714	0,721	0,725	0,724	0,724	0,722	0,718	0,739
	65	0,707	0,704	0,707	0,711	0,709	0,706	0,708	0,710	0,707	0,707	0,709	0,709	0,711	0,718	0,720	0,721	0,721	0,718	0,719	0,735
	70	0,707	0,706	0,707	0,709	0,710	0,709	0,708	0,710	0,708	0,711	0,714	0,712	0,717	0,717	0,718	0,719	0,720	0,718	0,719	0,733
	75	0,709	0,712	0,712	0,712	0,711	0,709	0,712	0,713	0,712	0,713	0,713	0,717	0,719	0,719	0,717	0,720	0,719	0,717	0,718	0,731
	80	0,713	0,713	0,716	0,720	0,720	0,722	0,718	0,721	0,719	0,719	0,720	0,722	0,722	0,722	0,723	0,721	0,722	0,721	0,721	0,730
	85	0,718	0,721	0,722	0,719	0,722	0,723	0,723	0,725	0,726	0,724	0,726	0,725	0,724	0,724	0,725	0,724	0,725	0,725	0,724	0,729
	90	0,717	0,722	0,724	0,724	0,726	0,726	0,729	0,727	0,728	0,725	0,725	0,727	0,723	0,725	0,728	0,726	0,725	0,727	0,725	0,728
	95	0,721	0,722	0,725	0,726	0,726	0,728	0,726	0,724	0,727	0,726	0,728	0,726	0,726	0,725	0,725	0,727	0,726	0,728	0,726	0,728
	Ø	0,770	0,766	0,764	0,762	0,761	0,760	0,759	0,758	0,756	0,753	0,751	0,751	0,749	0,747	0,746	0,745	0,742	0,738	0,735	0,753

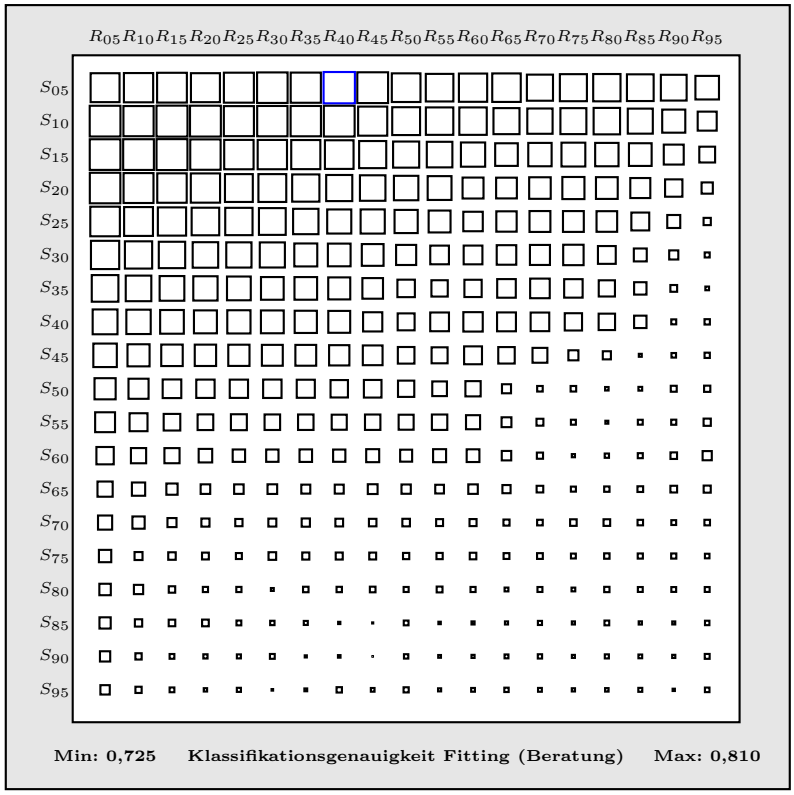


Abbildung 6.7: Kastendiagramm Klassifikationsgenauigkeit Fitting für die Downhillvarianten (Beratung)

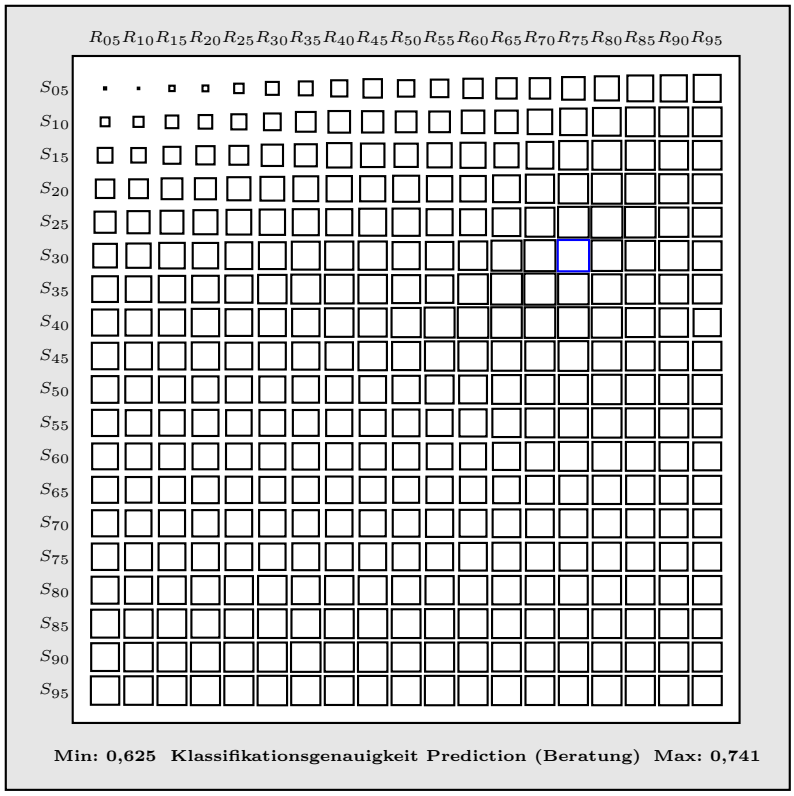


Abbildung 6.8: Kastendiagramm Klassifikationsgenauigkeit Prediction für die Downhillvarianten (Beratung)

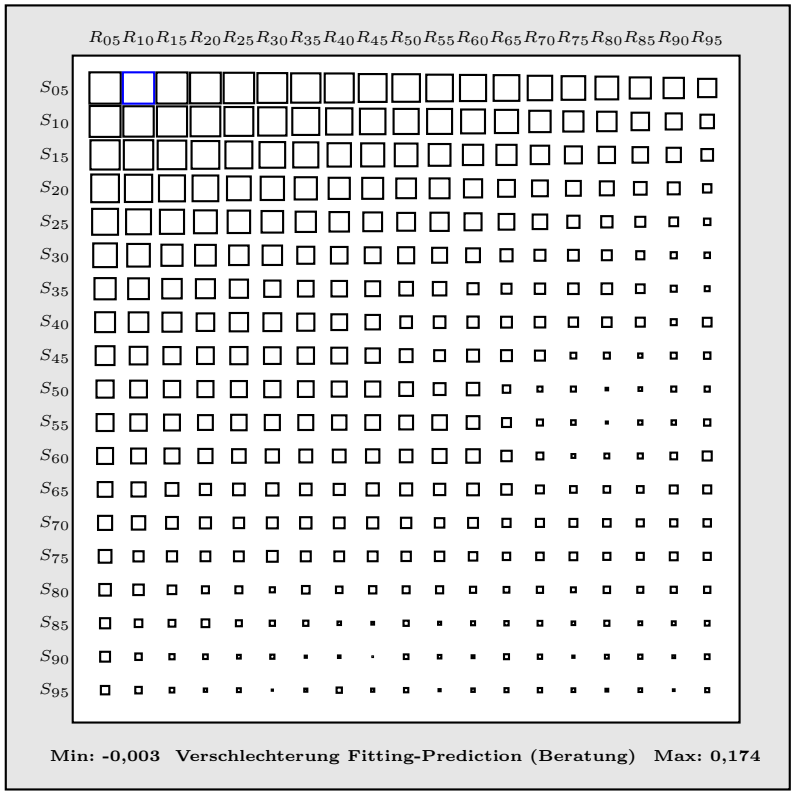


Abbildung 6.9: Kastendiagramm zur Verschlechterung Klassifikationsgenauigkeit für die Downhillvarianten (Beratungsunternehmen)

6.1.3 Ergebnisse hinsichtlich Δ_{UAPV}-Score und Regelanzahl

Neben der Klassifikationsgenauigkeit sollen die Ergebnisse der Algorithmen auch hinsichtlich des Δ_{UAPV}-Scores und der Regelzahl verglichen werden, da die Klassifikationsgenauigkeit nur eines der möglichen Gütekriterien für die Qualität eines Klassifikationsverfahrens darstellt. Auch diese Ergebnisse werden getrennt für die beiden Datensätze präsentiert.

6.1.3.1 Ergebnisse für den Datensatz Hochschule

Die Prediction-Ergebnisse für den Δ_{UAPV}-Score und die durschnittliche Anzahl erstellter Regeln sind in Tabelle 6.5 aufgeführt. Die in Abbildung 6.10 visualisierten *Fitting*-Ergebnisse für den Datensatz Hochschule sind dabei im Durchschnitt alle nichtnegativ, d.h. kein Verfahren liefert Ergebnisse, die schlechter sind als die einer Zufallsauswahl. Analog zu den Klassifikationsgenauigkeitsergebnissen werden die höchsten Werte bei kleinen Schritt- und Regelgrößenzahlen erzielt. Im Unterschied zu diesen gibt es eine zweite Region mit guten Ergebnissen bei kleinen Schrittzahlen und hohen Werten für die Regelgröße. Ab einer Schrittgröße von 50 liegen die Werte nahe bei 0. Die *Prediction*-Ergebnisse zeigen ähnliche Relationen zwischen den Algorithmenvarianten, allerdings liegt das Maximum der Ergebnisse bei einem Wert von 0,305 (vgl. Abbildung 6.11). Dies entspricht relativ genau der Hälfte des Wertes in der *Fitting*-Aufgabe. Die Heranziehung der Kriterien Genauigkeit und Δ_{UAPV}-Score liefert also unterschiedliche Intuitionen bezüglich der Rangfolge der Parameterkombinationen. Werden die Schrittzahlen hoch gewählt, so scheint die Klassifikationsgenauigkeit mit einer einseitigen Konzentration auf eine der beiden Klassen erkauft zu werden. Eine vernünftige Kombination scheint hier die Wahl einer großen Regelzahl kombiniert mit einer kleineren Schrittgröße zu sein. Ist dieser Wert allerdings zu klein, sinkt die Klassifikationsgenauigkeit.

Die Analyse der Regelanzahl liefert ein deutliches Bild (s. Abbildung 6.12): Die Anzahl der erstellten Regeln wird primär durch die gewählte Regelgröße bestimmt. Allerdings gibt es eine deutlich sichtbare Wechselwirkung zwischen Schritt- und Regelgröße: Bei einer mittleren Schrittgröße ist die Regelanzahl bei mittlerer Regelgröße höher als sowohl für kleine und große Wahlen für die Schrittgröße, sie sinkt ab einer Regelgröße von 55 stufenlos auf ungefähr 2 ab. Insgesamt dient damit der Parameter für die Regelgröße dem intendierten Zweck, die Anzahl der erstellten Regeln zu bestimmen. Weiterhin fällt auf, dass die Lösungen mit guten Generalisierungsergebnissen für beide Kriterien sich durch eine geringe Regelzahl

Tabelle 6.5: Δ_{UAPV}-Score Prediction für die Downhillvarianten (Hochschule)

		Regelgröße																			
		5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	Ø
	5	0,151	0,144	0,151	0,141	0,136	0,161	0,157	0,159	0,168	0,175	0,164	0,132	0,107	0,079	0,051	0,083	0,144	0,171	0,203	0,141
	10	0,154	0,144	0,135	0,134	0,137	0,150	0,157	0,160	0,183	0,200	0,189	0,142	0,090	0,033	0,026	0,065	0,110	0,161	0,170	0,134
	15	0,157	0,156	0,152	0,170	0,168	0,173	0,196	0,196	0,212	0,209	0,196	0,183	0,093	0,007	0,040	0,114	0,154	0,206	0,273	0,161
	20	0,141	0,152	0,157	0,151	0,188	0,202	0,183	0,196	0,217	0,214	0,209	0,205	0,063	0,019	0,043	0,054	0,066	0,150	0,305	0,153
	25	0,115	0,111	0,120	0,121	0,138	0,145	0,144	0,134	0,154	0,156	0,071	0,024	0,000	0,001	0,021	0,078	0,080	0,083	0,113	0,095
	30	0,103	0,118	0,131	0,117	0,121	0,129	0,107	0,146	0,109	0,096	0,092	0,071	0,012	0,012	0,015	0,020	0,033	0,058	0,174	0,088
	35	0,109	0,135	0,118	0,125	0,132	0,123	0,121	0,112	0,054	0,001	0,001	0,001	0,001	0,004	0,023	0,055	0,124	0,121	0,124	0,078
	40	0,094	0,095	0,082	0,110	0,090	0,079	0,092	0,101	0,025	0,003	0,002	0,000	0,000	0,009	0,037	0,039	0,030	0,042	0,037	0,051
	45	0,067	0,048	0,048	0,034	0,034	0,032	0,030	0,041	0,023	0,003	0,000	0,002	0,010	0,010	0,010	0,007	0,007	0,017	0,011	0,023
Schrittgröße	50	0,029	0,014	0,009	0,005	0,010	0,012	0,010	0,007	0,007	0,005	0,001	0,002	0,008	0,003	0,005	0,001	0,001	-0,001	0,002	0,007
	55	0,038	0,030	0,001	0,004	0,001	0,002	0,004	0,002	0,001	0,001	0,000	0,004	0,001	-0,001	0,001	0,002	0,001	0,002	0,002	0,005
	60	0,041	0,037	0,004	0,005	0,003	0,001	0,005	0,001	0,002	0,003	0,003	0,001	0,000	0,003	0,001	0,005	0,001	0,002	0,000	0,006
	65	0,013	0,011	0,000	0,000	-0,001	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-0,001	0,000	0,000	0,002	0,000	0,000	0,000	0,001
	70	0,002	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	75	0,001	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	80	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	85	0,000	-0,001	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	90	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	95	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	Ø	0,064	0,063	0,058	0,059	0,061	0,064	0,063	0,066	0,061	0,056	0,049	0,040	0,020	0,009	0,014	0,028	0,040	0,053	0,074	0,050

Tabelle 6.6: Regelzahl für die Downhillvarianten (Hochschule)

		Regelgröße																			
		5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	Ø
	5	8,782	6,822	5,888	5,142	4,710	4,346	3,930	3,854	3,856	3,144	2,978	2,998	2,986	2,954	2,544	2,000	2,000	2,000	2,000	3,839
	10	8,712	7,514	6,172	5,440	4,934	4,534	4,042	3,926	3,872	3,048	2,998	3,000	3,000	2,950	2,302	2,000	2,000	2,000	2,000	3,918
	15	8,156	7,092	6,548	5,166	4,718	4,410	3,916	3,896	3,628	3,030	2,998	3,000	3,000	2,826	2,308	2,000	2,000	2,000	2,000	3,826
	20	8,182	7,092	6,434	5,900	4,840	4,236	3,942	3,852	3,232	3,010	3,000	3,000	2,994	2,458	2,024	2,000	2,000	2,000	2,000	3,800
	25	7,914	6,728	6,196	5,692	5,206	4,030	3,890	3,524	3,414	3,140	3,000	3,000	2,964	2,878	2,428	2,000	2,000	2,000	2,000	3,790
	30	7,384	6,136	5,758	5,396	5,118	4,834	3,540	3,170	3,010	3,000	2,990	2,962	2,500	2,352	2,038	2,000	2,000	2,000	2,000	3,589
	35	6,760	5,216	4,932	4,926	4,902	4,822	4,546	3,250	3,010	3,000	2,994	2,980	2,980	2,912	2,286	2,000	2,000	2,000	2,000	3,553
	40	6,360	5,090	4,930	4,808	4,678	4,338	3,992	3,994	3,006	2,998	2,976	2,964	2,960	2,662	2,000	2,000	2,000	2,000	2,000	3,461
	45	6,246	5,042	4,492	4,002	3,992	3,988	3,988	3,986	3,984	2,994	2,974	2,758	2,002	2,000	2,000	2,000	2,000	2,000	2,000	3,287
Schrittgröße	50	5,442	4,434	4,016	3,998	3,980	3,932	3,940	3,916	3,900	3,710	2,118	2,060	2,006	2,000	2,000	2,000	2,000	2,000	2,000	3,129
	55	5,254	4,710	4,006	3,972	3,936	3,926	3,824	3,404	3,000	3,000	3,000	2,218	2,030	2,000	2,000	2,000	2,000	2,000	2,000	3,067
	60	5,230	4,888	4,006	3,938	3,886	3,662	3,000	3,000	3,000	3,000	3,000	3,000	2,100	2,008	2,000	2,000	2,000	2,000	2,000	3,038
	65	5,350	4,732	3,724	3,382	3,000	3,000	3,000	3,000	3,000	3,000	3,000	3,000	2,998	2,394	2,012	2,000	2,000	2,000	2,000	2,979
	70	4,980	4,232	3,332	3,000	3,000	3,000	3,000	3,000	3,000	3,000	3,000	3,000	3,000	3,000	2,316	2,000	2,000	2,000	2,000	2,940
	75	4,154	4,006	3,644	3,000	3,000	3,000	3,000	3,000	3,000	3,000	3,000	3,000	3,000	3,000	2,998	2,000	2,000	2,000	2,000	2,937
	80	4,034	4,002	3,800	3,032	3,000	3,000	3,000	3,000	3,000	3,000	3,000	3,000	3,000	3,000	2,000	2,000	2,000	2,000	2,000	2,888
	85	4,014	4,002	3,954	3,138	3,000	2,998	3,000	3,000	3,000	3,000	3,000	3,000	3,000	2,000	2,000	2,000	2,000	2,000	2,000	2,848
	90	4,000	4,000	3,978	3,324	3,000	3,000	3,000	3,000	3,000	3,000	2,998	3,000	2,000	2,000	2,000	2,000	2,000	2,000	2,000	2,805
	95	4,004	4,000	3,998	3,468	3,000	3,000	3,000	3,000	3,000	3,000	3,000	2,000	2,000	2,000	2,000	2,000	2,000	2,000	2,000	2,762
	Ø	6,050	5,249	4,727	4,249	3,995	3,792	3,555	3,409	3,259	3,057	2,949	2,839	2,659	2,494	2,171	2,000	2,000	2,000	2,000	3,287

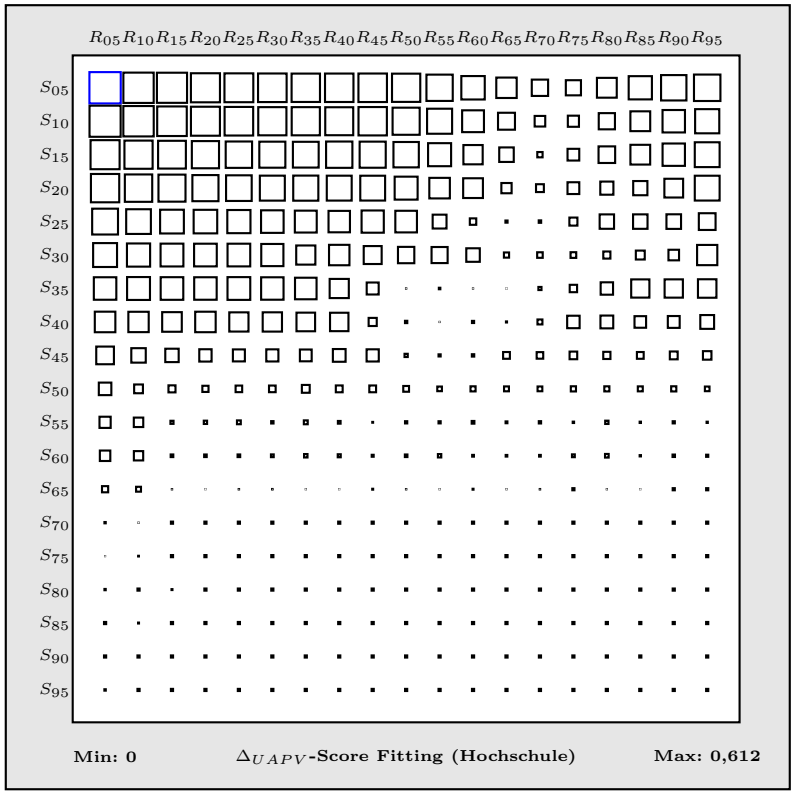


Abbildung 6.10: Kastendiagramm Δ_{UAPV}-Score Fitting für die Downhillvarianten (Hochschule)

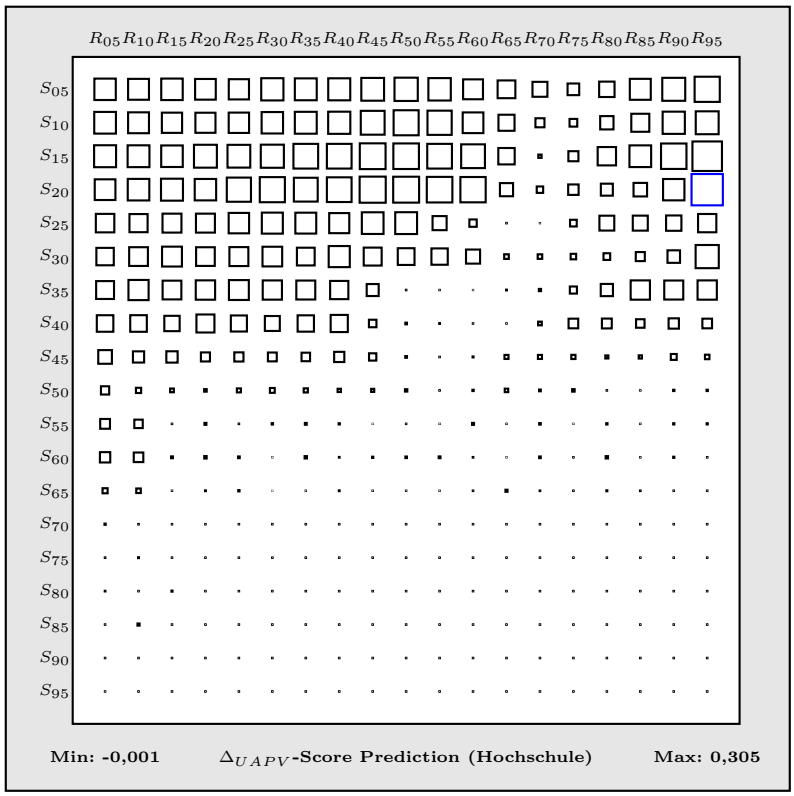


Abbildung 6.11: Kastendiagramm Δ_{UAPV}-Score Prediction für die Downhillvarianten (Hochschule)

mit dementsprechend hoher betroffener Fallzahl auszeichnen.

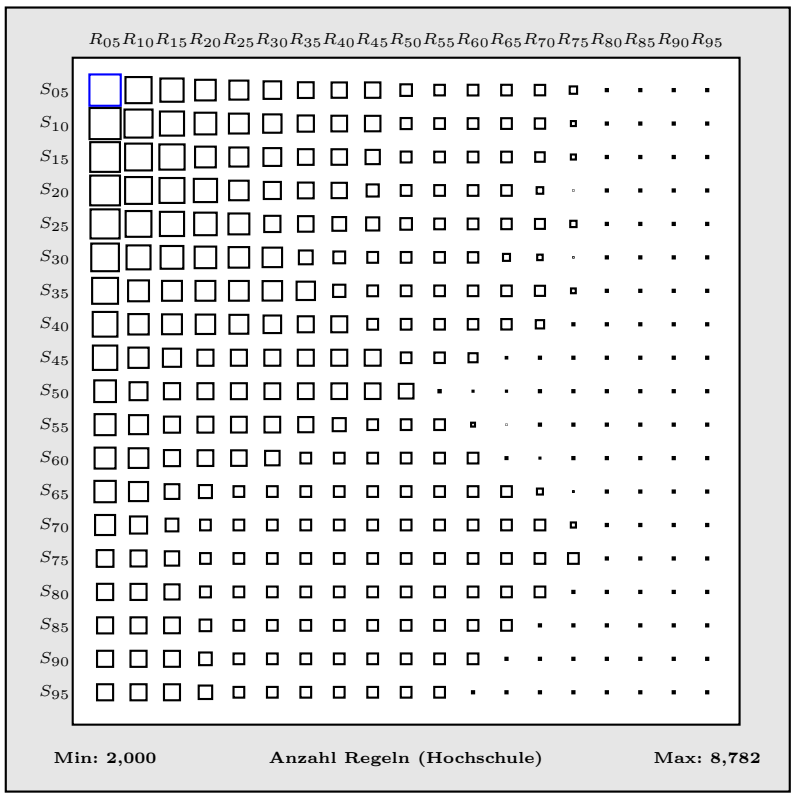


Abbildung 6.12: Kastendiagramm Regelanzahl für die Downhillvarianten (Hochschule)

6.1.3.2 Ergebnisse für den Datensatz Beratungsunternehmen

Die *Prediction*-Ergebnisse bezüglich des Δ_{UAPV}-Scores für den Datensatz *Beratungsunternehmen* sind in Tabelle 6.7 zusammengefasst, die Ergebnisse hinsichtlich der Regelanzahl in Tabelle 6.8. Die Ergebnisse in der *Fitting*-Aufgabe (s. Abbildung 6.13) und der *Prediction*-Aufgabe (s. Abbildung 6.14) zeigen erneut eine vergleichbare Struktur, erneut ungefähr eine Halbierung der absoluten Werte für Δ_{UAPV} zwischen *Fitting* und *Prediction*. Die Parameterkombination, die zu einer maximalen Klassifikationsgenauigkeit führt, liefert dabei auch akzeptable Werte für diesen Gütewert.

Auch die Visualisierung der Regelanzahl in Abhängigkeit von der Parameterkombination (s. Abbildung 6.15) zeigt ein vertrautes Bild: Die Anzahl wird demnach hauptsächlich von der Regelgröße bestimmt mit einer Interaktion wie oben beschrieben. Kombinationen mit guten Werte für beide Kennzahlen führen erneut zu einer geringeren Zahl von Regeln.

6.1.4 ROCS-Diagramme

Zuletzt soll anhand von ROCS-Diagrammen auf eine Besonderheit der beiden Datensätze verwiesen werden, die für die Interpretation der Ergebnisse berücksichtigt werden muss.

6.1.4.1 Ergebnisse für den Datensatz Hochschule

In Abbildung 6.16 sind die Ergebnisse aller Downhill-Varianten für das *Fitting* als Punkte in ein ROCS-Diagramm eingezeichnet worden. In der Darstellung wird ersichtlich, dass die Variation der Ergebnisse hinsichtlich der Sensitivität deutlich geringer ausfällt als die Variation hinsichtlich der Spezifität, bei der fast das gesamte Spektrum ausgenutzt wird. Die negative Klasse ist mit deutlich weniger Fällen besetzt, so dass sich alle Verfahren eher auf eine gute Prädiktion für die positive Klasse konzentrieren. Nahezu alle Ergebnisse liegen dabei oberhalb der Ergebnisse für eine Zufallsauswahl und oberhalb der Ergebnisse für die Wahl der Majoritätsklasse. Im Diagramm der *Prediction*-Ergebnisse (s. Abbildung 6.17) können die bereits in der Analyse der Kennwerte oben angestellten Beobachtungen besser interpretiert

Tabelle 6.7: Δ_{UAPV}-Score Prediction für die Downhillvarianten (Beratungsunternehmen)

		Regelgröße																			
		5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	Ø
	5	0,276	0,263	0,256	0,248	0,241	0,237	0,235	0,227	0,221	0,223	0,224	0,229	0,223	0,209	0,192	0,176	0,155	0,125	0,100	0,214
	10	0,240	0,240	0,233	0,227	0,227	0,214	0,219	0,218	0,209	0,214	0,217	0,225	0,230	0,236	0,220	0,193	0,163	0,132	0,085	0,207
	15	0,225	0,219	0,218	0,211	0,207	0,204	0,207	0,212	0,215	0,207	0,205	0,230	0,230	0,242	0,227	0,186	0,148	0,103	0,061	0,198
	20	0,221	0,212	0,221	0,201	0,213	0,202	0,203	0,205	0,203	0,205	0,215	0,221	0,244	0,239	0,219	0,188	0,143	0,093	0,052	0,195
	25	0,208	0,207	0,206	0,208	0,206	0,201	0,214	0,227	0,224	0,211	0,205	0,208	0,239	0,231	0,219	0,188	0,135	0,071	0,020	0,191
	30	0,198	0,202	0,206	0,198	0,206	0,201	0,226	0,221	0,220	0,200	0,214	0,242	0,243	0,246	0,219	0,162	0,078	0,042	0,010	0,186
	35	0,189	0,192	0,197	0,200	0,205	0,221	0,222	0,227	0,210	0,197	0,198	0,217	0,251	0,234	0,201	0,133	0,079	0,021	0,013	0,179
	40	0,193	0,191	0,194	0,202	0,199	0,201	0,208	0,222	0,199	0,208	0,237	0,240	0,223	0,214	0,191	0,173	0,095	0,011	0,029	0,181
	45	0,189	0,198	0,206	0,210	0,203	0,201	0,210	0,211	0,218	0,183	0,202	0,214	0,177	0,128	0,079	0,040	0,010	0,010	0,013	0,153
Schrittgröße	50	0,189	0,196	0,195	0,200	0,202	0,195	0,198	0,200	0,206	0,206	0,179	0,168	0,068	0,020	0,011	0,004	0,006	0,009	0,038	0,131
	55	0,177	0,173	0,188	0,200	0,190	0,185	0,186	0,191	0,182	0,188	0,185	0,138	0,076	0,027	0,010	0,005	0,016	0,028	0,070	0,127
	60	0,162	0,156	0,155	0,154	0,143	0,129	0,143	0,136	0,124	0,144	0,146	0,139	0,082	0,028	0,012	0,014	0,022	0,039	0,085	0,106
	65	0,138	0,120	0,111	0,090	0,087	0,083	0,092	0,084	0,088	0,096	0,087	0,087	0,077	0,038	0,025	0,026	0,038	0,059	0,044	0,077
	70	0,093	0,085	0,079	0,065	0,065	0,067	0,065	0,064	0,070	0,057	0,050	0,054	0,033	0,038	0,024	0,023	0,027	0,034	0,030	0,054
	75	0,077	0,062	0,054	0,049	0,055	0,053	0,047	0,045	0,046	0,048	0,039	0,032	0,022	0,023	0,030	0,020	0,024	0,031	0,031	0,041
	80	0,054	0,038	0,031	0,019	0,020	0,018	0,020	0,018	0,020	0,022	0,015	0,013	0,017	0,016	0,011	0,015	0,016	0,014	0,013	0,021
	85	0,039	0,025	0,011	0,011	0,010	0,010	0,010	0,008	0,008	0,007	0,006	0,009	0,007	0,007	0,006	0,007	0,006	0,008	0,007	0,011
	90	0,028	0,015	0,006	0,005	0,003	0,001	0,002	0,003	0,003	0,003	0,005	0,002	0,007	0,004	0,001	0,002	0,005	0,002	0,003	0,005
	95	0,018	0,010	0,004	0,003	0,002	0,002	0,003	0,003	0,001	0,000	0,000	0,002	0,002	0,004	0,004	0,001	0,002	0,001	0,002	0,003
	Ø	0,153	0,148	0,146	0,142	0,141	0,138	0,143	0,143	0,141	0,138	0,138	0,140	0,129	0,115	0,100	0,082	0,062	0,044	0,037	0,120

Tabelle 6.8: Regelzahl für die Downhillvarianten (Beratungsunternehmen)

		Regelgröße																			
		5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	Ø
	5	10,420	8,124	6,856	6,040	5,624	5,232	4,964	4,776	4,594	4,282	3,964	3,928	3,848	3,436	2,998	3,000	2,992	2,986	2,966	4,791
	10	9,532	8,320	6,914	6,126	5,614	5,320	4,998	4,772	4,510	4,170	3,970	3,920	3,794	3,174	3,000	3,000	3,000	2,994	2,982	4,743
	15	9,272	8,312	7,580	6,356	5,654	5,334	5,000	4,686	4,506	4,078	3,930	3,826	3,588	3,064	2,998	2,998	2,992	2,988	2,952	4,743
	20	9,132	8,034	7,280	6,924	5,694	5,280	4,932	4,730	4,348	3,972	3,902	3,740	3,404	3,046	3,000	2,996	2,988	2,980	2,934	4,701
	25	8,706	7,642	7,146	6,604	6,296	5,292	4,778	4,588	4,324	4,032	3,894	3,596	3,184	3,014	3,000	2,988	2,996	2,982	2,900	4,630
	30	8,520	7,452	6,776	6,290	6,032	5,706	4,794	4,460	4,062	3,896	3,728	3,598	3,152	3,010	2,992	2,998	2,986	2,926	2,812	4,536
	35	8,130	7,072	6,436	6,094	5,746	5,380	5,182	4,396	4,044	3,766	3,596	3,358	3,128	3,036	3,000	2,996	2,998	2,924	2,676	4,419
	40	7,650	6,452	5,954	5,736	5,638	5,484	5,220	4,932	4,002	3,570	3,190	3,038	3,000	3,000	2,996	2,998	2,996	2,850	2,394	4,268
	45	7,502	6,314	5,752	5,494	5,260	5,058	4,898	4,746	4,630	3,706	3,386	3,082	3,000	2,998	2,994	2,996	2,996	2,996	2,722	4,238
Schrittgröße	50	6,842	5,712	5,260	5,000	4,922	4,904	4,846	4,768	4,602	4,332	3,464	3,030	3,008	2,998	2,996	2,988	2,968	2,930	2,770	4,123
	55	6,444	5,526	5,112	5,004	4,904	4,826	4,674	4,570	4,344	4,078	4,000	3,074	3,012	2,994	2,950	2,914	2,840	2,774	2,524	4,030
	60	6,178	5,366	5,092	4,774	4,522	4,356	4,170	3,998	3,998	3,992	3,982	3,960	3,084	3,010	2,912	2,656	2,516	2,456	2,192	3,853
	65	5,964	5,036	4,608	4,248	4,062	4,008	3,988	3,992	3,988	3,940	3,834	3,768	3,516	3,010	2,892	2,556	2,182	2,004	2,000	3,663
	70	5,640	4,632	4,276	4,068	4,000	3,998	3,994	3,970	3,950	3,862	3,692	3,526	3,348	3,240	2,872	2,474	2,154	2,008	2,000	3,563
	75	5,586	4,484	4,164	4,034	3,990	3,992	3,968	3,932	3,754	3,716	3,550	3,406	3,186	2,998	3,000	2,636	2,328	2,088	2,004	3,517
	80	5,768	4,552	4,226	4,030	3,928	3,824	3,702	3,660	3,496	3,426	3,210	3,000	3,000	3,000	3,000	3,000	2,598	2,404	2,158	3,473
	85	5,950	4,570	4,126	3,892	3,688	3,608	3,402	3,276	3,174	3,000	3,000	3,000	3,000	2,998	3,000	3,000	3,000	2,670	2,440	3,410
	90	5,958	4,462	3,946	3,692	3,560	3,370	3,160	2,998	3,000	3,000	2,998	3,000	3,000	3,000	3,000	3,000	3,000	3,000	2,574	3,354
	95	5,876	4,280	3,724	3,506	3,244	3,000	3,000	3,000	3,000	3,000	2,998	3,000	3,000	3,000	3,000	3,000	3,000	3,000	3,000	3,296
	Ø	7,319	6,123	5,538	5,153	4,862	4,630	4,404	4,224	4,017	3,780	3,594	3,413	3,224	3,054	2,979	2,905	2,817	2,735	2,579	4,071

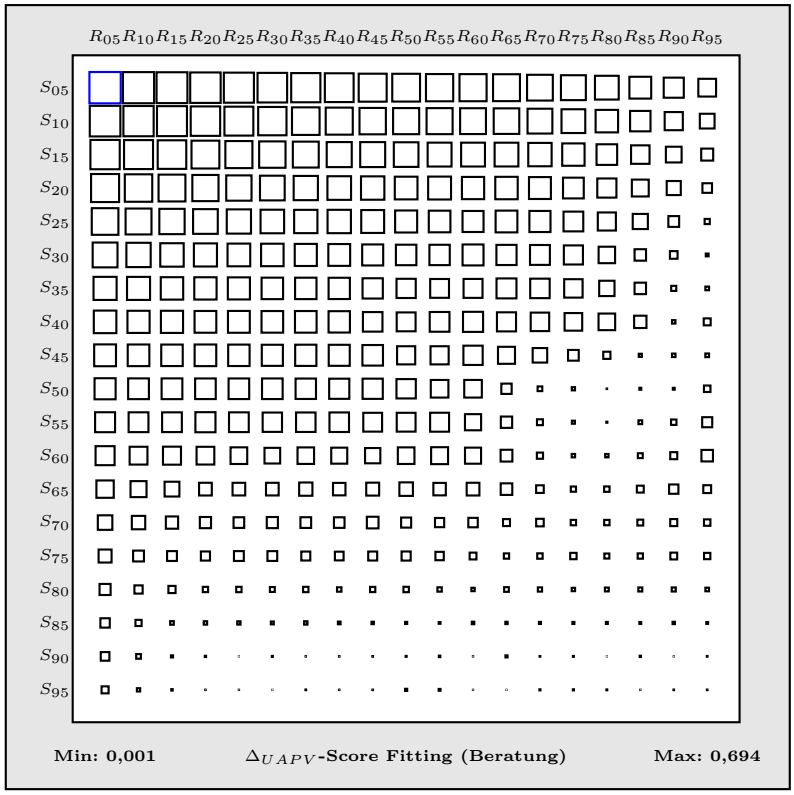


Abbildung 6.13: Kastendiagramm Δ_{UAPV}-Score Fitting für die Downhillvarianten (Beratungsunternehmen)

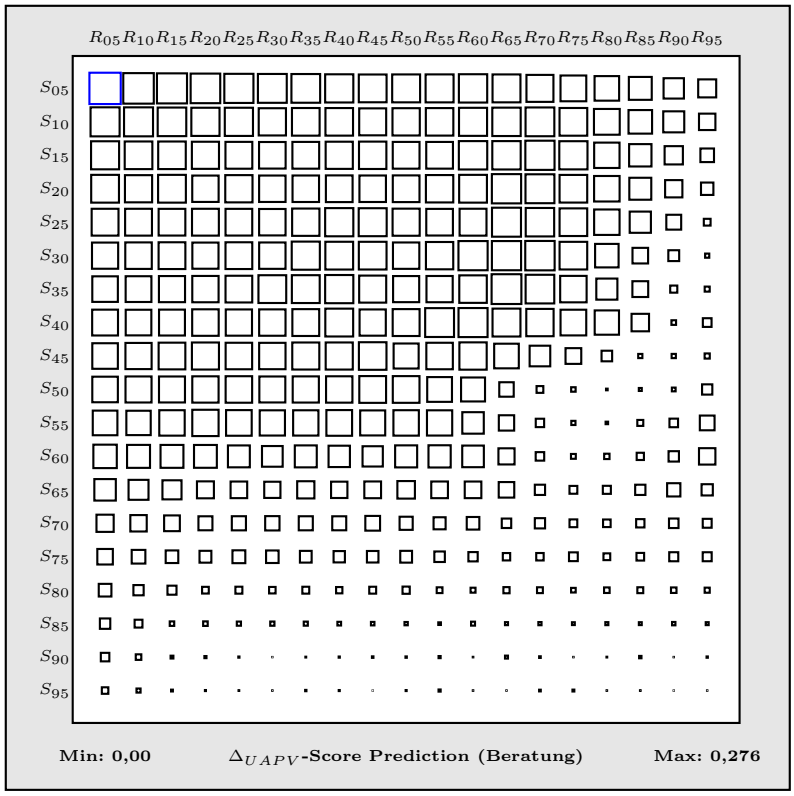


Abbildung 6.14: Kastendiagramm Δ_{UAPV}-Score Prediction für die Downhillvarianten (Beratungsunternehmen)

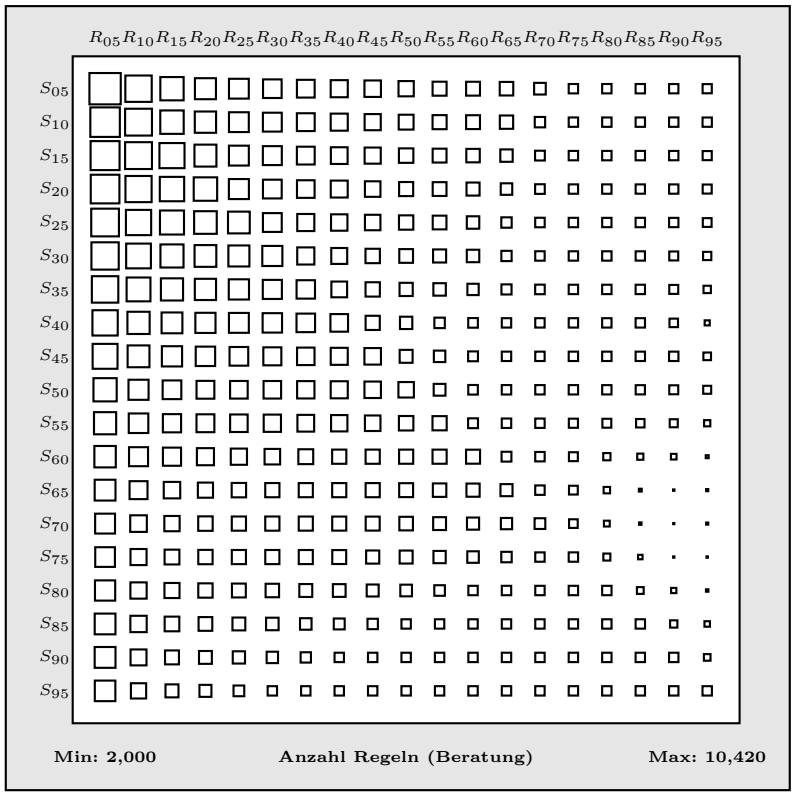


Abbildung 6.15: Kastendiagramm Regelanzahl für die Downhillvarianten (Beratungsunternehmen)

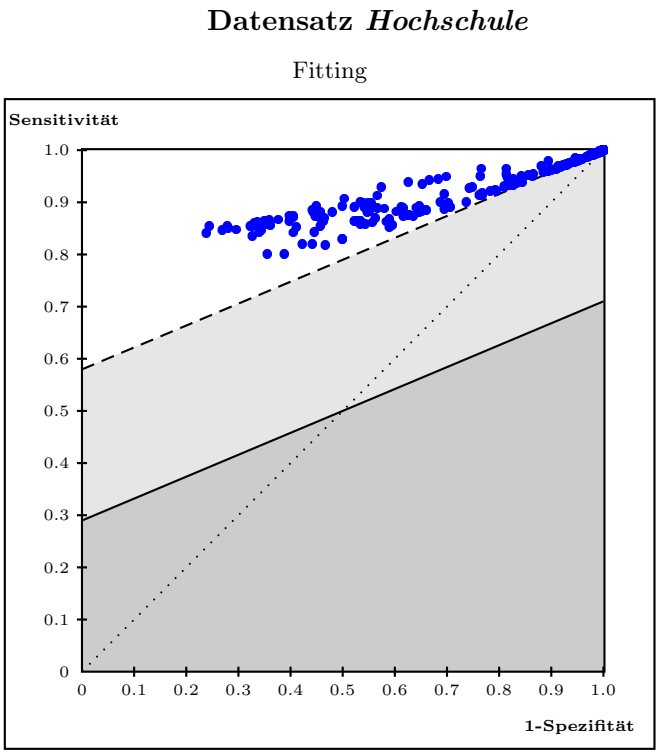


Abbildung 6.16: ROCS-Diagramm für die Fitting-Ergebnisse der Downhill-Varianten (Hochschule)

Datensatz *Hochschule*

Prediction

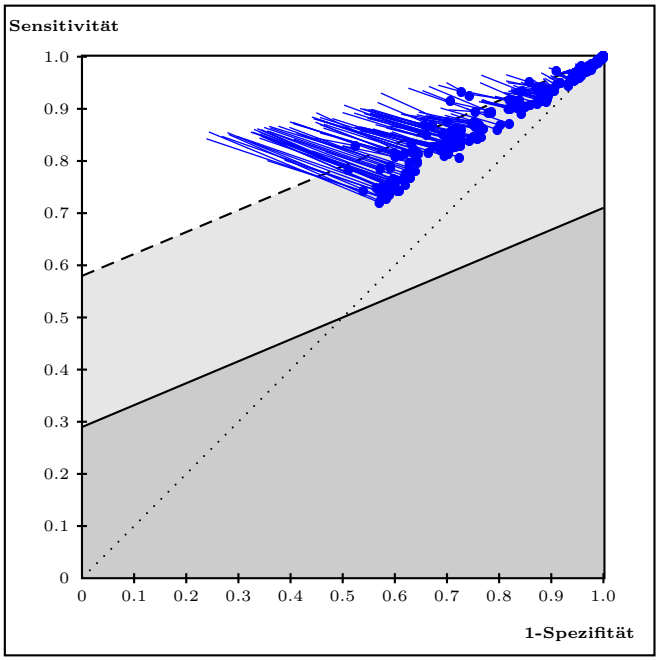


Abbildung 6.17: ROCS-Diagramm für die Prediction-Ergebnisse der Downhill-Varianten (Hochschule)

werden. Die eingezeichneten Linien entsprechen der Strecke zwischen Fitting- und Prediction-Ergebnis einer Parameterkombination.

Es wird deutlich, dass die Linien zueinander größtenteils parallel sind und dass die Veränderungsrichtung ungefähr orthogonal zur Winkelhalbierenden des Diagramms verläuft. Dies entspricht inhaltlich einer größenvergleichbaren Verschlechterung von Sensitivität und Spezifität. Gleichzeitig ist erkennbar, dass nahezu alle Verfahren den Abstand zur Winkelhabierenden zwischen Fitting und Prediction ungefähr halbieren, so dass sich die Rangfolge der Algorithmen hinsichtlich des Δ_{UAPV}-Kriteriums kaum verändert. Dies wurde beim Vergleich der Kastendiagramme oben bereits angemerkt, kann am ROCS-Diagramm aber hinsichtlich der Aufteilung auf Sensitivität und Spezifität präzisiert werden.

Die Verschlechterung hat in nahezu allen Fällen ebenfalls die Konsequenz, dass die neue Prädiktionsleistung hinsichtlich der Klassifikationsgenauigkeit unterhalb der Leistung der Majoritätsregel liegt, mit anderen Worten, die Verfahren sind nur im *Fitting* genauer als die einfache Klassifikation in die Mehrheitsklasse, in der *Prediction*-Aufgabe sind sie zum Teil sogar deutlich schlechter. Dieses Ergebnis kann nicht über zufällige Aufteilungen der Stichprobe erklärt werden, da jedem Datenpunkt 500 Aufteilungen zu Grunde liegen. Dementsprechend muss festgehalten werden, dass jene Parameterkombinationen, die zu einer besseren Anpassung an die Daten führen, gleichzeitig am stärksten ein *Overfitting*-Problem aufweisen. Es sind nur wenige Parameterkombinationen erkennbar, die sich geringer verschlechtern als die übrigen und auch in der Prediction-Aufgabe oberhalb der Ergebnisse der Majoritätsregel liegen.

6.1.4.2 Ergebnisse für den Datensatz Beratungsunternehmen

Für den Datensatz *Beratungsunternehmen* sind die *Fitting*-Ergebnisse in Abbildung 6.18, die *Prediction*-Ergebnisse in Abbildung 6.19 veranschaulicht. Im *Fitting* fällt auf, dass die Spannbreite der Ergebnisse hinsichtlich der Spezifität größer, hinsichtlich der Sensitivität kleiner ist als für den Datensatz *Hochschule*. Die Punktwolke ist insgesamt geschlossener, einzelne Punkt weisen kleinere Abstände zu benachbarten Punkten auf.

Bei der Analyse der Prediction-Ergebnisse gibt es zunächst Ähnlichkeiten mit den Prediction-Ergebnissen für den Datensatz *Hochschule*: Die Richtung der Veränderungen ist für fast alle Verfahren ungefähr orthogonal zur Winkelhalbierenden, dabei dementsprechend zueinander nahezu parallel. Der Abstand zur Winkelhalbierenden wird erneut im Prediction-Task drastisch verkürzt, was auf eine deutliche Verschlechterung der Ergebnisse und Overfitting hinweist. Im Unterschied zum Beispiel zuvor sind allerdings keine

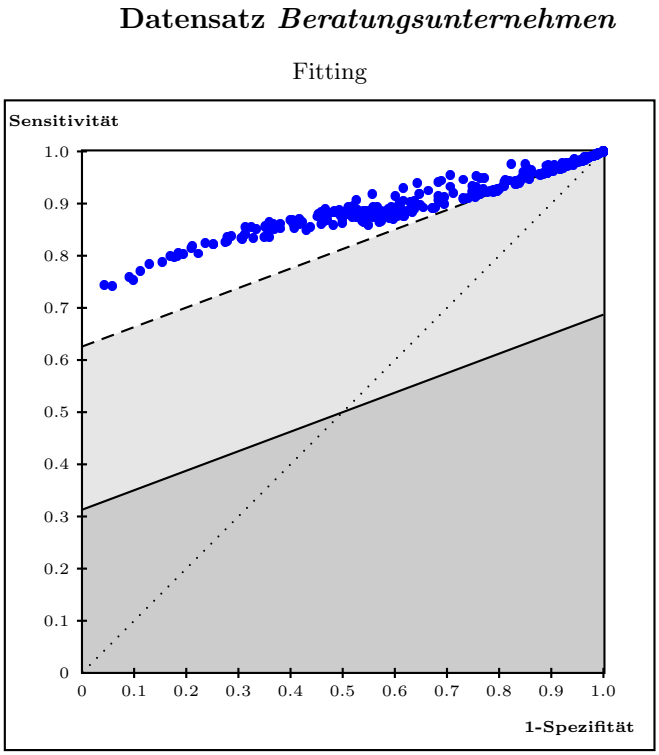


Abbildung 6.18: ROCS-Diagramm für die Fitting-Ergebnisse der Downhill-Varianten (Beratungsunternehmen)

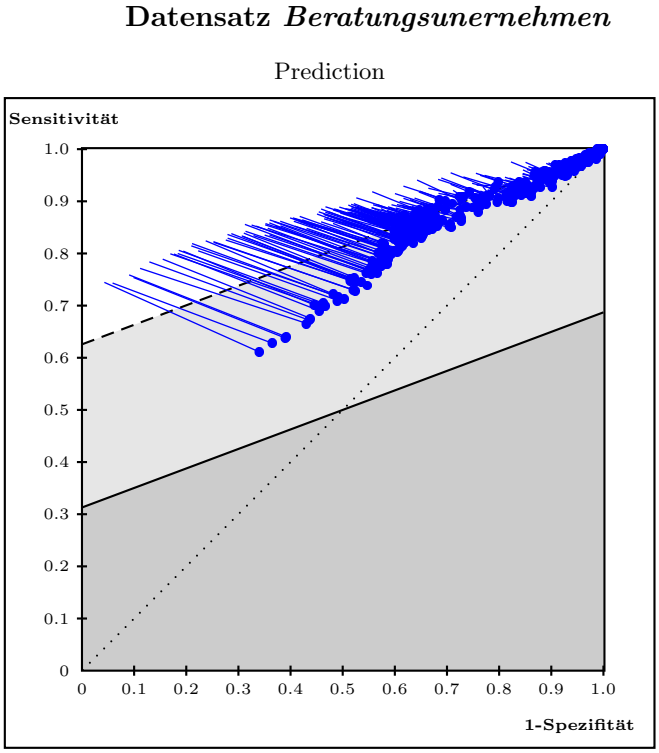


Abbildung 6.19: ROCS-Diagramm für die Prediction-Ergebnisse der Downhill-Varianten (Beratungsunternehmen)

Parameterkombinationen erkennbar, die sich deutlich geringer verschlechtern als die übrigen, und auch in der Prediction-Aufgabe oberhalb der Ergebnisse der Majoritätsregel liegen.

6.1.5 Diskussion der Ergebnisse und Empfehlungen für die Parameterwahl

Die vorliegenden Ergebnisse basieren lediglich auf der Untersuchung von zwei Datensätzen mit jeweils einer festgesetzten Größe des Training-Test-Splits. Es bietet sich hier aber trotz der methodischen Abweichung (Kreuzvalidierung versus Stichprobentrennung) ein Vergleich mit den Ergebnissen von Montel (2005) an (vgl. Abbildung 2.12): Sieht man von der Leistung des CART-Algorithmus ab, die durch die erfolgte Stabilitätsprüfung am gesamten Datensatz zudem in ihrer Vergleichbarkeit problematisiert werden muss, dann kann festgestellt werden, dass die Downhill-Ergebnisse in *Fitting* und *Prediction* in einem ähnlichen Bereich liegen wie die Ergebnisse der Standard-Verfahren. Dies ist insbesondere deshalb bemerkenswert, weil dem Downhill-Algorithmen ein kleinerer Teil der Stichprobe zur Verfügung gestellt wurde als den Algorithmen, die eine Leave-One-Out-Kreuzvalidierung durchlaufen haben. Fokussiert man auf die Beobachtung, dass keiner der Algorithmen in beiden Studien zufriedenstellende Werte für Fitting und Prediction erbringen konnte, stellt sich an dieser Stelle ernsthaft die Frage, ob die Datensätze selbst genug Struktur aufweisen, die von einem Algorithmus erfasst werden kann. In dem Moment, in dem jede positive Abweichung beim Fitting einer Verschlechterung in der Prediction-Task gegenübersteht, beginnen die Ergebnisse in ihrer Struktur den Resultaten zu ähneln, die bei der Anwendung auf zufällig generierte Daten zu erwarten wären. Die Frage nach der Strukturiertheit der Datensätze ließe sich aber nur bei Auffinden eines Algorithmus mit besseren Ergebnissen als den beobachteten positiv beantworten. An dieser Stelle muss deshalb zusammengefasst werden, dass aufgrund der ersten beiden Anwendungen nicht endgültig über die Qualität und potenzielle Nützlichkeit des Downhill-Algorithmus geurteilt werden kann. Hierzu sind weitere Anwendungen an empirischen Datensätzen in Ergänzung mit Anwendungen an konstruierten Datensätzen mit bekannter Struktur notwendig.

Für die Parameterwahl kann ebenfalls nur ein vorsichtiger Rat gegeben werden: Die Regelgröße sollte nicht zu klein gewählt werden, um Overfitting zu vermeiden. Für die Schrittgröße scheint es hingegen eine untere und eine obere Grenze zu geben: Wird die Schrittgröße zu klein gewählt, ist Overfitting die zu erwartende Konsequenz. Wird die Schrittgröße zu groß gewählt, ist das Verfahren nicht flexibel genug, um sich an Datenstrukturen anzupassen. Hierunter leidet bei ungleicher Klassenbesetzung der Δ_{UAPV} wahrscheinlich mehr als die Klassifikationsgenauigkeit. Trotz der negativen Einräumungen kann festgehalten werden, dass unabhängig von der Erstellung von Klassifikationsmodellen unter Umständen viel über einen Datensatz gelernt werden kann, wenn der Downhill-Algorithmus mit variierten Parameterkombinationen systematisch eingesetzt wird.

6.2 Untersuchung der Klassifikationsleistung von Fast and Frugal Trees an dreißig Datensätzen

Um die Leistungsfähigkeit von Algorithmen sinnvoll zu vergleichen, ist die Wahl einer angemessenen Datenumgebung unverzichtbar. Prechelt (1995) kritisiert die Beschränkung auf wenige Datensätze in vielen Studien zum Vergleich von neuronalen Netzwerk-Lernalgorithmen (nur 6% der von ihm untersuchten Studien berichten Ergebnisse von mehr als zwei Datensätzen). Er schlägt ebenfalls eine Mischung aus konstruierten und empirischen Datensätzen vor. Ebenso wird ein Vergleich invaldiert, wenn Datensätze mit bekannter Struktur gewählt werden, die optimal auf eine Teilmenge der zu vergleichenden Algorithmen abgestimmt ist. Weitere Desiderata bilden die Möglichkit der Reproduktion von Simulationen durch Angabe der verwendeten Parameterwerte sowie die Verfügbarkeit der verwendeten Datensätze. Prechelt (1995) wünscht sich zuletzt auch die Veröffentlichung negativer Resultate, damit sich unveröffentlichte Versuche nicht wiederholen müssen.

Der vorliegende Algorithmenvergleich ist auf empirische Datensätze beschränkt, verwendet dafür einen Satz von dreißig Datensätzen, um die notwendige Breite zu erreichen. Die Algorithmen sind spezifiziert, die Datensätze sind größtenteils öffentlich zugänglichen Datensatzsammlungen entnommen. Alle eingesetzten Algorithmen werden in diesem Abschnitt dokumentiert. Die Ergebnisse dieses Teils der Studie gehen auch in zwei geplante Veröffentlichungen ein (Katsikopoulos et al., i. Vorb.; Martignon et al., i. Vorb.). Im Folgenden sollen die insgesamt sechzehn Algorithmen der Konkurrenz vorgestellt werden.

6.2.1 Darstellung der Wettbewerbsteilnehmer

6.2.1.1 Context Model und PROBEX als Beispiele für Exemplar Models

Auch die in diesem Abschnitt vorgestellten *Exemplar models* werden von den Autoren als kompatibel mit dem Konzept *Bounded Rationality* vorgestellt (Juslin & Persson, 2002). An dieser Stelle soll lediglich der zentrale Algorithmus des *Context Model* (Medin & Schaffer, 1978) für die Klassifikation von Objekten vorgestellt werden:

Gegeben sei eine Menge von n Objekten aus zwei Kategorien mit jeweils k binären Cuewerten für jedes Objekt. Ein Objekt, das mit seinen Cuewerten durch eine Person gespeichert worden ist, wird in diesem Ansatz als *Exemplar* bezeichnet. Es sei eine Person gegeben, die n_A *Exemplars* der Kategorie A ($\{\overline{a}_1, \overline{a}_2, \ldots, \overline{a}_{n_A}\}$), und n_B *Exemplars* der Kategorie B ($\{\overline{b}_1, \overline{b}_2, \ldots, \overline{b}_{n_B}\}$) gespeichert hat, wobei jedes Exemplar $\overline{x}_i$ als Vektor seiner Cuewerte definiert ist, also $\overline{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ik})$. Zudem wird für *Exemplars* ein Ähnlichkeitsmaß definiert, das für das Vektorenpaar $\overline{x}$ und $\overline{y}$ den Wert

$$S(\overline{x}, \overline{y}) = \prod_{i=1}^{k} d_i, \; d_i = \begin{cases} 1, & x_i = y_i \\ s, & x_i \neq y_i \end{cases} \tag{6.2}$$

annimmt. Dabei ist $s \in [0; 1]$ ein freier Parameter, über den reguliert werden kann werden kann, wie groß der Einfluss einzelner diskrepanter Merkmale sein soll. Es gilt $S(\overline{x}, \overline{x}) = 1$ und $s^k \leq S(\overline{x}, \overline{y}) \leq 1$. Soll ein Exemplar $\overline{t}$ mit unbekannter Kategorie klassifiziert werden, so werden zwei Wahrscheinlichkeiten $P(A)$ und $P(B)$ mit $(P(A) + P(B) = 1$ berechnet, die Klassifikationswahrscheinlichkeiten des Objekts in jeweils eine der beiden Kategorien angeben mit

$$P(A) = \frac{\sum_{i=1}^{n_A} S(\overline{t}, \overline{a}_i)}{\sum_{i=1}^{n_A} S(\overline{t}, \overline{a}_i) + \sum_{i=1}^{n_B} S(\overline{t}, \overline{b}_i)} \tag{6.3}$$

und

$$P(B) = \frac{\sum_{i=1}^{n_B} S(\overline{t}, \overline{b}_i)}{\sum_{i=1}^{n_A} S(\overline{t}, \overline{a}_i) + \sum_{i=1}^{n_B} S(\overline{t}, \overline{b}_i)} \tag{6.4}$$

Eine deterministische Variante ergibt sich durch Klassifikation des Objekts in die Kategorie, für die die Wahrscheinlichkeit größer ist, bei Zufallskategorisierung im Falle der Gleichwahrscheinlichkeit. Juslin und Persson (2002) weisen darauf hin, dass die Klassifikationswahrscheinlichkeiten im Allgemeinen sowohl durch Ähnlichkeit als auch durch Häufigkeit beeinflusst werden, die Summen werden nicht durch die Anzahl der gespeicherten *Exemplars* jeder Klasse geteilt. Für $s = 1$ hingegen ergibt sich

$$P(A) = \frac{n_a}{n_a + n_b} \tag{6.5}$$

und dementsprechend in der deterministischen Variante eine Einordnung in die Majoritätsklasse, also eine Abhängigkeit der Klassifikation allein von der Häufigkeit. Für $n_a = n_b$ ergibt sich im Gegenzug eine Abhängigkeit allein von der durchschnittlichen Ähnlichkeit für $s < 1$. Bis zu diesem Punkt entspricht das (nicht-deterministische) Modell dem *Context Model* von Medin und Schaffer (1978), das später von Nosofsky (1984, 1986) zum *Generalized Context Model* (GCM) für kontinuierliche Dimensionen erweitert wurde.

Der von Juslin und Persson (2002) entwickelte PROBEX-Algorithmus (PROBabilities from *Exemplars*) beinhaltet einige Modifikationen des *Context Model*. Terminologisch wird hier unterschieden zwischen Objekten in der Umgebung, den Gedächtnisrepräsentationen dieser Objekte (*Exemplars*) und Objektnamen in einer Urteilsaufgabe (Probes). Das Kriterium kann entweder kontinuierlich oder binär sein. Zu den Modifikationen gehört die Einführung der *retrieval probability* p_r, die angibt, mit welcher Wahrscheinlichkeit Cuewerte von Objekten der Umgebung aus der Wissensbasis abgerufen werden können. Das Ähnlichkeitsmaß wird so modifiziert, dass bei unbekannten Wert d_i auf 1 gesetzt wird, wenn beide Werte unbekannt sind, d_i hingegen auf $\sqrt{s}$, wenn genau ein Wert bekannt ist. Weiterhin werden nicht die Ähnlichkeiten zu allen Objekten der Umgebung berechnet, Objekte werden sequenziell aktiviert und verglichen, wobei das Auswahlverfahren einem *Roulette-Wheel*-Verfahren entspricht, bei dem die Taschengröße durch die Ähnlichkeit zwischen Probe und anderen Objekten bestimmt wird, wobei eine *Stopping Rule* die Abfrage von Objekten beendet, wenn die Veränderung von Schätzmaßen zwischen Abfragen einen kritischen Wert

unterschreitet. Die Autoren diskutieren in direktem Bezug auf den Ansatz der ABC-Gruppe die Bedingungen für die ökologische Rationalität der Heuristik und zeigen in einer Städtevergleichsaufgabe ähnlich gute Ergebnisse für PROBEX und *Take The Best*, die sich gegen das Verfahren der multiplen Regression durchsetzen können.

6.2.1.2 Exemplar Models im Wettbewerb

Im Rahmen der Simulation sollen nur Varianten des *Context Model* zum Einsatz kommen: Hierzu werden die Objekte der Trainingsstichprobe als gespeicherte *Exemplars* betrachtet. Dabei wird auch in der *Fitting*-Aufgabe jedes Objekt anhand des *Context Model* klassifiziert, was die etwas seltsam anmutende Konsequenz hat, dass Kriterienwerte für die Aufteilung der *Exemplars* herangezogen werden, die dann als Ergebnis der Modellanwendung vorhergesagt werden. Trotzdem entspricht dieses Vorgehen dem Vorgehen bei anderen Heuristiken und Algorithmen, deren Parameter aufgrund der Trainingsstichprobe geschätzt werden. Das *Fitting*-Ergebnis drückt eher den Grad der Rekonstruierbarkeit aufgrund des Modells aus als eine genuine Vorhersageleistung. Für die Generalisierungsaufgabe werden weiterhin ausschließlich die Objekte der Trainingsstichprobe als *Exemplars* berücksichtigt. Eine Abbruchregel wird nicht verwendet, und alle Cuewerte der Objekte der Trainingsstichprobe werden als bekannt vorausgesetzt. Es werden drei Varianten des *Context Model* zum Einsatz gebracht, die sich durch die Werte des Parameters s unterscheiden: Diese drei Verfahren sollen als **Ex3** ($s = 0,3$), **Ex4** ($s = 0,4$) und **Ex5** ($s = 0,5$) bezeichnet werden. Dabei soll deutlich unterstrichen werden, dass die gewählten Algorithmen nicht die „Referenzklasse“ der Exemplar-Modelle darstellen[1]: Zum einen werden die Prädiktoren nur in binärer Form genutzt, zum anderen ist die Wahl der Parameter nicht notwendigerweise optimal. Alle Prädiktoren werden weiterhin gleichgewichtet. Aufgrund der Einfachheit bieten diese Modelle allerdings einen Vergleichspunkt zu den ähnlich einfachen *Fast and Frugal Trees*. Die Algorithmen wurden in die Testumgebung GIER$_{[1.0]}$.simple_heuristics integriert und im direkten Vergleich mit den Fast and Frugal Trees ausgeführt. Die Prädiktoren werden künstlich mediandichotomisiert, bei Variablen, deren Minimum 0 und deren Maximum 1 ist, wird 0,5 als Trennpunkt angesetzt.

6.2.1.3 CART und Logistische Regression

Der von Breiman et al. (1984) entwickelte CART-Algorithmus (*Classification and Regression Trees*) wird in der Matlab-Implementierung in Matlab 7.0 zum Einsatz gebracht und soll als **CART** bezeichnet werden. CART ist ebenfalls ein baumgenerierendes Verfahren, dessen Bäume allerdings nicht auf eine bestimmte Form beschränkt sind. Die Aufbauprozedur beginnt an der Wurzel und setzt sich je nach Auffindbarkeit eines geeigneten Aufteilungspunktes für einen der Prädiktoren rekursiv bis zu den Endknoten fort, an denen entweder die Fallzahl der Teststichprobe in dieser Gruppe zu klein wird oder kein geeigneter Prädiktor mehr gefunden werden kann (vgl. Woike, 2000). Dem Namen entsprechend können CART-Bäume sowohl für Regressions- als auch für Klassifikationsaufgaben eingesetzt werden. Als Beispiele für zahlreiche praktische Anwendungen von CART-Algorithmen in vielfältigen Bereichen können Barriga, Hamman, Hoag, Marshall und Shetterly (1996), Germanson, Lanzino, Kongable, Torner und Kassell (1998), Markham, Mathieu und Wray (1998), Marx et al. (1998), Krivokapich, Child, Walter und Garfinkel (1999) und Rainer, Lam, Wong und Cocks (1999) dienen. Es wird ein nicht geprunter CART-Baum erstellt bei Verwendung des Gini-Impurity-Maßes. Die Befehlszeilen in der gewählten Implementierung sind in Listing 6.1 präsentiert:

```
t = treefit(daten(:,2:matrixvals(2)),daten(:,1),'method','classification');
YFIT = treeval(t,daten(:,2:matrixvals(2)));
```

Listing 6.1: Matlab-Code für CART

Die Logistische Regression (vgl. Hosmer & Lemeshow, 1989; Montel, 2005) wird - als **LR** bezeichnet - wie zuvor ebenfalls in der in Matlab 7.0 enthaltenen Implementierung zum Einsatz gebracht. Die Befehlszeilen können Listing 6.2 entnommen werden:

[1] Für den Hinweis auf die Möglichkeit dieses Missverständnisses bin ich Henrik Olsson (persönliche Kommunikation, Berlin, 2006) dankbar.

```
[jkb,jkdev,jkstats] = glmfit(daten(:,2:matrixvals(2)),
                        [daten(:,1) unityvec'], 'binomial', 'logit');
[yfitlog] = glmval(jkb,daten(:,2:matrixvals(2)),'logit');
yfitpred(:,1)=(yfitlog > 0.5);
```

Listing 6.2: Matlab-Code für die logistische Regression

Diese beiden Verfahren sollen die „klassischen Auswertungsverfahren“ vertreten. Die Verwendung der logistischen Regression wird in vielen Fällen beinahe als normativ angezeigt betrachtet. Beide Verfahren sind um ein Vielfaches aufwändiger als die einfachen Heuristiken, im Fall der logistischen Regression liegt der Aufwand in der Schätzung der Parameter, beim CART-Algorithmus führen die wiederholten Prüfungen und Tests auf Aufteilung ebenfalls zu einem rechenintensiven Verfahren.

6.2.1.4 Eingesetzte Varianten der Fast and Frugal Trees

Es werden die im vorangegangenen Kapitel beschriebenen elf Varianten der *Fast and Frugal Trees* eingesetzt, die sich lediglich hinsichtlich der Konstruktionsregel unterscheiden. Die fünf ursprünglichen Verfahren sind: FaF MaxSens (**MS**), FaF ZigSens (**ZS**), FaF MaxVal (**MV**), FaF ZigVal (**ZV**) und FaF MaxAcc (**MA**). Die sechs im zweiten Schritt konstruierten *Plus*-, *Majority*- und *Fold*-Varianten der Bäume sind FaF ZigSens Plus (**SP**), FaF ZigSens Majority (**SM**), FaF ZigSens Fold (**SF**), FaF ZigVal Plus (**VP**), FaF ZigVal Majority (**VM**) und FaF ZigVal Fold (**VF**). Auch für die Bäume werden die Prädiktoren mediandichotomisiert. Für das Schätzen der Baumstruktur und für die Bestimmung der Cuerichtungen steht ausschließlich die Trainingsstichprobe zur Verfügung.

6.2.2 Datensätze und Wettbewerbsmodus

In diesem Abschnitt sollen zum einen die Datensätze des Wettbewerbs und zum anderen der Modus ihrer Vorgabe und des Vergleichs dargestellt werden.

6.2.2.1 Überblick über die verwendeten Datensätze

Für den Vergleich der Algorithmen ist eine Testsuite aus dreißig Datensätzen konstruiert worden, die sich hinsichtlich der Bereiche, aus denen sie stammen, der Prädiktorenart und -anzahl sowie der Stichprobengröße hinreichend unterscheiden. Bei einigen Datensätzen war es notwendig, ein mehrkategoriales Kriterium durch Zusammenfügen von Kategorien zu binarisieren. Datensätze mit fehlenden Daten wurden teilweise eliminiert.

Die verwendeten Datensätze sollen im Einzelnen kurz vorgestellt werden (vgl. Tabelle 6.9). Sechzehn Datensätze stammen dabei aus dem *UCI Machine Learning Repository* (Asuncion & Newman, 2007). Dies sind die folgenden Datensätze:

- *Echocardiogram*: Kriterium in diesem Datensatz ist das Überleben von Patienten nach einem Herzinfarkt (12 Monate oder mehr), dazu werden sieben Prädiktoren genutzt, unter anderem Kardiogramm-Daten. Im Datensatz sind 62 Fälle (Referenz: Salzberg, 1988).
- *Post-Operative*: In diesem Datensatz wird die Entscheidung vorhergesagt, einen Patienten nach der Operation aus dem Krankenhaus zu entlassen. Dazu stehen acht Prädiktoren zur Verfügung, der Datensatz enthält 86 Fälle (Referenz: Woolery, Grzymala-Busse, Summers & Budihardjo, 1991).
- *Zoo*: Hier soll in die Kategorien „Säugetier“ und „kein Säugetier“ klassifiziert werden, zu den sechzehn Prädiktoren gehören Angaben zu den Eigenschaften von 101 Tieren (Quelle: Richard S. Forsyth).
- *Iris*: Zu klassifizieren sind hier Pflanzen in verschiedene Iris-Varianten. Dabei soll das Vorliegen einer der drei Arten bestimmt werden unter Verwendung von vier Pflanzeneigenschaften. Es gibt 150 Pflanzen im Datensatz (Referenz: Fisher, 1936).
- *Hepatitis*: Vorhergesagt wird das Überleben von Hepatitis-Patienten aufgrund von sieben Prädiktoren, zu denen demografische Informationen und Symptome gehören. Im Datensatz sind 153 Fälle (Referenz: Cestnik, Konenenko & Bratko, 1987).

Tabelle 6.9: Überblick über 30 Datensätze der Testsuite: Angegeben sind jeweils die Fallzahlen, die Kategorienauteilung, der Anteil der kleineren Kategorie und die Aufteilung der Prädiktoren auf binäre, kategoriale und kontinuierliche Variablen. Als Quelle ist die Herkunft des Datensatzes angegeben (UCI: *UCI Machine Learning Repository*, ACD: *Analyis of Categorical Data*, O: Andere Quelle, siehe Text).

	Beschreibung					Prädiktorart			Quelle
Datensatz	n	Cues	n_1	n_0	kleinerer Anteil	bin.	kat.	kont.	
Bankruptcy	50	5	25	25	0,500	-	-	5	ACD
Solvency	52	8	27	25	0,481	-	-	8	ACD
Echocardiogram	62	7	18	44	0,290	1	-	6	UCI
Post-Operative	86	8	24	62	0,279	-	7	1	UCI
Broadway	89	5	27	62	0,303	-	3	2	ACD
Credit Score	100	6	73	27	0,270	2	1	3	ACD
Zoo	101	16	41	60	0,406	15	1	-	UCI
Iris	150	4	50	100	0,333	-	-	4	UCI
Hepatitis	153	7	121	32	0,209	6	-	1	UCI
Strikes	163	4	86	77	0,472	2	-	2	ACD
Wine	178	13	71	107	0,399	-	-	13	UCI
Baseball CYY	189	9	43	146	0,228	2	1	6	ACD
Forensic Glass	214	9	146	68	0,318	-	-	9	UCI
Horse Colic	218	6	128	90	0,413	1	1	4	UCI
SPECT	267	22	212	55	0,206	22	-	-	UCI
Heart (Hungaria)	294	10	106	188	0,361	1	4	5	UCI
Heart (Cleveland)	297	13	137	160	0,461	3	5	5	UCI
Alcohol	345	5	88	257	0,255	-	-	5	UCI
Ionosphere	351	33	225	126	0,359	-	-	33	UCI
Tourism	365	4	85	280	0,233	1	-	3	ACD
Voting	402	16	248	154	0,383	-	16	-	UCI
Education	435	4	211	224	0,485	2	-	2	O
Cardiac	558	8	469	89	0,159	1	-	7	O
Football	672	4	451	221	0,329	1	-	3	O
Breast Cancer	683	9	444	239	0,350	-	9	-	UCI
Diabetes	786	6	268	500	0,349	-	-	6	UCI
Authorship	841	69	173	668	0,259	-	-	69	ACD
Baseball HoF	1320	15	124	1196	0,094	-	-	15	ACD
Yeast	1484	8	463	1021	0,312	-	-	8	UCI
Supreme Court	4052	7	170	3882	0,042	5	1	1	ACD
Mittelwert (gerundet)	499	11	-	-	0,318	2	2	8	-

- *Wine*: Anhand von dreizehn chemischen Analyseergebnissen soll die Herkunft eines Weines bestimmt werden. Dabei wird vorhergesagt, ob eine Probe aus einem bestimmten von drei Herkunftsorten stammt oder nicht. Es sind 178 Fälle in der Stichprobe (Referenz: Aeberhard, Coomans & de Vel, 1992).
- *Forensic Glass*: In diesem Datensatz soll klassifiziert werden, ob eine Glasprobe Fensterglas ist oder zu einer von fünf anderen Glaskategorie gehört. Acht Prädiktoren werden durch die Bestimmung der Bestandteile gebildet. Es sind 214 Proben in der Stichprobe (Quelle: B. German).
- *SPECT*: Aufgrund der Merkmale einer *Single Proton Emission Computed Tomography (SPECT)* soll das Vorliegen eines Herzproblems diagnostiziert werden, dabei stehen 22 Prädiktoren bei 267 Fällen zur Verfügung (Referenz: Kurgan, Cios, Tadeusiewicz, Ogiela & Goodenday, 2001).
- *Heart (Hungaria)*: In diesem Satz wird das Vorliegen einer Herzkrankheit aufgrund von zehn Prädiktoren vorhergesagt bei 294 Fällen (Referenz: Detrano et al., 1989).
- *Heart (Cleveland)*: Dieser Datensatz enthält eine andere Prädiktorenmenge als der *Heart(Hungaria)* Datensatz, aber die gleiche Klassifikationsaufgabe. Mit dreizehn Prädiktoren werden 297 Fälle klassifiziert (Referenz: Detrano et al., 1989).
- *Alcohol*: Als Kriterium dient hier der Alkoholkonsum eines Patienten, klassifiziert in die beiden Gruppen „mehr als 2,5 Pints pro Tag“ oder „weniger als 2,5 Pints pro Tag“. Dazu stehen fünf diagnostische Prädiktoren bei 345 Fällen zur Verfügung (Quelle: BUPA Medical Research Ltd.).
- *Ionosphere*: In diesem Datensatz geht um die Klassifikation von „guten“ und „schlechten“ Radarergebnissen, wobei „gute“ Radarergebnisse auf Strukturiertheit von Signalen in der Ionosphäre hindeuten. Mit 33 Prädiktoren werden 351 Signale klassifiziert (Referenz: Sigillito, Wing, Hutton & Baker, 1989).
- *Voting*: Anhand von Antworten auf sechzehn Fragen wird die Wahlentscheidung (demokratisch vs. republikanisch) vorhergesagt bei 402 Personen (Referenz: Schlimmer, 1987).
- *Breast Cancer*: Aufgrund von diagnostischen Zellenwerten werden Tumore als gutartig oder bösartig klassifiziert. Dazu können neun Prädiktoren bei 683 Fällen genutzt werden (Referenz: Mangasarian & Wolberg, 1990).
- *Diabetes*: Es werden Fälle von Diabetes klassifiziert, dafür können sechs Prädiktoren verwendet werden, 786 Personen sind einzuteilen (Quelle: National Institute of Diabetes and Digestive and Kidney Diseases).
- *Yeast*: Vorhergesagt wird die Lokalisierungsstelle von Proteinen. Dazu wird in die Kategorien „zytosolisch oder zytoskeletal“ und „andere Stelle“ klassifiziert. Es gibt acht Prädiktoren und 1484 Fälle (Referenz: Horton & Nakai, 1996).

Zehn Datensätze entstammen den Begleitdatensätzen zu Simonoffs *Analyzing Categorical Data* (Simonoff, 2003)[2]. Dies sind die folgenden Klassifikationsaufgaben:

- *Bankruptcy*: Aufgrund von fünf finanziellen Indikatoren wird der Bankrott von Firmen vorhergesagt, es gibt fünfzig Fälle (Referenz: Simonoff, 2003).
- *Solvency*: Die Solvenz von Firmen wird aufgrund von acht Prognosewerten bestimmt bei 52 Fällen (Referenz: Simonoff, 2003).
- *Broadway*: Hier soll klassifiziert werden, ob ein Broadway-Musical sechs Monate oder mehr fortgesetzt wird. Dazu stehen Ratingdaten und Ticketverkäufe zur Verfügung. Im Datensatz sind 89 Musicals (Referenz: Simonoff, 2003).
- *Credit Score*: Vorhergesagt wird die Annahme von Kreditanträgen unter Berücksichtigung von Angaben zum Antragsteller. Es sind 100 Anträge in der Stichprobe (Referenz: Simonoff, 2003).

[2]Die Datensätze wurden von der Seite http://www.stern.nyu.edu/~jsimonof/AnalCatData.html heruntergeladen.

- *Strikes*: Vorhergesagt wird die Anzahl von Streiks bei Verwendung von Informationen zum möglichen Streikfall. Vier Prädiktoren und 163 Fälle bilden den Datensatz (Referenz: Simonoff, 2003).
- *Baseball CYY*: In diesem Datensatz stellt das Kriterium die Vergabe des CY Young-Awards für Baseball-Pitcher dar. Neun Prädiktoren zu den Leistungen von 189 Sportlern werden berücksichtigt (Referenz: Simonoff, 2003).
- *Tourism*: Aufgrund von demografischen Daten und finanziellen Informationen soll vorhergesagt werden, ob Reisen unternommen werden oder nicht. Vier Prädiktoren bei 365 Fällen sind enthalten (Referenz: Simonoff, 2003).
- *Authorship*: Angaben zu 69 Worthäufigkeiten in Texten sollen zur Bestimmung genutzt werden, ob der Autor eines Textes Shakespeare ist oder nicht. Insgesamt sollen 841 Texte klassifiziert werden (Referenz: Simonoff, 2003).
- *Baseball HoF*: Spielstatistiken dienen zur Bestimmung, ob ein Baseballspieler in die *Hall of Fame* aufgenommen wird oder nicht. 15 Prädiktoren werden für die Klassifikation von 1320 Spielern eingesetzt (Referenz: Simonoff, 2003).
- *Supreme Court*: Hier soll bestimmt werden, ob Handlungen des Abgeordnetenhauses oder Senats als Antwort auf Entscheidungen des amerikanischen obersten Gerichtshofs unternommen werden. Als Prädiktoren dienen sieben Eigenschaften der Entscheidungen. Im Datensatz sind 4052 Entscheidungen aufgeführt (Referenz: Eskridge, 1991).

Drei weitere Datensätze komplettieren die Sammlung, die aus anderen Quellen hinzugefügt wurden:

- *Education*: Im Datensatz sind Informationen zu 435 irischen Schulkindern im Alter von elf Jahren. Die Kinder sollen klassifiziert werden hinsichtlich des Erwerbs eines Schulabschlusszertifikats, Prädiktoren sind Schulform, ein Testergebnis, das Geschlecht und das Prestige des vom Vater ausgeübten Berufs (Referenz: Greaney & Kelleghan, 1984).
- *Cardiac*: In diesem Datensatz soll eine Herzerkrankung vorhergesagt werden durch Verwendung demografischer und diagnostischer Informationen. Es liegen Daten von 558 Patienten vor mit acht Prädiktoren[3] (Referenz: Garfinkel & Mitarbeiter, 1999).
- *Football*: In diesem Datensatz sind Football-Spiele der NFL-Liga dahingehend zu klassifizieren, ob das Heimteam einen Sieg erringt oder nicht. Als Prädiktoren dienen Wettquoten und Informationen über die Spiele. Zu 672 Spielen liegen Informationen auf vier Prädiktoren vor[4] (Quelle: Hal Stern).

Mit diesen dreißig Datensätzen ist ein ausführlicherer Methodenvergleich möglich. Es sollte darauf hingewiesen werden, dass sich die Struktur der Datensätze von den zuvor diskutierten Datensätzen dahingehend unterscheidet, dass die Prädiktoren nicht auf Werte zwischen 0 und 100 festgelegt sind, teilweise auch kontinuierliche Werte annehmen können.

6.2.2.2 Turniermodus

Um einen Trade-Off herzustellen zwischen unzumutbaren Laufzeiten und instabilen Ergebnissen wurde folgende Lösung gewählt. Jeder Datensatz wird in 19 verschiedenen Aufteilungen genutzt, die sich aus unterschiedlichen Größenverhältnissen von Trainings- und Teststichprobe ergeben. Bei jedem Datensatz erfolgt die Aufteilung relativ zur Ursprungsgröße, nicht absolut. Die kleinste Trainingsstichprobe enthält dabei 5% der Fälle bei einer entsprechenden Teststichprobe mit den 95% restlichen Fällen. Die größte Trainingsstichprobe enthält 95% der Fälle bei einer Teststichprobe mit 5% der Fälle. Die Größe der Trainingsstichprobe wird in 5%-Inkrementschritten variiert, so dass 19 verschiedene Aufteilungen genutzt werden können. Es wird jede Aufteilung jedem Verfahren der Konkurrenz bei jedem der dreißig Datensätze in 1000 Trials vorgegeben, so dass jedem Verfahren $30 \times 19 \times 1000 = 570000$ Klassifikationsaufgaben

[3] Dieser Datensatz wurde von der Seite des *UCLA Department of Statistics*, http://www.stat.ucla.edu/datasets, heruntergeladen.

[4] Dieser Datensatz wurde der *StatLib*-Sammlung entnommen, http://lib.stat.cmu.edu/datasets/.

gestellt werden. In jedem Trial wird allein die Trainingsstichprobe für die Schätzung von Parametern oder die Konstruktion von Bäumen genutzt, die Teststichprobe zur Überprüfung des Ergebnisses. Dabei wird in jedem Trial sowohl ein *Fitting*- als auch ein *Prediction*-Ergebnis erhalten, somit gibt es pro Verfahren 1 140 000 Einzelergebnisse. Insgesamt nehmen 16 Verfahren an der Konkurrenz teil, so dass in der Simulation insgesamt 18 240 000 Ergebnisse produziert werden. Für jedes Ergebnis wird die Klassifikationsgenauigkeit A berechnet, die als Gütemaßstab im Turnier dienen soll. Der Übersichtlichkeit halber sollen die Ergebnisse von jeweils 1000 strukturidentischen Trials als Mittelwert interpretiert werden. Die Trialanzahl ermöglicht dabei eine hinreichend genaue Schätzung des Durchschnittswertes aller möglichen Aufteilungen bei einem gegebenen Trainings-Test-Verhältnis. Gleichzeitig ermöglicht diese Vorgehensweise einen Einblick in die Leistungsfähigkeit der Verfahren bei unterschiedlich großen Stichproben. Dabei werden die Berechnungen für die Verfahren **CT** und **LR** in Matlab durchgeführt, die Berechnungen für die siebzehn übrigen Verfahren in $\text{GIER}_{[1.0]}$.simple_heuristics.

Es muss angemerkt werden, dass bei der gewählten Aufteilung gerade bei kleinen Datensätzen oder Datensätzen mit vielen nichtkontinuierlichen Prädiktoren nicht in jedem Trial ein Ergebnis berechnet werden konnte, da im Fall der logistischen Regression Singularitäten in Matrizen zu Fehlermeldungen führten oder im Fall der Bäume zu wenige Fälle zur Berechnung zur Verfügung standen. Der Matlab-Lauf war dabei sensibler als die Baumsoftware, so dass bei einer zu hohen Ausfallquote ($> 25\%$) die entsprechenden Angaben vollständig eliminiert wurden. Damit sind bei kleinen Stichproben die Mengen der berücksichtigten Datensätze nicht identisch, was nur für den Vergleich zwischen Bäumen und Konkurrenzverfahren eine Rolle spielen wird. Die Simulationsstudien wurden im Verlauf von mehreren Monaten teilweise auf PCs und Laptops und teilweise in einem CITRIX-Netzwerk durchgeführt[5].

6.2.3 Ergebnisse des Wettbewerbs

Eine ausführliche Darstellung aller erhaltenen Ergebnissen befindet sich in Anhang D. Der Übersichtlichkeit halber sollen an dieser Stelle nicht alle 19 380 Ergebnisse präsentiert werden, sondern eine begründete Auswahl. In der Darstellung der Ergebnisse sollen zunächst die Ergebnisse vorgestellt werden, die den Vergleich der elf verschiedenen Konstruktionsverfahren ermöglichen. Im Anschluss wird ein Vergleich mit den Konkurrenzverfahren durchgeführt.

6.2.3.1 Ergebnisse hinsichtlich der Wahl des Klassifikationsverfahrens

Die Ergebnisse der fünf *Fast and Frugal Trees* mit den ursprünglichen Konstruktionsverfahren sind in Abbildung 6.20 zusammengefasst, wobei über alle dreißig Datensätze gemittelt wurde. *Fitting*- und *Prediction*-Ergebnisse für jedes der fünf Verfahren jeweils in einem Diagramm sind für jede der 19 Aufteilungen ersichtlich, ein sechstes Diagramm zeigt den direkten Vergleich der Verfahren hinsichtlich der Generalisierungsleistung. Jeder Datenpunkt in den Diagrammen repräsentiert dabei den Durchschnittswert aus $30 \times 1000 = 30\ 000$ Trials.

Bei allen Verfahren fällt ein *Overfitting* auf, das erwartungsgemäß mit der Trainingsstichprobengröße abnimmt, bis ca. 25% sehr schnell, dann langsamer, ab 65% scheint die Differenz zwischen *Fitting*- und *Prediction*-Ergebnissen auf einem ähnlichen Niveau zu bleiben. Die Verfahren, die auf Validität beruhen, zeigen bessere Ergebnisse als die Verfahren, die auf Sensitivität beruhen, FaF MaxVal kann sich gegen die anderen Verfahren durchsetzen, die ZigZag-Verfahren liegen zudem hinter FaF MaxAcc zurück. FaF MaxSens fällt dabei deutlich ab und aus der Konkurrenz heraus.

Das vergleichsweise gute Abschneiden von FaF MaxAcc ist zunächst überraschend, wenn berücksichtigt wird, dass Entscheidungen auf für die Verwendung eines Cues irrelevanten Fällen basieren: Scheinbar sind Sensitivität und Spezifität der Cues günstig verteilt, so dass sich dieser Umstand nicht als größeres Problem darstellt. Das bessere Abschneiden der validitätsbasierten Verfahren hängt, wie im vorangegangenen Kapitel diskutiert, sehr wahrscheinlich mit dem Umstand zusammen, dass die Validitäten im Gegensatz zu Sensitivitäten und Spezifitäten direkt für die Verwendung im Baum relevant sind. In der einfachen Variante können die ZigZag-Verfahren - wie demonstriert - nicht überzeugen. Allerdings waren mögliche Probleme der Algorithmen diskutiert worden, weshalb ebenfalls die modifizierten ZigZag-Verfahren getestet wurden. Dabei sind folgende Erwartungen plausibel:

[5] Für die Studie konnten teilweise die Ressourcen des Max-Planck-Instituts für Bildungsforschung in Berlin genutzt werden. Die reine Laufzeit der Algorithmen (ohne Auswertungen und Einrichtung) bewegt sich im Bereich von zahlreichen Wochen, weshalb die Studie ohne diese Unterstützung nicht möglich gewesen wäre.

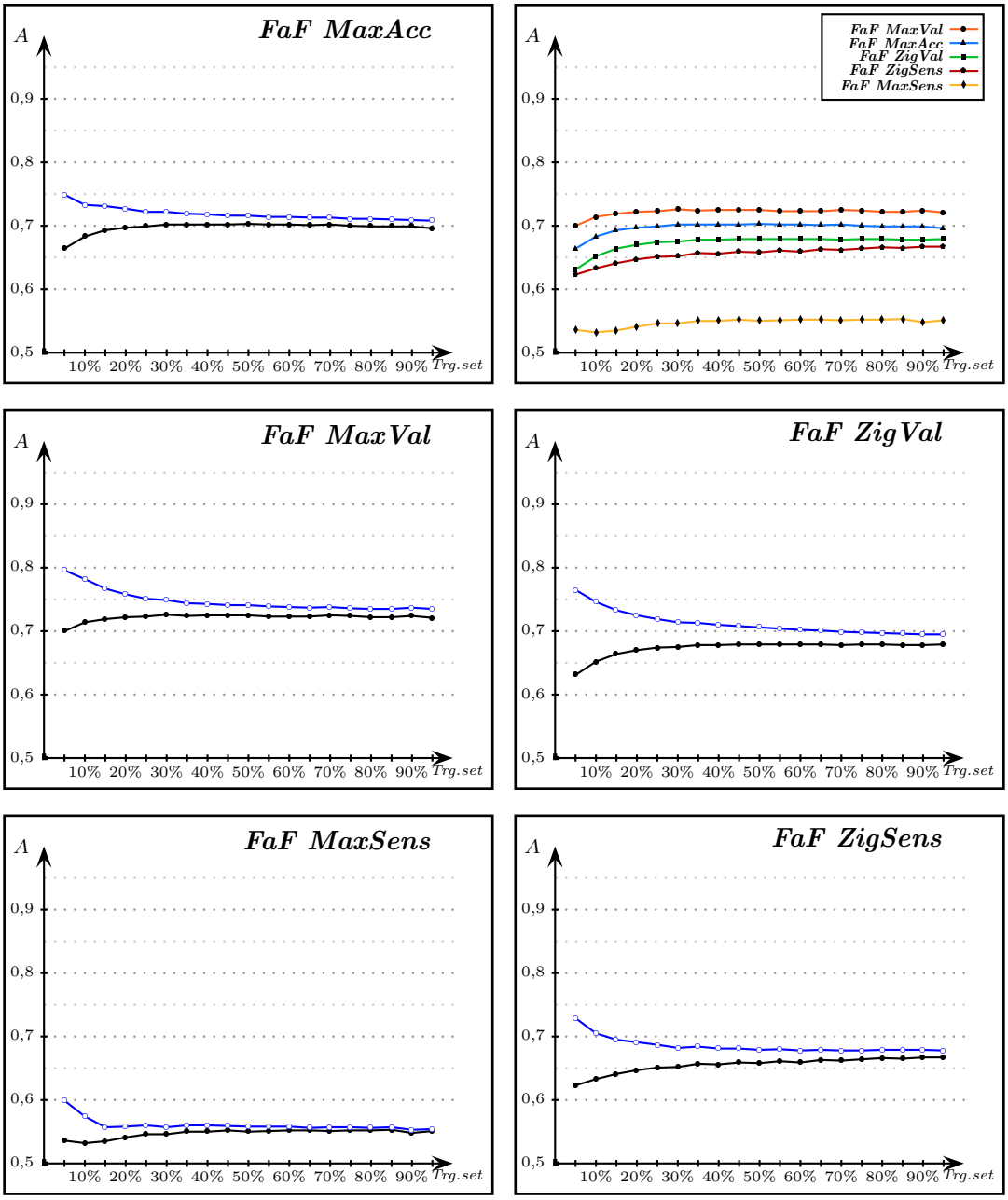


Abbildung 6.20: Fitting- und Prediction-Ergebnisse der *Fast and Frugal Trees* mit den fünf ursprünglichen Konstruktionsverfahren: In jedem der fünf Diagramme sind die Fitting- und die Prediction-Ergebnisse der Verfahren eingezeichnet, aufgeteilt nach Trainingsstichprobengrößen, dabei zeigt die blaue Linie die Fitting-Ergebnisse, die schwarze Linie die Prediction-Ergebnisse. Im Diagramm rechts oben sind die Prediction-Ergebnisse aller Verfahren zusammengetragen.

- Bei ungleichen Klassenverteilungen und mediandichotomisierten Cues ergibt sich bei den Validitäten (und nicht notwendigerweise bei Sensitivitäten und Spezifitäten) die Eigenschaft, dass die auf die kleinere Klasse bezogenen Validitäten im Allgemeinen kleiner sein werden als die auf die größere Klasse bezogenen Validitäten, da bei Gleichverteilung positiver und negativer Cuewerte und Ungleichverteilung von positiver und negativer Klassenzugehörigkeit für eine der beiden Validitäten eine obere Grenze unterhalb von 1 existiert.
- In der Konsequenz sind sich das FaF ZigVal Plus und das FaF ZigVal Majority-Verfahren sehr ähnlich, da die Entscheidung für die größere Klasse und die Entscheidung für den höchsten Validitätswert im Allgemeinen in die gleiche Richtung weisen sollten.
- Für die ZigSens-Verfahren ergibt sich ein anderes Problem: Aufgrund der Überlegungen zu differenzieller Streuung der Schätzwerte für Sensitivität und Spezifität eines Cues wird eine Bevorzugung der kleineren Klasse zu erwarten sein, was sich gegenüber einer Setzung der ersten Entscheidungsrichtung sogar eher nachteilig auswirken könnte.
- Das ZigSens Majority-Verfahren sollte hingegen gegenüber dem ZigSens Plus-Verfahren besser abschneiden, da hier die Wahl der größeren Klasse sichergestellt wird und deshalb keine „Default-Fehler“ im ersten Schritt zu erwarten sind, die sich bei Gleichverteilung der Cuewerte und Wahl der kleineren Klasse ergeben.
- Die Fold-Varianten sollten gegenüber den Majority-Varianten bessere Ergebnis erwarten lassen.

Das Erwartungsmuster findet sich in den Ergebnissen, die in Abbildung 6.21 zusammengetragen sind. Das Overfitting nimmt eine ähnliche Form und ähnliche Ausmaße an wie zuvor. Ein Vergleich der *Prediction*-Leistung der sechs Algorithmen mit den beiden ursprünglichen ZigVal-Verfahren zeigt Folgendes:

- FaF ZigSens und FaF ZigSens Plus liegen ungefähr auf gleicher Höhe, wenn überhaupt schneidet FaF ZigSens Plus schlechter ab, vor allem bei sehr kleinen Stichproben. Dies entspricht den Erwartungen.
- Die Ergebniskurven von FaF ZigVal Plus und FaF ZigVal Majority sind beinahe deckungsgleich, mit beiden Regeln wird im Sinne der Erwartungen das gleiche Ziel erreicht.
- Die Fold-Varianten sind in beiden Fällen die besten Varianten, ungleiche Klassengrößen scheinen damit eines der Hauptprobleme der Algorithmen zu sein, und die Fold-Regel wird den Erfordernissen am besten gerecht.
- Die FaF ZigVal Fold-Variante hat ein besseres Ergebnis als die FaF ZigSens Fold-Variante, was ähnlich erklärt werden kann wie bei den Prinzipienvergleichen zuvor: Der Validitätskennwert entspricht der Entscheidungsstruktur des Baumes besser als die Sensitivität/Spezifität.

Bevor eine detailliertere Analyse der Einzelergebnis unternommen werden kann, sollen die Ergebnisse der Konkurrenzalgorithmen betrachtet werden.

6.2.3.2 Ergebnisse hinsichtlich des Vergleichs mit konkurrierenden Algorithmen

Die Ergebnisse der Konkurrenten sind in Abbildung 6.22 dargestellt. Dabei muss angemerkt werden, dass für LR und CART bei kleinen Stichproben einige Datensätze nicht verarbeitet werden konnten, was den unregelmäßigen Verlauf der beiden Kurven erklären kann.

Bei der Betrachtung der Einzelkurven fällt das hohe *Overfitting* von LR und CART auf sowie das geringe *Overfitting* bei den Exemplar-Modellen. Die sehr einfachen Exemplar-Modelle werden allerdings von den übrigen Verfahren - wie erwartet - klar geschlagen. Die Parameterwahl hat einen Einfluss auf das Ergebnis, auf eine weitere Variation wird aber verzichtet. Irregulär ist der Verlauf der Kurve allein für CART, was durch zwei Umstände erklärt werden kann: die fehlenden Datensätze für die kleinsten Stichprobengrößen und den Umstand, dass CART für die rekursive Aufspaltungsprozedur von entsprechenden Fallzahlen abhängig ist, die eine Mehrfachaufteilung der Stichprobe erlauben. Markant ist weiterhin aber auch die rechte Seite der Kurven: Nimmt das beträchtliche Overfitting für LR mit steigender Stichprobengröße ab - auch wenn eine größere Differenz bei 95% verbleibt als bei den Bäumen - bleibt es für CART nahezu in gleicher Höhe.

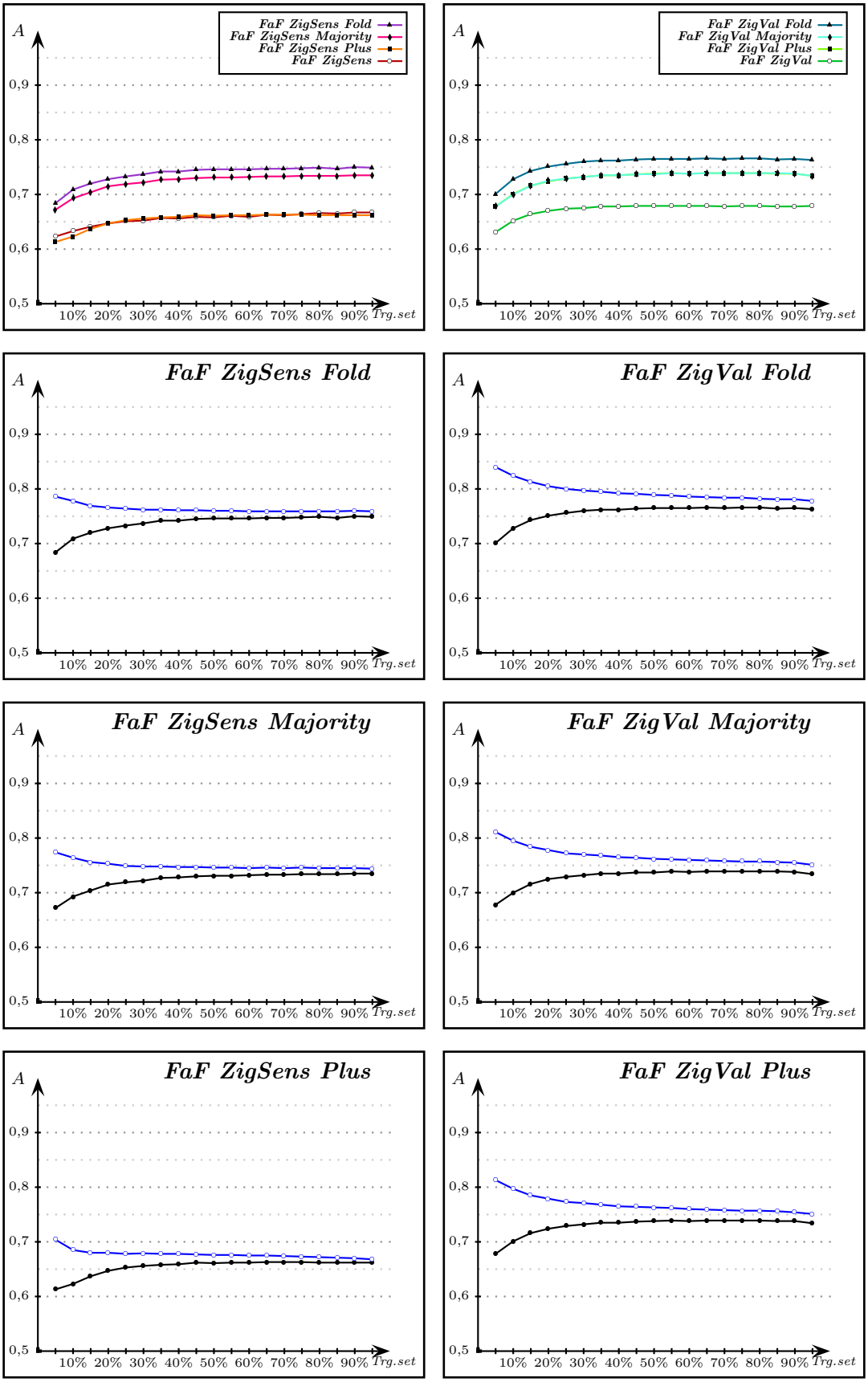


Abbildung 6.21: Ergebnisse der *Fast and Frugal Trees* mit modifizierten Konstruktionsverfahren: In den Diagrammen der unteren drei Reihen sind die Fitting-Ergebnisse (blau) und die Prediction-Ergebnisse (schwarz) der Verfahren eingezeichnet; die Diagramme in der oberen Reihe zeigen einen Vergleich der Prediction-Ergebnisse der auf Sensitivität/Spezifität beruhenden Verfahren (oben links) und der auf Validität beruhenden Verfahren (oben rechts).

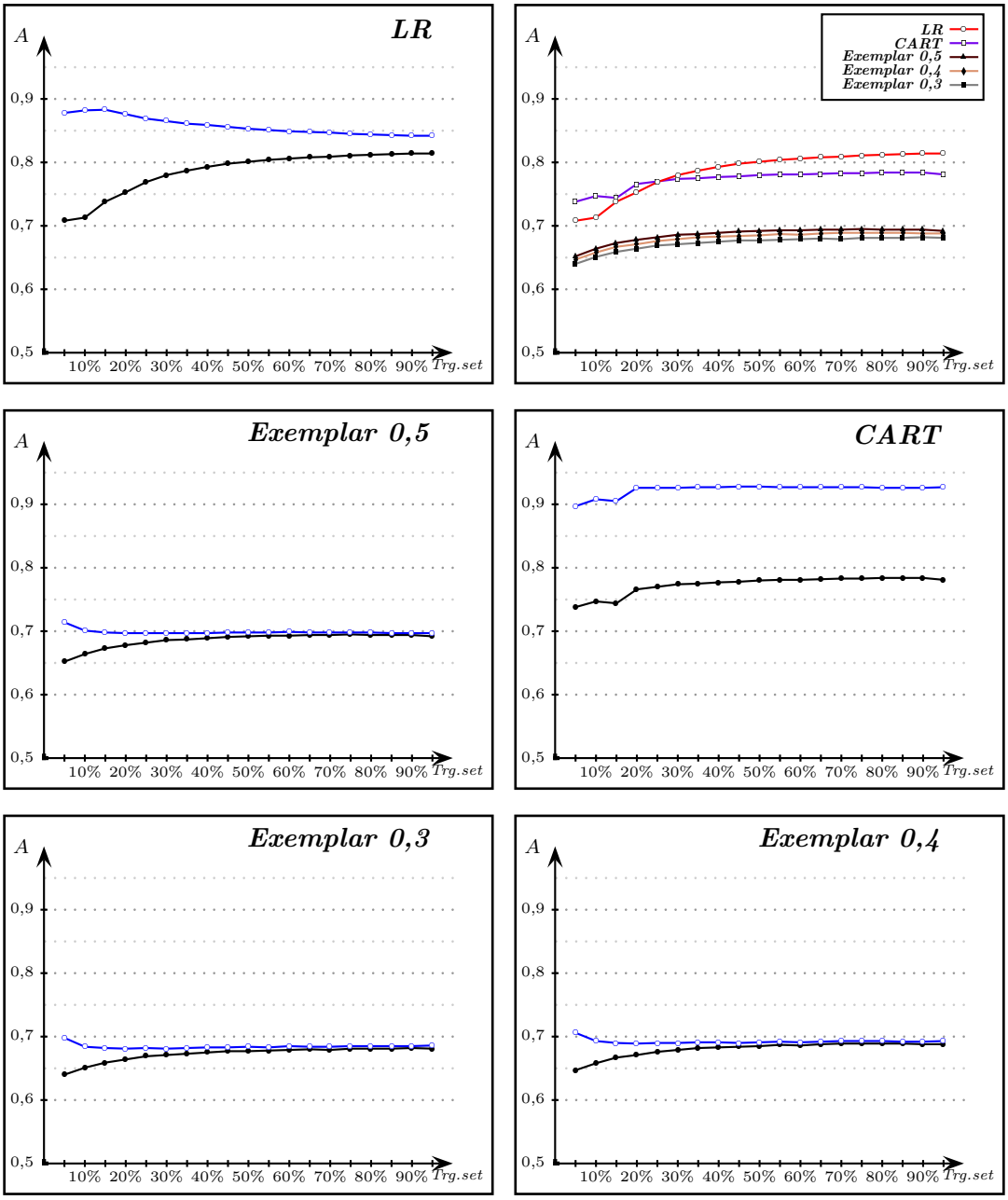


Abbildung 6.22: Ergebnisse der konkurrierenden Algorithmen: In den Diagrammen der unteren drei Reihen sind die Fitting-Ergebnisse (blau) und die Prediction-Ergebnisse (schwarz) der Verfahren - gemittelt über dreißig Datensätze und jeweils 1000 Trials - eingezeichnet; im Diagramm oben rechts sind die fünf Prediction-Ergebnisse zusammengetragen.

Es bleibt ein Vergleich der besten Ergebnisse: In Abbildung 6.23 ist eine Auswahl der Algorithmen für den direkten Vergleich zusammengestellt. Auffallend ist die deutliche Abhängigkeit von der Stichprobengröße[6]. Bei ca. 25% kann LR sich vor CART platzieren und das Feld dominieren. Bei sehr kleinen Stichproben hat CART das beste Ergebnis, allerdings ist FaF ZigVal Fold durchaus konkurrenzfähig und kann bei 15% nicht deutlich geschlagen werden. Berücksichtigt man die Heterogenität der Datensatzmengen bei kleineren Stichproben und die sehr unterschiedliche Komplexität der Verfahren, ist dieses Ergebnis durchaus bemerkenswert.

Eine detailliertere Analyse soll untersuchen, ob sich die gefundene Ordnung der Algorithmen über die Datensätze als stabil erweist, um zu Auswahlempfehlungen für Klassifikationsprobleme zu gelangen.

6.2.4 Hinweise auf ökologische Rationalität von Fast and Frugal Trees

Zum Vergleich der Algorithmen sei eine einfachere Darstellung vorgestellt, die sich auf vier Datenpunkte aus den oben gewählten Diagrammen beschränkt: Zum einen wird das *Fitting*-Ergebnis dargestellt, das an einer Trainingsstichprobe von 95% gewonnen wurde, zum anderen werden drei *Prediction*-Ergebnisse abgebildet, und zwar für 90%, 50% und 15% (bei einer Trainingsstichprobe von 90% bleibt bspw. eine Teststichprobe von 10%). Es wird keine kleinere Trainingsstichprobe gewählt, da hier das Problem heterogener Datensatzmengen aufträte. Mit diesen vier Punkten können die bisher beobachteten Diagramme im Wesentlichen beschrieben werden, die Darstellung als Balkendiagramm soll die Lesbarkeit verbessern. Der Vergleich beschränkt sich auf vier Algorithem: CART und LR wurden als die bisher leistungsstärkten Algorithmen der Konkurrenz gewählt, zum anderen wurden zwei Baumverfahren ausgewählt: FaF ZigVal Fold und FaF MaxVal. Diese Auswahl kann damit begründet werden, dass die validitätsbasierten Verfahren besser abschneiden als die übrigen Verfahren und FaF ZigVal Fold das bisher beste Baumverfahren darstellt. FaF MaxVal wurde aufgrund des fundamental unterschiedlichen Konstruktionsverfahren hinzugenommen. Über alle 30 Datensätze kann das durschnittliche Ergebnis der Algorithmen damit in Abbildung 6.24 präsentiert werden.

Dem Diagramm kann sowohl der *Overfitting*-Effekt entnommen werden als auch der Reihenfolgewechsel der Verfahren bei kleineren Trainingsstichproben. Weiterhin kann nachvollzogen werden, dass der Vorsprung der klassischen und rechenintensiven Verfahren gegenüber den Bäumen bei sinkender Stichprobengröße abnimmt.

6.2.4.1 Differenzierte Darstellung der Klassifikationsleistungen

Eine nach Datensätzen differenzierte Leistung der Algorithmen in den vier gewählten Kategorien ist in den Tabellen 6.10, 6.11 und 6.12 aufgeführt. Der Überblick erbringt die Beobachtung, dass die Reihenfolge der Verfahren über die Datensätze hinweg nicht stabil bleibt. Damit ist es ebenfalls möglich, Hinweise zu gewinnen auf die ökologische Rationalität der *Fast and Frugal Trees*.

6.2.4.2 Medizinische Datensätze

Die hier vorgestellten Klassifikationsverfahren wurden ursprünglich im medizinischen Bereich entwickelt, was zur Frage führt, ob sich diese Verfahren in besonderer Weise in diesem Bereich beweisen können. In der Menge der Datensätze befinden sich elf Datensätze, die aus dem (tier-)medizinischen Bereich stammen. Es sind der Stichprobengröße nach: *Echocardiogram*, *Post-Operative*, *Hepatitis*, *Horse Colic*, *SPECT*, *Heart (Hungaria)*, *Heart (Cleveland)*, *Alcohol*, *Cardiac*, *Breast Cancer* und *Diabetes*.

Die Daten wurden für diese Teilgruppe der Datensätze noch einmal gesondert analysiert. Die Ergebnisse sind in Abbildung 6.25 zusammengefasst. Hier fällt auf, dass der CART-Algorithmus schlechtere Ergebnisse erbringt als im vorherigen Vergleich und dass bei kleinen Datensätzen die besten Ergebnisse von FaF ZigVal Fold demonstriert werden, wobei der Baum in jeder Prediction-Aufgabe bessere Ergebnisse erbringt als CART. Bei kleinen Stichproben zeigt zudem FaF MaxVal keine schlechteren Ergebnisse als LR und CART. Damit bestätigt sich der Verdacht, dass medizinische Stichproben ein geeigneter Anwendungsbereich für die Bäume sein könnten, wobei auch innerhalb dieser Teilmenge noch heterogene

[6] Bei einer Leave-One-Out-Kreuzvalidierung wäre ein Ergebnis nahe dem rechten Rand erhalten worden, was den Blick auf diese Dynamik verstellt hätte.

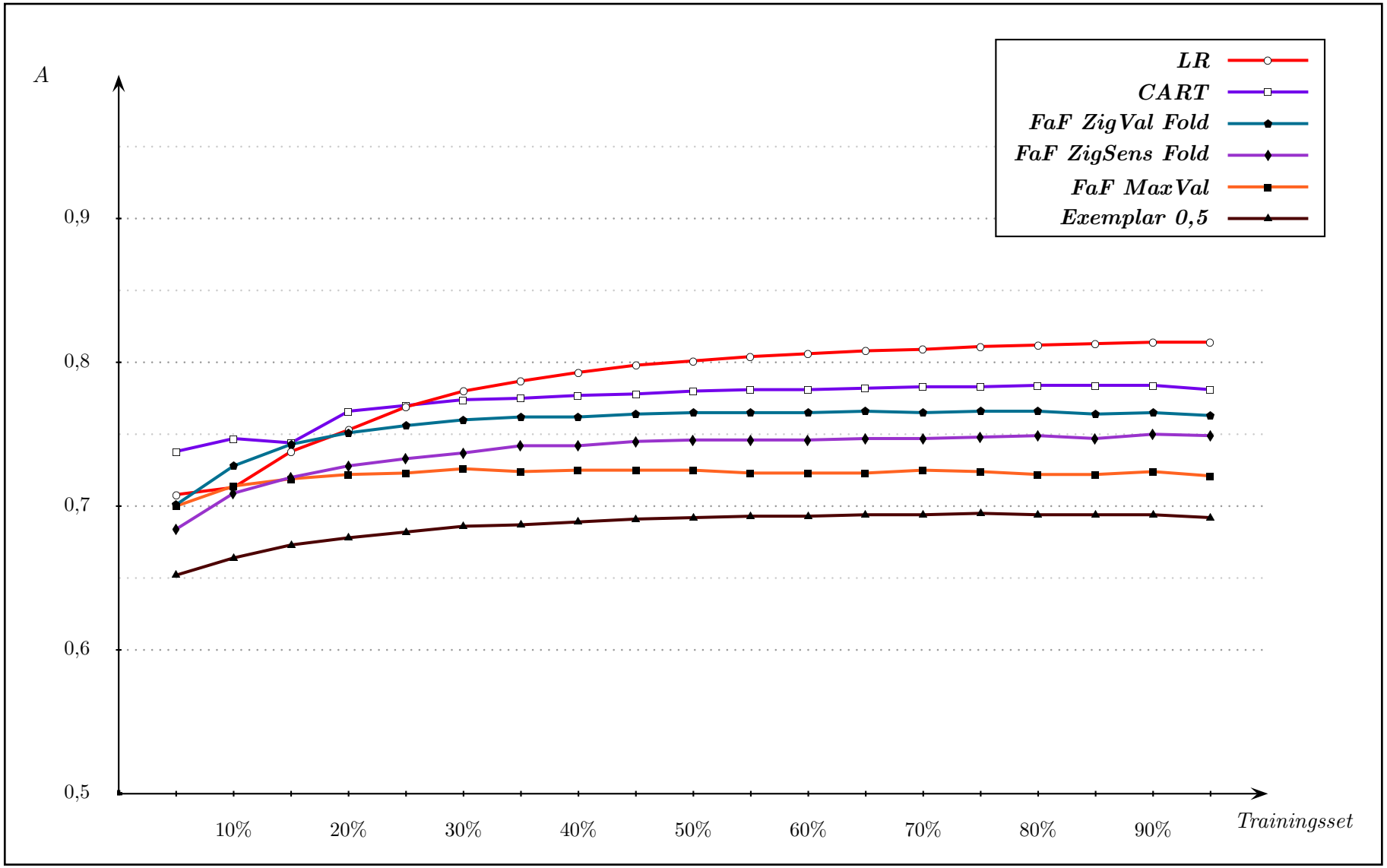


Abbildung 6.23: Vergleich der besten Prediction-Leistungen über 30 Datensätze und jeweils 1000 Trials gemittelt.

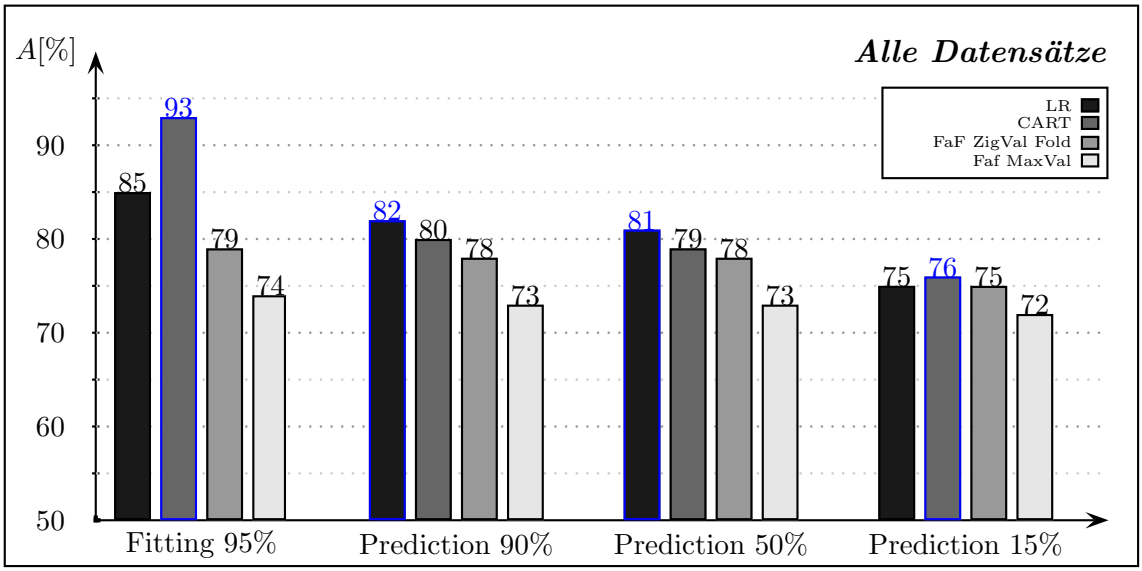


Abbildung 6.24: Balkendiagramm für den Algorithmenvergleich über alle 30 Datensätze: Die Leistung der vier Klassifikationsverfahren hinsichtlich Fitting und Prediction ist dabei auf ganze Prozentwerte gerundet worden. Die jeweils beste Leistung ist farbig umrandet dargestellt.

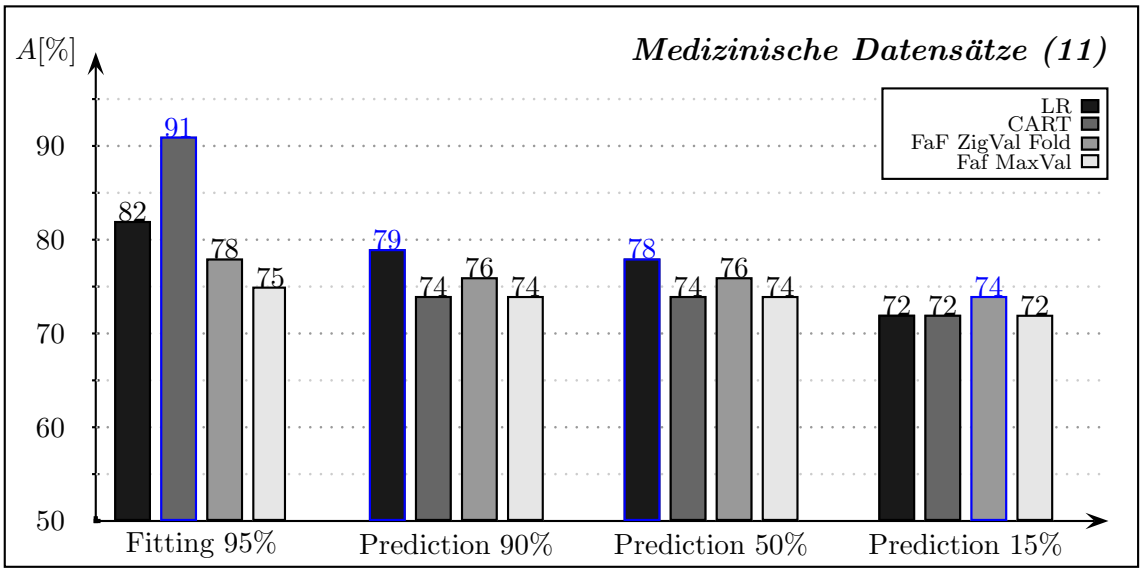


Abbildung 6.25: Balkendiagramm für den Algorithmenvergleich über 11 medizinische Datensätze: Die Leistung der vier Klassifikationsverfahren hinsichtlich Fitting und Prediction ist dabei auf ganze Prozentwerte gerundet worden. Die jeweils beste Leistung ist farbig umrandet dargestellt.

Tabelle 6.10: Ausgewählte Wettbewerbsergebnisse für die fünf *Fast and Frugal Tree*-Algorithmen mit den ursprünglichen Konstruktionsverfahren: In jeder Zeile finden sich die Angaben zu einem Datensatz, es sind die Werte angegeben für die Fitting-Leistung bei Trainingsdatensätzen mit 95% der Fälle sowie die Generalisierungsleistung bei Trainingsdatensätzen mit 15%, 50% und 90% der Fälle, erhoben an Testdatensätzen mit entsprechend 85%, 50% und 10% der Fälle eines Datensatzes.

			Fitting (95%)					Prediction (15%)					Prediction (50%)					Prediction (90%)				
Datensatz	n	Cues	MS	ZS	MV	ZV	MA	MS	ZS	MV	ZV	MA	MS	ZS	MV	ZV	MA	MS	ZS	MV	ZV	MA
Bankruptcy	50	5	61	83	61	83	61	59	70	54	71	54	62	82	56	82	57	58	81	52	81	56
Solvency	52	8	56	84	59	84	58	-	-	-	-	-	61	77	53	77	53	56	79	52	78	41
Echocardiogram	62	7	71	72	72	67	72	51	50	64	56	71	64	60	71	60	72	69	72	72	64	72
Post-Operative	86	8	72	52	72	61	72	66	48	68	50	71	72	47	72	50	72	72	41	72	51	72
Broadway	89	5	71	72	72	67	72	51	50	64	56	71	63	59	71	60	72	68	71	72	63	72
Credit Score	100	6	75	76	73	83	73	68	77	75	78	71	75	76	73	81	73	76	76	73	81	73
Zoo	101	16	41	68	100	89	60	-	-	-	-	-	44	71	98	89	62	40	66	99	87	60
Iris	150	4	43	66	69	53	69	45	54	66	53	62	41	58	67	48	66	42	65	68	48	68
Hepatitis	153	7	53	71	79	69	79	58	72	79	69	76	57	71	79	68	79	52	71	79	68	79
Strikes	163	4	71	72	77	75	71	69	72	67	74	65	71	72	71	75	67	71	71	75	74	68
Wine	178	13	44	77	60	77	60	50	74	61	75	60	46	77	60	77	61	43	77	60	77	60
Baseball CYY	189	9	30	55	78	65	78	35	59	77	61	77	31	59	78	63	78	29	54	77	64	77
Forensic Glass	214	9	78	78	73	81	73	70	79	72	78	68	78	79	73	81	70	78	78	73	80	72
Horse Colic	218	6	45	74	61	73	50	49	59	60	67	54	49	63	60	74	54	44	71	60	73	50
SPECT	267	22	36	74	79	66	79	31	71	79	61	75	33	72	79	64	79	34	73	79	64	79
Heart (Hungaria)	294	10	67	56	65	81	64	64	60	70	79	64	66	56	68	81	64	66	56	65	81	64
Heart (Cleveland)	297	13	55	62	75	79	54	52	58	62	75	55	53	60	66	77	54	55	62	70	75	54
Alcohol	345	5	37	56	79	55	79	47	56	78	56	78	43	57	79	56	79	37	55	79	55	79
Ionosphere	351	33	75	73	75	74	65	75	73	75	71	63	75	73	75	72	64	75	72	75	71	64
Tourism	365	4	28	41	75	45	75	38	41	74	41	74	33	41	74	41	74	29	38	74	42	74
Voting	402	16	63	96	64	95	87	64	94	70	92	80	61	96	64	95	85	62	97	64	96	87
Education	435	4	51	75	61	72	52	58	68	64	70	57	55	68	63	70	57	51	74	59	70	52
Cardiac	558	5	54	71	84	60	84	32	68	84	61	84	46	71	84	61	84	50	71	84	59	84
Football	672	4	63	63	67	64	67	58	62	66	62	64	62	63	67	64	67	63	63	67	64	67
Breast Cancer	683	9	83	85	87	95	74	82	88	85	94	74	84	86	84	94	74	83	85	85	95	74
Diabetes	786	6	54	64	68	64	68	55	64	68	64	68	52	63	68	64	68	54	63	68	64	68
Authorship	841	69	21	60	79	61	79	21	61	79	63	79	21	61	79	62	79	21	59	79	61	79
Baseball HoF	1320	15	16	57	89	53	89	16	56	90	53	90	16	57	90	53	90	16	56	89	52	89
Yeast	1484	8	69	57	69	53	69	69	55	69	55	69	69	55	69	55	69	69	56	69	52	69
Supreme Court	4052	7	96	54	96	52	96	96	54	96	59	96	96	54	96	55	96	96	54	96	52	96
Mittelwert			56	68	74	70	71	55	64	72	66	70	56	66	73	68	71	55	67	73	68	70

Tabelle 6.11: Ausgewählte Wettbewerbsergebnisse für die sechs modifizierten *Fast and Frugal Tree*-Algorithmen: In jeder Zeile finden sich die Angaben zu einem Datensatz, es sind die Werte angegeben für die Fitting-Leistung bei Trainingsdatensätzen mit 95% der Fälle sowie die Generalisierungsleistung bei Trainingsdatensätzen mit 15%, 50% und 90% der Fälle, erhoben an Testdatensätzen mit entsprechend 85%, 50% und 10% der Fälle eines Datensatzes.

			Fitting (95%)						Prediction (15%)						Prediction (50%)						Prediction (90%)					
Datensatz	n	Cues	SP	SM	SF	VP	VM	VF	SP	SM	SF	VP	VM	VF	SP	SM	SF	VP	VM	VF	SP	SM	SF	VP	VM	VF
Bankruptcy	50	5	82	85	83	85	85	83	70	65	65	73	73	70	83	83	83	84	83	83	82	80	80	82	80	80
Solvency	52	8	83	84	84	84	84	84	-	-	-	-	-	-	79	78	78	78	78	78	80	77	77	78	78	77
Echocardiogram	62	7	73	73	76	79	79	78	57	67	68	64	65	66	66	73	73	71	71	72	70	72	74	75	75	74
Post-Operative	86	8	72	72	72	69	69	72	67	70	70	60	62	64	72	72	72	64	64	68	73	73	73	65	66	71
Broadway	89	5	81	81	79	76	76	81	72	76	75	70	72	73	79	80	80	74	75	77	81	81	79	73	73	80
Credit Score	100	6	73	76	73	83	83	76	70	77	75	77	78	77	74	76	73	81	81	77	74	76	73	81	81	76
Zoo	101	16	68	100	100	96	96	96	-	-	-	-	-	-	70	98	97	95	95	94	67	100	100	95	95	94
Iris	150	4	66	77	75	72	72	72	55	60	61	65	65	65	57	67	66	71	71	71	64	75	70	70	70	68
Hepatitis	153	7	53	71	78	69	69	74	59	72	77	69	69	74	56	71	78	68	68	75	53	71	77	68	68	72
Strikes	163	4	72	72	72	75	75	75	71	70	71	72	72	72	72	71	71	73	74	74	72	71	71	73	74	74
Wine	178	13	77	92	92	92	92	92	73	77	77	84	84	84	77	89	90	89	89	90	77	92	92	92	92	92
Baseball CYY	189	9	55	69	70	66	66	68	59	68	71	69	69	72	59	68	70	67	67	68	55	66	67	65	65	67
Forensic Glass	214	9	78	78	80	81	81	84	73	78	77	77	78	79	78	79	78	80	80	81	78	78	79	80	81	82
Horse Colic	218	6	59	74	74	73	73	73	56	58	58	67	66	66	58	63	63	73	73	73	59	71	71	74	74	74
SPECT	267	22	51	74	77	66	66	70	47	71	77	61	61	70	48	72	78	64	64	71	49	73	77	64	64	69
Heart (Hungaria)	294	10	66	66	66	83	83	83	66	68	67	80	80	79	66	66	66	81	81	81	66	66	66	82	82	82
Heart (Cleveland)	297	13	55	55	55	80	80	80	53	52	52	75	74	74	54	54	53	76	77	77	54	54	54	76	77	77
Alcohol	345	5	56	71	79	75	75	78	60	73	77	71	71	76	58	74	78	72	72	78	56	72	79	74	74	78
Ionosphere	351	33	73	73	73	73	74	74	73	73	72	71	71	71	72	72	72	73	72	72	72	72	72	72	71	71
Tourism	365	4	41	61	69	64	64	71	44	60	68	61	61	68	42	59	67	62	62	68	38	57	66	61	61	68
Voting	402	16	93	97	97	95	95	95	86	94	94	92	92	92	91	96	96	95	95	95	93	96	96	95	95	95
Education	435	4	51	51	51	75	74	74	59	57	57	74	71	71	54	54	55	74	72	72	51	51	51	75	75	74
Cardiac	558	5	57	71	82	60	60	78	47	68	79	61	61	76	54	70	81	61	61	77	56	71	81	59	59	76
Football	672	4	63	63	64	64	64	65	57	62	63	62	62	63	61	63	64	64	64	64	62	63	64	64	64	65
Breast Cancer	683	9	85	80	80	96	96	96	86	83	82	95	95	95	86	81	81	96	96	96	86	80	80	96	96	96
Diabetes	786	6	64	72	72	71	71	71	64	70	70	71	71	71	63	71	71	71	71	71	64	72	72	72	72	72
Authorship	841	69	60	86	79	80	80	87	60	78	81	80	80	88	61	82	79	79	79	87	60	86	79	79	79	86
Baseball HoF	1320	15	57	74	84	65	65	72	56	73	83	65	65	75	57	74	84	64	65	72	57	74	84	65	65	71
Yeast	1484	8	69	69	69	63	63	68	69	69	69	64	65	67	69	69	69	64	64	68	69	69	69	63	63	67
Supreme Court	4052	7	96	96	96	73	73	96	96	96	96	82	82	96	96	96	96	79	79	96	96	96	96	77	77	96
Mittelwert			68	75	77	76	76	79	64	71	73	72	72	75	67	74	75	75	75	78	67	74	76	75	75	78

Tabelle 6.12: Ausgewählte Wettbewerbsergebnisse für die *Context Models* sowie LR und CART: In jeder Zeile finden sich die Angaben zu einem Datensatz, es sind die Werte angegeben für die Fitting-Leistung bei Trainingsdatensätzen mit 95% der Fälle sowie die Generalisierungsleistung bei Trainingsdatensätzen mit 15%, 50% und 90% der Fälle, erhoben an Testdatensätzen mit entsprechend 85%, 50% und 10% der Fälle eines Datensatzes.

			Fitting (95%)					Prediction (15%)					Prediction (50%)					Prediction (90%)				
Datensatz	n	Cues	LR	CT	Ex3	Ex4	Ex5	LR	CT	Ex3	Ex4	Ex5	LR	CT	Ex3	Ex4	Ex5	LR	CT	Ex3	Ex4	Ex5
Bankruptcy	50	5	96	97	69	69	68	69	48	59	59	58	83	80	66	65	65	87	80	66	66	66
Solvency	52	8	94	96	80	80	81	-	-	-	-	-	80	74	77	78	78	82	78	79	79	79
Echocardiogram	62	7	83	93	72	73	72	60	68	64	65	66	70	68	68	69	70	73	68	71	71	71
Post-Operative	86	8	74	80	72	72	72	54	61	65	67	68	63	61	71	72	72	70	65	72	72	72
Broadway	89	5	79	88	72	73	72	65	66	63	64	66	73	69	69	70	70	76	71	71	71	70
Credit Score	100	6	95	99	71	73	73	54	96	65	66	69	86	97	66	70	72	91	98	70	73	73
Zoo	101	16	100	100	55	56	55	-	-	-	-	-	98	99	55	55	55	99	100	55	55	55
Iris	150	4	74	98	59	65	67	70	83	58	60	62	73	94	59	63	66	71	95	58	63	66
Hepatitis	153	7	83	88	79	79	79	71	72	77	78	79	78	74	79	79	79	81	75	79	79	79
Strikes	163	4	73	89	73	73	72	68	69	68	67	65	70	72	71	71	71	70	70	72	72	72
Wine	178	13	100	98	60	62	62	89	87	59	59	60	95	91	59	61	61	96	90	61	62	61
Baseball CYY	189	9	86	95	75	77	77	75	78	72	74	75	83	82	74	76	77	83	83	75	77	77
Forensic Glass	214	9	86	95	58	60	63	77	77	58	60	62	83	79	59	61	63	84	82	59	60	63
Horse Colic	218	6	69	90	58	58	59	62	62	55	55	56	66	66	57	58	58	66	66	58	58	58
SPECT	267	22	88	89	75	77	80	67	76	71	72	75	79	80	74	76	78	82	81	74	76	79
Heart (Hungaria)	294	10	85	91	57	59	60	76	74	58	59	61	82	76	57	58	60	83	76	56	57	60
Heart (Cleveland)	297	13	85	91	60	61	61	74	71	60	60	60	82	74	59	60	60	83	73	60	59	60
Alcohol	345	5	78	91	75	74	74	75	70	74	75	75	77	71	75	74	74	77	71	74	75	75
Ionosphere	351	33	94	98	59	60	61	76	84	58	59	60	85	87	58	60	60	88	88	58	60	61
Tourism	365	4	77	88	77	77	77	75	69	77	77	77	77	70	77	77	77	77	70	77	77	77
Voting	402	16	98	98	49	49	49	93	94	48	48	50	95	95	48	48	49	96	95	49	49	49
Education	435	4	77	86	56	56	56	75	71	54	55	53	76	71	55	55	55	77	71	56	56	55
Cardiac	558	5	84	93	84	84	84	81	74	84	84	84	84	74	84	84	84	84	74	84	84	84
Football	672	4	68	79	67	67	67	66	59	67	67	67	67	59	67	67	67	68	57	67	67	67
Breast Cancer	683	9	97	98	50	52	52	94	93	57	59	61	96	94	53	53	58	97	95	50	51	54
Diabetes	786	6	77	92	69	69	68	76	69	68	68	67	77	71	69	68	68	77	72	69	69	68
Authorship	841	69	100	99	73	73	75	96	93	72	72	73	98	95	72	73	74	98	95	72	73	74
Baseball HoF	1320	15	95	98	91	91	91	92	92	90	90	91	94	93	91	91	91	95	93	91	91	91
Yeast	1484	8	68	91	69	69	69	68	67	68	69	69	68	69	69	69	69	68	70	69	69	69
Supreme Court	4052	7	96	97	96	96	96	96	96	96	96	96	96	96	96	96	96	96	96	96	96	96
Mittelwert			85	93	69	69	70	75	76	67	67	68	81	79	68	69	69	82	80	68	69	69

Untergruppen identifiziert werden könnten. Zuletzt soll an zwei einzelnen (nicht-medizinischen) Datensätzen demonstriert werden, dass sich auch außerhalb des Bereichs Anwendungsmöglichkeiten finden lassen können.

6.2.4.3 Einzelne Datensätze

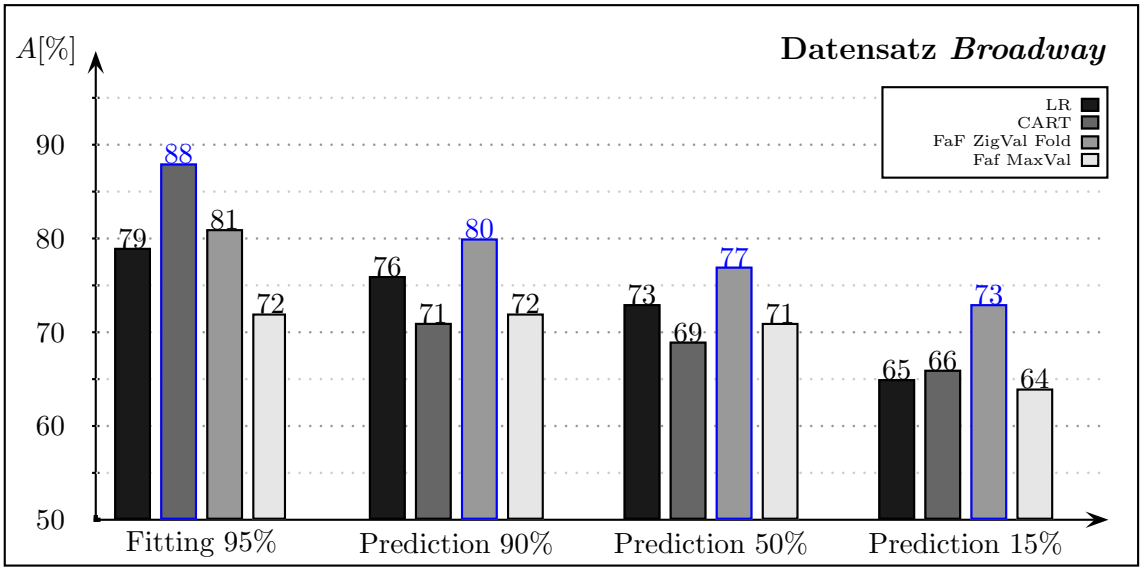


Abbildung 6.26: Balkendiagramm für den Algorithmenvergleich für den Datensatz *Broadway*: Die Leistung der vier Klassifikationsverfahren hinsichtlich Fitting und Prediction ist dabei auf ganze Prozentwerte gerundet worden. Die jeweils beste Leistung ist farbig umrandet dargestellt.

Zunächst sollen die Ergebnisse für den Datensatz *Broadway* - dargestellt in Abbildung 6.26 diskutiert werden: An diesem Datensatz lässt sich die Dominanz des FaF ZigVal Fold-Verfahrens über alle drei Stichprobengrößen hinweg demonstrieren. Dabei ist der Algorithmus auch im Fitting besser als LR, was darauf hindeutet, dass das Verfahren in der Konkurrenz nicht nur vom Overfitting der übrigen Verfahren profitiert. Damit nicht der Eindruck entstehen kann, dass zumindest die Baumverfahren sich in eine datensatzinvariante Ordnung sortieren lassen, sollen aber auch die Ergebnisse für den Datensatz *Yeast* gesondert dargestellt werden (s. Abbildung 6.27), bei denen sich Faf MaxVal bei kleinen Datensätzen an die Spitze des Feldes setzt, in jedem Fall aber bessere Ergebnisse erbringt als Faf ZigVal Fold. Besonders drastisch ist bei diesem Datensatz das *Overfitting* durch CART, während LR und FaF MaxVal in den Ergebnissen zwischen *Fitting* und *Prediction* konstant bleiben. Was an diesem Datensatz zudem demonstriert werden kann, ist, dass die Leistung der Bäume nicht allein von der absoluten Größe der Stichprobe abhängt; der Datensatz *Yeast* gehört mit 1484 Fällen zu den größten Datensätzen der Konkurrenz.

6.2.4.4 Ergebnisse für die drei Ausgangsdatensätze

Fast and Frugal Trees stellen ebenfalls hierarchische Regelmengen dar, weshalb diese Methode auch zur Bearbeitung der Ausgangsfragestellung eingesetzt werden kann. Um zu vergleichbaren Ergebnissen zu gelangen, wurden die Verfahren CART, LR, FaF MaxVal und FaF ZigVal Fold für die Bedingungen Fitting (95%) und Prediction (15%, 50%, 90%) mit jeweils 1000 Trials an den Datensätzen *Hochschule*, *Beratungsunternehmen* und *Schüler* getestet. Die auf glatte Prozentwerte gerundeten Ergebnisse sind in Tabelle 6.13 augeführt.

Die durschnittlichen Ergebnisse über die drei Datensätze gemittelt sind zudem in Abbildung 6.28 visualisiert[7]. Die Ergebnisse ähneln denen für den Datensatz *Yeast*. CART zeichnet sich durch ein immenses *Overfitting* aus, gerade bei kleineren Stichproben sind die Bäume sehr konkurrenzfähig. Eine Beobachtung der von FaF MaxVal konstruierten Bäume muss den Erfolg ein wenig schmälern: Die Ergebnisse des Verfahrens beruhen zu einem Großteil auf der schlichten Vorgehensweise, alle vorhandenen Fälle in den

[7] Im Vergleich zu den Ergebnissen von Montel (2005) sei darauf hingewiesen, dass die hier verwendete CART-Variante keine Pruning-Prozedur enthielt, dementsprechend aber auch ausschließlich die Trainingsstichprobe zur Erstellung des Modells genutzt werden konnte; dementsprechend ist das schlechtere Ergebnis von CART so zu erklären.

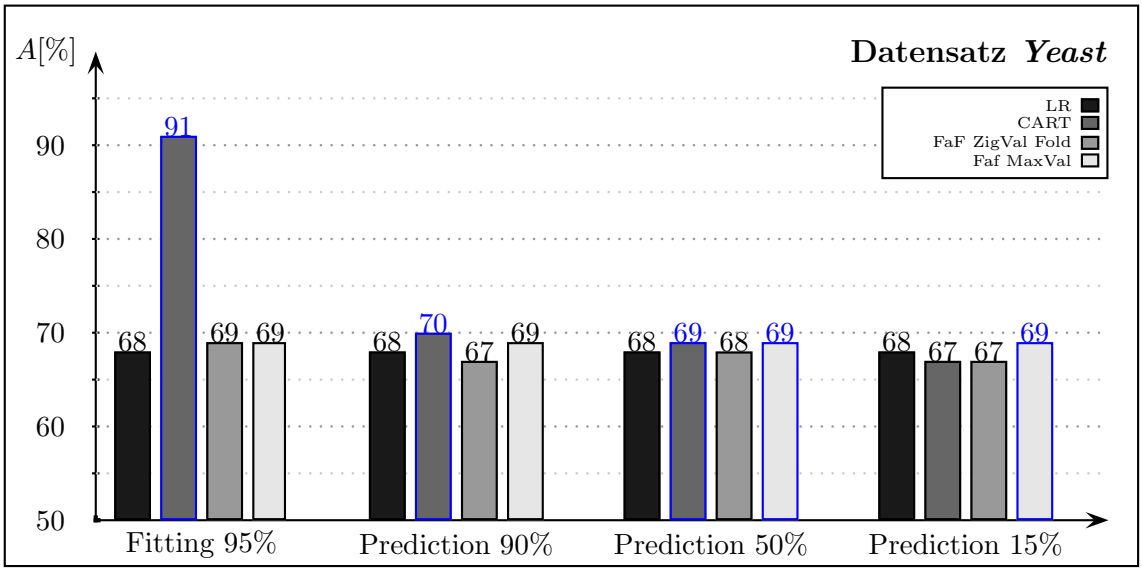


Abbildung 6.27: Balkendiagramm für den Algorithmenvergleich für den Datensatz *Yeast*: Die Leistung der vier Klassifikationsverfahren hinsichtlich Fitting und Prediction ist dabei auf ganze Prozentwerte gerundet worden. Die jeweils beste Leistung ist farbig umrandet dargestellt.

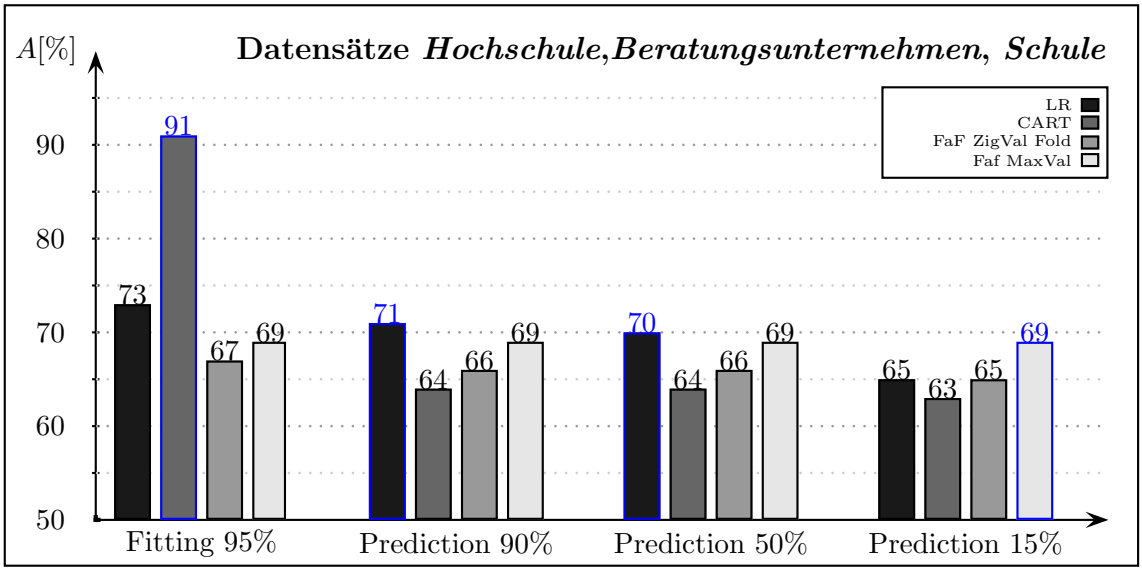


Abbildung 6.28: Balkendiagramm für den Algorithmenvergleich für die Datensätze *Hochschule*, *Beratungsunternehmen* und *Schule*): Die über Datensätze und Trials gemittelte Leistung der vier Klassifikationsverfahren hinsichtlich Fitting und Prediction ist dabei auf ganze Prozentwerte gerundet worden. Die jeweils beste Leistung ist farbig umrandet dargestellt.

meisten Trials in die Majoritätsklasse einzusortieren. Im Sinne der statistischen Problemstellung ist dieses Vorgehen erfolgreich, im Sinne einer praktischen Problemlösung aber selten nützlich. Dies trifft durch die alternierende Struktur nicht auf die Bäume von FaF ZigVal Fold zu. Auch CART und LR können nicht bessere Ergebnisse erzielen als das in diesem Fall triviale Prinzip, was bereits bei der Analyse der Ergebnisse von Montel (2005) bemerkt worden war und auch bei den Erprobungen anderer Algorithmen zu Tage trat. An dieser Stelle wird die Hypothese plausibel, dass die in den Daten befindliche Struktur für die Klassifikation entweder sich allen verwendeten Verfahren zu einem substanziellen Teil entzieht oder aber nicht in einem ausreichenden Maße vorhanden ist. Die Stichproben in der 15%-Bedingung haben die Größen $n \approx 30$, $n \approx 65$ und $n \approx 300$, stellen also für manche Fragestellungen durchaus plausible Entscheidungsgrundlagen dar.

Damit zeigt sich in jedem Fall, dass kein Verfahren die Konkurrenz dominieren kann, dass also nur in Anbetracht der Aufgabenstellung bestimmt werden kann, welches Verfahren präferiert werden sollte. *Fast and Frugal Trees* können hier ihre ökologische Rationalität in bestimmten Anwendungsbereichen positiv demonstrieren und bieten sich somit zur Aufnahme auch in den Methodenkanon des psychologischen Diagnostikers an.

6.3 Untersuchung der Klassifikationsleistung von gier.ga und gier.swarm

In diesem Abschnitt soll über Ergebnisse zu Algorithmen aus dem Bereich des Künstlichen Lebens berichtet werden. Dazu wurden zwei Algorithmenklassen an den bereits für die Downhill-Varianten eingesetzten Datensätzen erprobt.

6.3.1 Präzisierung der Aufgabenstellung

Das zu lösende Problem ist erneut ein Klassifikationsproblem. Die eingesetzten Algorithmen erhalten erneut eine Teilmenge der Daten und müssen hierarchische Regelmengen erstellen, deren Genauigkeit an Testdatenmengen erprobt wird. Ziel der Simulation ist es, einen Eindruck über die relative Eignung der Verfahren und ihrer verwendeten Varianten zu erhalten.

6.3.1.1 Verwendete Algorithmen

In der Simulationsstudie werden Varianten des $\text{GIER}_{[1.0]}$.ga-Algorithmus und des $\text{GIER}_{[1.0]}$.swarm-Algorithmus erprobt. Die verwendeten $\text{GIER}_{[1.0]}$.ga-Algorithmen können in der VBk-Darstellung durch Abbildung 6.29 repräsentiert werden. Der entsprechende $\text{GIER}_{[1.0]}$.script-Code (vgl. Anhang C) ist in Listing 6.3 wiedergegeben:

```
gier_ga_M/
(gier.ga_script)
Mind.-n(Regel)=M/ Restf.=True/ Pop.gr.=20/ Neu-Pop.gr.=20/ Generationen=0/
             Datens.=False/ Regels.=False/ Aktivg=False/ abd.v.reg.=False/
             Max.Var.=0/ Elit.=True/
    t_creation_rule**  U.repair=False/ R.repair=True/ Aktivg=False/ Max.var=0/
             Snowbd=False/
        t_creation_decision_all_x**  x=2/
        no_snowboard**
    select_all**
    t_objective_tq**
    t_fitness_positivevalues**
    t_scalefitness_truncation**  f=2/ Sigma=1,5/
        t_partition_noactive**
    Operatoren**  Anzahl=2/ Mutationen=1/ Cross-Overs=1/ Mut. Aktivgen=0/
             Cr.-Ov.    Aktivgen=0/ mu/ 0/ co/ 0/
    t_mutation_standard**  p=0,01/
    t_cross_over_rule**  cross_two=True/ species_rule=False/
          t_cross_detail_uniform_50**  probability=0,3/
    t_reparatur_randomrepair**  One-Scale=True/
```

Tabelle 6.13: Ergebnisse von vier Verfahren für die drei Ausgangsdatensätze: In jeder Zeile finden sich die Angaben zu einem Datensatz, es sind die Werte angegeben für die Fitting-Leistung bei Trainingsdatensätzen mit 95% der Fälle sowie die Generalisierungsleistung bei Trainingsdatensätzen mit 15%, 50% und 90% der Fälle, erhoben an Testdatensätzen mit entsprechend 85%, 50% und 10% der Fälle eines Datensatzes.

			Fitting (95%)				Prediction (90%)				Prediction (50%)				Prediction (15%)			
Datensatz	n	Cues	LR	CT	ZV	MV	LR	CT	ZV	MV	LR	CT	ZV	MV	LR	CT	ZV	MV
Hochschule	191	14	75	91	68	72	71	67	65	72	69	66	66	72	60	64	64	72
Beratungsunternehmen	437	16	76	92	71	73	74	65	71	73	73	66	70	73	68	66	68	73
Schule	1993	15	67	90	63	61	66	59	63	61	67	59	63	61	66	58	62	61
Mittelwert			73	91	67	69	71	64	66	69	70	64	66	69	65	63	65	69

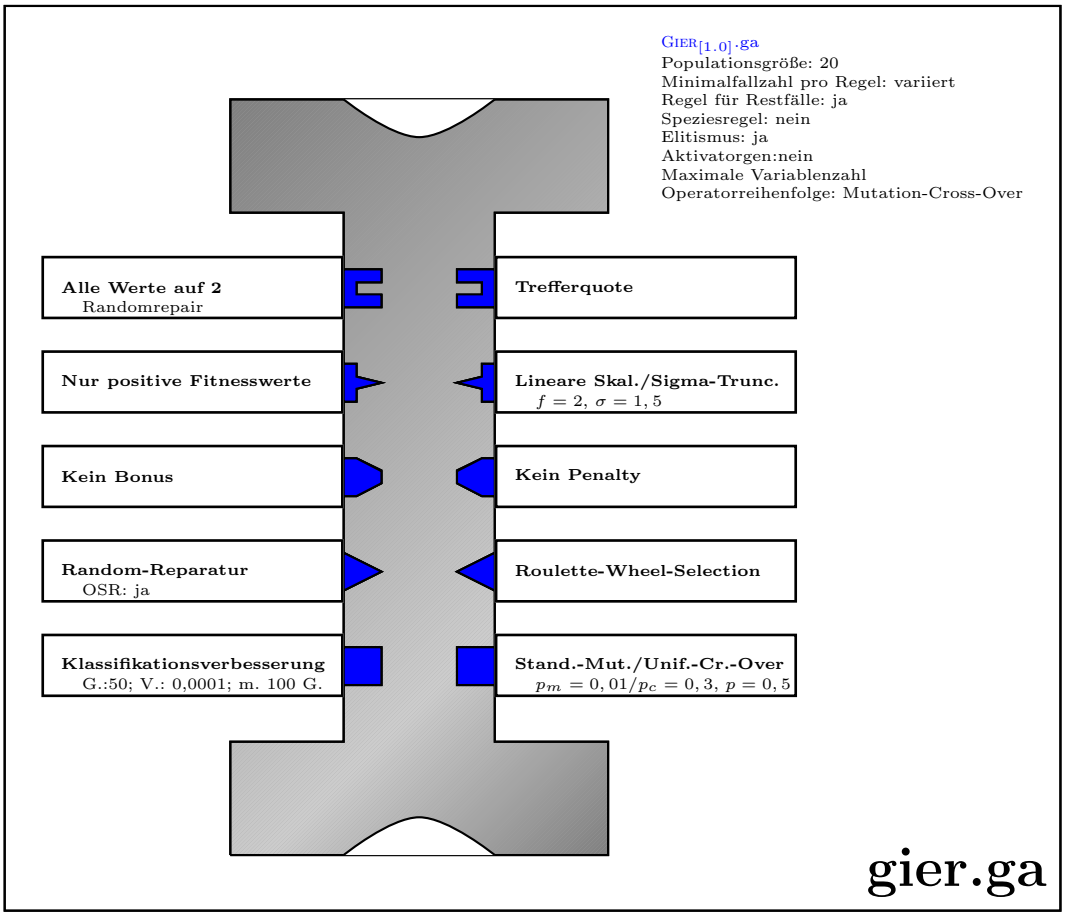


Abbildung 6.29: VBk-Darstellung der eingesetzten Varianten GIER[1.0].ga

```
t_bonus_fitness_no_bonus**
t_penalty_fitness_no_penalty**
t_selection_wheel**
    t_partition_noactive**
t_stopcheck_change_class**  mind.Generationen=100/
      hoechst.Generationen=10000000/ Generationen=50/
      Klassifikationsverbesserung=0,0001/
```

Listing 6.3: Script-Code für die genetischen Algorithmen: An die Stelle von M wird jeweils die Mindestfallzahl eingesetzt

Die Varianten des Algorithmus unterscheiden sich dabei ausschließlich hinsichtlich der Mindestfallzahl pro Regel, die Werte zwischen 20 und 80 annehmen kann. Die 20 Individuen der Population werden alle mit variabler Klassifikationsrichtung initialisiert, also ohne Speziesregel. Anfangsindividuen werden dabei mit dem *Random-Repair*-Verfahren korrigiert. Der Algorithmus verwendet einen Mutations- und einen *Cross-Over*-Operator: Der Mutationsoperator ist die Standardmutation mit $p_m = 0,01$, der *Cross-Over*-Operator ein *Uniform-Repair* mit $p_c = 0,3$. Die Fitnesswerte werden skaliert, das Selektionsverfahren ist die *Roulette-Wheel-Selection*. Es werden weder Bonus- noch Penalty-Regeln verwendet, dafür wird eine *Random-Repair*-Prozedur mit *One-Scale-Repair* eingesetzt. Das Verfahren wird solange iteriert, bis in fünfzig aufeinanderfolgenden Generationen keine Klassifikationsverbesserung von mindestens 0,0001 mehr erreicht werden kann, mindestens jedoch 100 Generationen. Eine Variablenauswahl wird nicht vorgenommen. Die Varianten des Schwarmalgorithmus werden in der VBk-Darstellung in Abbildung 6.30 dargestellt und in Listing 6.4 wiedergegeben:

```
gier_swarm_M/
(gier.swarm_script)
    Mind.-n(Regel)=M/ Restf.=True/ Schwarmgr.=25/ Timest.=0/ Datens.=False/
```

```
        Regels.=False/ Geschw.=5/
t_swarm_init_connections_circle**
t_creation_rule**  U.repair=False/ R.repair=True/ Aktivg=False/ Max.var=0/
        Snowbd=False/
    t_creation_decision_all_x**  x=2/
    no_snowboard**
t_swarm_move_random**  vmax=5/ phi=4/ rand=True/
    t_swarm_inertia_none**
t_objective_tq**
t_fitness_notransformation**
t_bonus_fitness_no_bonus**
t_penalty_fitness_size_linear**  Cut-Off=[M]/ b=0,05/ c=0,2/
t_stopcheck_change_class**  mind.Generationen=0/ hoechst.Generationen=10000000/
        Generationen=50/ Klassifikationsverbesserung=0,0001/
```

Listing 6.4: Script-Code für die Schwarmalgorithmen: An die Stelle von M wird jeweils die Mindestfallzahl eingesetzt

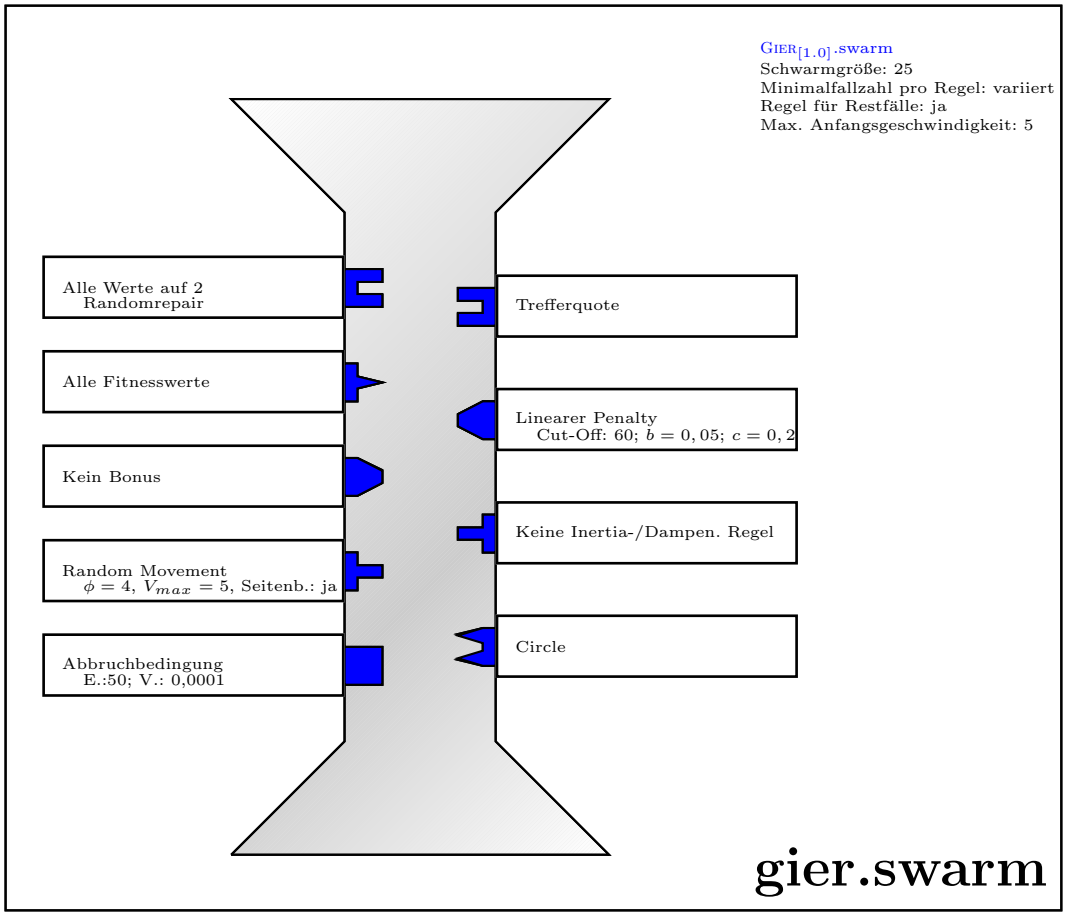


Abbildung 6.30: VBk-Darstellung der eingesetzten Varianten GIER[1.0].swarm

Auch hier unterscheiden sich die Varianten lediglich in der Mindestfallzahl - die hier zwischen 20 und 90 liegt - mit einer zusätzlichen Anpassung im Penalty-Modul. Die Initialisierung des Schwarms entspricht der Initialisierung der GA-Anfangspopulationen, allerdings zusätzlich mit einer Anfangshöchstgeschwindigkeit von 5 und einer Circle-Topologie. Das Bewegungsmodul setzt ein *Random-Movement* mit $\phi = 4$, $V_{max} = 5$ und Seitenbegrenzung ein, eine *Inertia*- oder *Dampening*-Regel wird nicht verwandt. Zur Bewertung der Lösungen wird eine Penaltyregel mit einem Cut-Off entsprechend der Mindestfallzahl und $b = 0,05$ und $c = 0,2$ eingesetzt. Die Abbruchbedingung wird ebenfalls auf 50 Epochen ohne Verbesserung von mindestens 0,0001 angesetzt.

Die Algorithmen werden jeweils in mehreren Varianten vorgegeben, die sich hinsichtlich der Mindestfallzahl pro Regel unterschieden. Bei den Swarm-Algorithmen wird auch die Penalty-Bedingung entsprechend angepasst. Als Vergleichsverfahren werden erneut LR und CART berücksichtigt, weiterhin die beiden Verfahren Faf MaxVal und FaF ZigVal Fold. Zum Teil waren die Analysen bereits bei der Untersuchung der *Fast and Frugal* Trees angestellt worden, für den 80%:20%-Datensatz-Split wurden jeweils 1000 Trials zusätzlich durchgeführt.

6.3.1.2 Datensätze und Vorgabemodus

Tabelle 6.14: Trialübersicht im Algorithmenvergleich von GIER$_{[1.0]}$.ga und GIER$_{[1.0]}$.swarm

		Datensatz		
		Hochschule		*Beratungsunternehmen*
Algorithmus	Mindestfallzahl	50%:50%	80%:20%	50%:50%
GIER$_{[1.0]}$.GA	20	50 Trials	-	-
	30	50 Trials	-	-
	40	50 Trials	-	50 Trials
	50	50 Trials	-	50 Trials
	60	-	-	50 Trials
	70	-	-	50 Trials
	80	-	-	50 Trials
GIER$_{[1.0]}$.swarm	20	500 Trials	-	-
	30	500 Trials	500 Trials	-
	40	500 Trials	500 Trials	500 Trials
	50	500 Trials	500 Trials	500 Trials
	60	-	500 Trials	500 Trials
	70	-	-	500 Trials
	80	-	-	500 Trials
	90	-	500 Trials	-
LR	-	1000 Trials	1000 Trials	1000 Trials
CART	-	1000 Trials	1000 Trials	1000 Trials
FaF MaxVal	-	1000 Trials	1000 Trials	1000 Trials
FaF ZigVal Fold	-	1000 Trials	1000 Trials	1000 Trials

Als Prüfsteine für den Methodenvergleich dienen erneut die beiden Datensätze *Hochschule* und *Beratungsunternehmen*. Der Datensatz *Hochschule* wird mit einem Trainingsset-Testset-Split von 50%:50% verwendet, der Datensatz Beratungsunternehmen einmal mit einem Split von 50%:50% und einmal mit einem Split von 80%:20%. Aufgrund der Laufzeitdifferenzen wird die letztgenannte Aufteilung nur den GIER$_{[1.0]}$.swarm-Algorithmen vorgegeben. Für GIER$_{[1.0]}$.ga-Algorithmen werden pro Datensatz und Algorithmus 50 Trials angesetzt, für die schnelleren GIER$_{[1.0]}$.swarm-Algorithmen 500 Trials[8]. Erhoben wurde für die Schwarm-Algorithmen und die genetischen Algorithmen in jedem Trial die Anzahl der Regeln in der erzeugten Regelmenge sowie Klassifikationsgenauigkeit, Sensitivität und Spezifität für *Fitting* und *Prediction*. Insgesamt wurden die in Tabelle 6.14 zusammengefassten Trials durchgeführt. Damit stellen insgesamt 2×19450 Ergebnisse die Basis für die folgenden Auswertungen dar.

6.3.2 Ergebnisse für den Datensatz Hochschule

An dieser Stelle sollen die Ergebnisse zunächst getrennt nach Algorithmenklassen dargestellt werden, wobei zunächst die GIER$_{[1.0]}$.ga-Ergebnisse und im Anschluss die Ergebnisse der Schwärme präsentiert werden.

[8] Ein Durchgang mit 500 Trials benötigt auf dem verwendeten PC bei vier Schwarmvarianten etwa 300 Minuten Laufzeit, ein Durchgang mit 50 Trials bei vier GA-Varianten benötigt hingegen ca. 400 Minuten. Für diesen Teil der Untersuchung standen leider keine weiteren Computerressourcen zur Verfügung, deswegen beschränken sich die Ergebnisse auf den beschriebenen Umfang.

6.3.2.1 Ergebnisse für gier.ga

Tabelle 6.15: Ergebnisse für die $\text{GIER}_{[1.0]}$.ga-Varianten für den Datensatz *Hochschule* mit einem Trainings-/Testset-Split von 50%-50%: Klassifikationsgenauigkeit, Sensitivität und Spezifität sind als gerundete Durchschnittswerte in Prozent angegeben, die Regelzahl entspricht dem Mittelwert über 50 Trials.

Datensatz *Hochschule* 50%			Fitting			Prediction		
Algorithmus	Mindestfälle	Regelzahl	A	S^+	S^-	A	S^+	S^-
$\text{GIER}_{[1.0]}$.ga	20	4,98	90	94	82	65	75	37
$\text{GIER}_{[1.0]}$.ga	30	3,82	85	88	79	65	73	43
$\text{GIER}_{[1.0]}$.ga	40	3,00	85	92	56	68	81	34
$\text{GIER}_{[1.0]}$.ga	50	2,00	82	91	57	68	81	34

Die gerundeten durchschnittlichen Ergebnisse der $\text{GIER}_{[1.0]}$.ga Varianten für den Datensatz *Hochschule* (50%) sind in Tabelle 6.15 aufgeführt, die Ergebnisse für den Datensatz *Beratungsunternehmen* (50%) in Tabelle 6.16.

Tabelle 6.16: Ergebnisse für die $\text{GIER}_{[1.0]}$.ga-Varianten für den Datensatz *Beratungsunternehmen* mit einem Trainings-/Testset-Split von 50%-50%: Klassifikationsgenauigkeit, Sensitivität und Spezifität sind als gerundete Durchschnittswerte in Prozent angegeben, die Regelzahl entspricht dem Mittelwert über 50 Trials.

Datensatz *Beratungsunternehmen* 50%			Fitting			Prediction		
Algorithmus	Mindestfälle	Regelzahl	A	S^+	S^-	A	S^+	S^-
$\text{GIER}_{[1.0]}$.ga	40	6,00	85	92	66	69	82	37
$\text{GIER}_{[1.0]}$.ga	50	5,00	84	91	66	71	82	41
$\text{GIER}_{[1.0]}$.ga	60	4,00	81	89	61	70	81	40
$\text{GIER}_{[1.0]}$.ga	70	3,88	81	89	60	70	81	39
$\text{GIER}_{[1.0]}$.ga	80	3,00	79	87	58	71	81	43

Erwartungsgemäß sind die *Fitting*-Ergebnisse bei kleineren Mindestfallzahlen besser: Ein Regelsystem mit mehr Regeln, die evolviert werden, und kleinerer Fallabdeckung pro Regel kann sich besser an die Trainingsdaten anpassen. Die durschnittliche Sensitivität ist dabei höher als die durchschnittliche Spezifität, was durch die Kombination aus einer Verwendung von Trefferquoten als Fitnesskriterium und der Ungleichverteilung der Klassen im Datensatz erklärt werden kann. In der Generalisierung zeigt sich dabei ein deutliches *Overfitting* vor allem für den (kleineren) Datensatz *Hochschule*: Die Leistungsdifferenzen zwischen den Verfahren sind insgesamt niedrig, tendenziell zeigen Verfahren mit weniger Regeln bessere Ergebnisse. Insbesondere geht die Spezifität in der Generalisierung auf Werte unterhalb von 0,5 zurück, d.h. durch die Klassifikation werden negative Fälle schlechter identifiziert als durch eine Zufallsauswahl.

6.3.2.2 Ergebnisse für gier.swarm

Tabelle 6.17: Ergebnisse für die $\text{GIER}_{[1.0]}$.swarm-Varianten für den Datensatz *Hochschule* mit einem Trainings-/Testset-Split von 50%-50%: Klassifikationsgenauigkeit, Sensitivität und Spezifität sind als gerundete Durchschnittswerte in Prozent angegeben, die Regelzahl entspricht dem auf zwei Nachkommastellen gerundeten Mittelwert über 500 Trials.

Datensatz *Hochschule* 50%			Fitting			Prediction		
Algorithmus	Mindestfälle	Regelzahl	A	S^+	S^-	A	S^+	S^-
$\text{GIER}_{[1.0]}$.swarm	20	4,98	89	93	79	65	78	33
$\text{GIER}_{[1.0]}$.swarm	30	3,73	83	87	74	65	74	39
$\text{GIER}_{[1.0]}$.swarm	40	3,00	82	94	49	69	85	27
$\text{GIER}_{[1.0]}$.swarm	50	2,00	74	87	38	64	81	24

Tabelle 6.18: Ergebnisse für die GIER$_{[1.0]}$.swarm-Varianten für den Datensatz *Hochschule* mit einem Trainings-/Testset-Split von 80%-20%: Klassifikationsgenauigkeit, Sensitivität und Spezifität sind als gerundete Durchschnittswerte in Prozent angegeben, die Regelzahl entspricht dem auf zwei Nachkommastellen gerundeten Mittelwert über 500 Trials.

Datensatz *Hochschule* 80%			Fitting			Prediction		
Algorithmus	Mindestfälle	Regelzahl	A	S^+	S^-	A	S^+	S^-
GIER$_{[1.0]}$.swarm	20	5,11	85	92	67	66	80	32
GIER$_{[1.0]}$.swarm	30	4,00	83	90	66	67	80	35
GIER$_{[1.0]}$.swarm	40	3,36	79	84	68	66	75	41
GIER$_{[1.0]}$.swarm	50	3,00	81	93	50	71	86	31

Die Ergebnisse der Schwarmalgorithmen sind in Tabelle 6.17 und Tabelle 6.18 für den Datensatz *Hochschule* und in Tabelle 6.19 für den Datensatz *Beratungsunternehmen* zusammengefasst. Das Fitting betreffend zeigt sich der gleiche Trend wie bei den genetischen Algorithmen, die Varianten mit kleinerer Mindestfallzahl zeigen bessere Ergebnisse. Die Verhältnisse in der Generalisierung sind weniger geordnet: zwar rücken auch hier die Verfahren nahe zusammen, die Reihenfolge der Algorithmen kehrt sich aber nicht vollständig um. Dies ist in dem Moment nicht zu erwarten, in denen eine Struktur in den Daten vorhanden ist, auf deren Grundlage eine Generalisierung möglich wird, die auf bestimmte Regelgrößen zugeschnitten ist. Weiterhin stellen große Regeln unter Umständen zu unflexible Regeln dar, um sich adäquat noch an Datenteilmengen anpassen zu können. Insbesondere bei den größeren Fallzahlen besteht das Ergebnis im Extremfall aus einer einzigen Regel und einer Restregel. In diesen Fällen ist es eher beunruhigend festzustellen, dass diese Lösung von Lösungen mit mehr Regeln in der Generalisierung nicht übertroffen werden kann. Der Vergleich zwischen den beiden Aufteilungen des Datensatzes *Hochschule* zeigt zum einen eine größere benötigte Regelzahl in der 80%-Bedingung. Weiterhin sind die Fitting-Ergebnisse im größeren Trainingsdatensatz schlechter, dafür sind die Generalisierungsleistungen tendenziell besser.

Die Generalisierungsergebnisse für den Datensatz *Beratungsunternehmen* sind auf ähnlicher Stufe über alle Bedingungen hinweg. Hier ist allerdings die Trainingsstichprobe deutlich größer als im vorherigen Datensatz. Die Variante für die Bedingung mit 70 Fällen fällt etwas aus der Ordnung heraus: Die Sensitivität in der Generalisierungsbedingung ist relativ niedrig, die Spezifität höher.

Tabelle 6.19: Ergebnisse für die GIER$_{[1.0]}$.swarm-Varianten für den Datensatz *Beratungsunternehmen* mit einem Trainings-/Testset-Split von 50%-50%: Klassifikationsgenauigkeit, Sensitivität und Spezifität sind als gerundete Durchschnittswerte in Prozent angegeben, die Regelzahl entspricht dem auf zwei Nachkommastellen gerundeten Mittelwert über 500 Trials.

Datensatz *Hochschule* 80%			Fitting			Prediction		
Algorithmus	Mindestfälle	Regelzahl	A	S^+	S^-	A	S^+	S^-
GIER$_{[1.0]}$.swarm	30	5,11	85	92	67	66	80	32
GIER$_{[1.0]}$.swarm	40	4,00	83	90	66	67	80	35
GIER$_{[1.0]}$.swarm	50	3,36	79	84	68	66	75	41
GIER$_{[1.0]}$.swarm	60	3,00	81	93	50	71	86	31
GIER$_{[1.0]}$.swarm	90	2,00	73	88	33	66	85	22

6.3.3 Vergleich der Algorithmen

Im Folgenden sollen die Ergebnisse der Algorithmenklassen zunächst miteinander verglichen werden, um im Anschluss auch mit der erweiterten Konkurrenz konfrontiert zu werden.

6.3.3.1 Vergleich von gier.ga und gier.swarm

Um Ähnlichkeiten und Unterschiede zu verdeutlichen, sind die Ergebnisse der Algorithmen für die beiden vergleichbaren Aufgaben in Liniendiagrammen gegeneinander aufgetragen worden. Abbildung 6.31

lässt den Vergleich der durchschnittlichen Regelzahlen der resultierenden Regelmengen zu: Die Abweichungen zwischen den Algorithmenklassen sind unbeträchtlich. Beide Verfahren bewegen sich zudem eng an den Mindestgrenzen, so dass die Regeln ähnlich große Teilmengen der Fälle abdecken können. In Abbildung 6.32 sind die *Fitting*-Leistungen der Algorithmen gegeneinander aufgetragen: Hier zeigen die genetischen Algorithmen für den Datensatz *Hochschule* leicht bessere Ergebnisse, für den anderen Datensatz sind die Resultate vergleichbar. Abbildung 6.33 ermöglicht weiterhin den Vergleich der Vorhersageleistungen der Algorithmenklassen: Hier büßen die genetischen Algorithmen ihren Vorsprung im Fitting für den Datensatz *Hochschule* ein, für den Datensatz *Beratungsunternehmen* sind die Ergebnisse der Schwarmalgorithmen in drei Varianten sogar etwas besser.

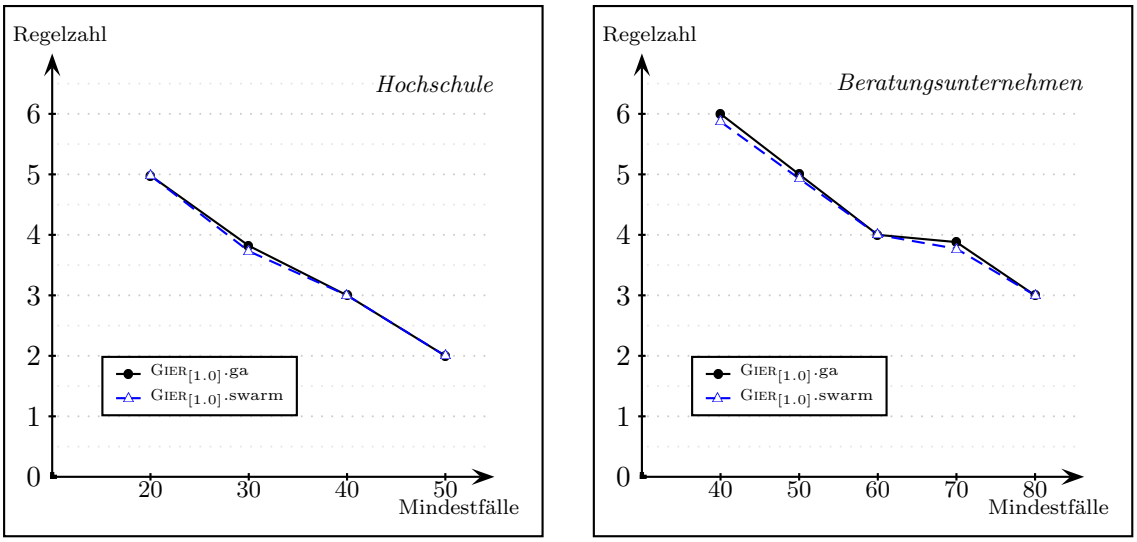


Abbildung 6.31: Vergleich der durchschnittlichen Regelzahlen für gier.ga und gier.swarm: Für jede der gier.ga- und gier.swarm-Varianten ist die durchschnittliche Regelzahl für die Datensätze *Hochschule* (50%) auf der linken und *Beratungsunternehmen* (50%) auf der rechten Seite eingetragen.

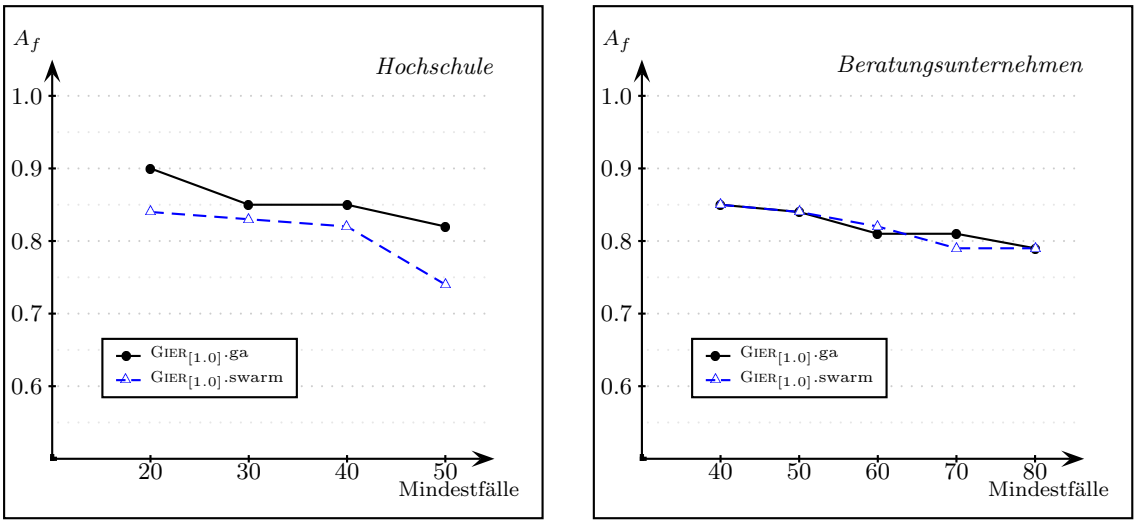


Abbildung 6.32: Vergleich der durchschnittlichen Fitting-Ergebnisse für gier.ga und gier.swarm: Für jede der gier.ga- und gier.swarm-Varianten ist die durchschnittliche Regelzahl für die Datensätze *Hochschule* (50%) auf der linken und *Beratungsunternehmen* (50%) auf der rechten Seite eingetragen.

6.3.3.2 ROCS-Diagramme für die Algorithmen

Die ROCS-Darstellung ermöglicht eine getrennte Betrachtung von Sensitivität und Spezifität im Vergleich mit Zufallsauswahl und Majoritätsregel. Abbildung 6.34 demonstriert die Ergebnisse für den Datensatz

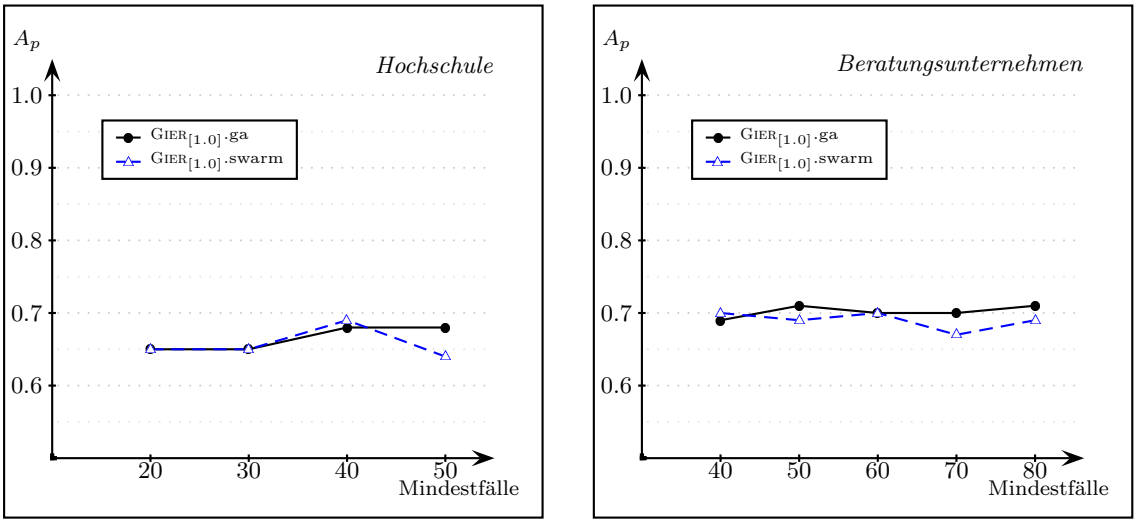


Abbildung 6.33: Vergleich der durchschnittlichen Prediction-Ergebnisse für gier.ga und gier.swarm: Für jede der gier.ga- und gier.swarm-Varianten ist die durchschnittliche Regelzahl für die Datensätze *Hochschule* (50%) auf der linken und *Beratungsunternehmen* (50%) auf der rechten Seite eingetragen.

Datensatz *Hochschule* $(n_1 = 138, n_0 = 53)$

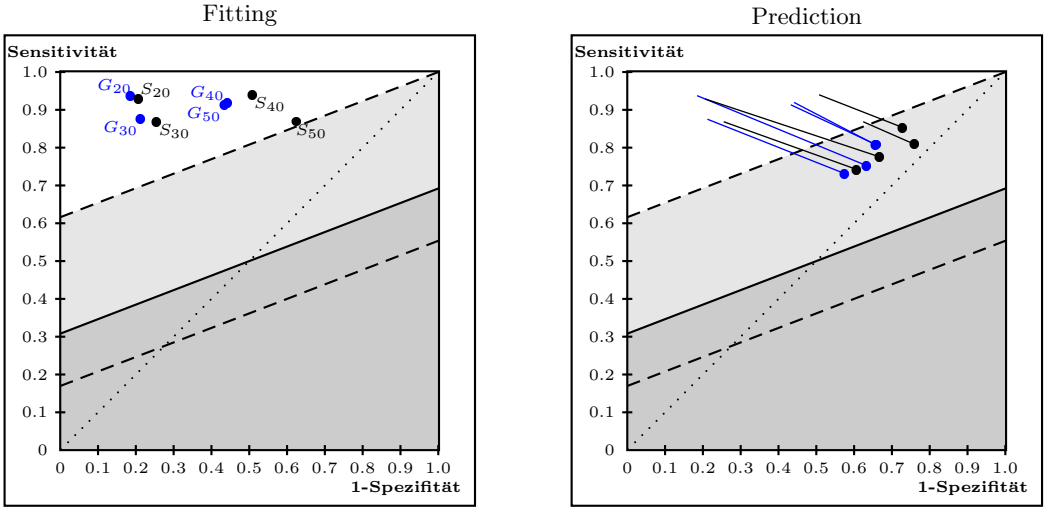


Abbildung 6.34: ROCS-Diagramm für die Ergebnisse von gier.ga und gier.swarm für den Datensatz *Hochschule*: Im Prediction-Diagramm ist jeweils die Linie zwischen Fitting- und Prediction-Ergebnis eingezeichnet, die Beschriftung wird aus Gründen der Übersichtlichkeit auf eine Diagrammhälfte begrenzt.

Hochschule, Abbildung 6.35 für den Datensatz *Beratungsunternehmen*. Dabei kann in einer ersten Würdigung bereits erkannt werden, dass die Ergebnisse in der *Prediction*-Bedingung ohne Ausnahme eine schlechtere Leistung erbringen als die Majoritätsregel, aber eine bessere als die Zufallsauswahl. Die bereits beobachtete Systematik setzt sich also nicht nur für die $\text{GIER}_{[1.0]}$.downhill-Algorithmen, sondern auch für $\text{GIER}_{[1.0]}$.ga und $\text{GIER}_{[1.0]}$.swarm fort. Für den Datensatz *Hochschule* kann weiterhin beobachtet werden, dass die Streuung der Spezifitäten größer ist als die der Sensitivitäten, was tendenziell auch für den zweiten Datensatz gilt. In der *Prediction*-Bedingung rücken die Algorithmen näher zusammen als in der Fitting-Bedingung.

Datensatz *Beratungsunternehmen* ($n_1 = 318, n_0 = 119$)

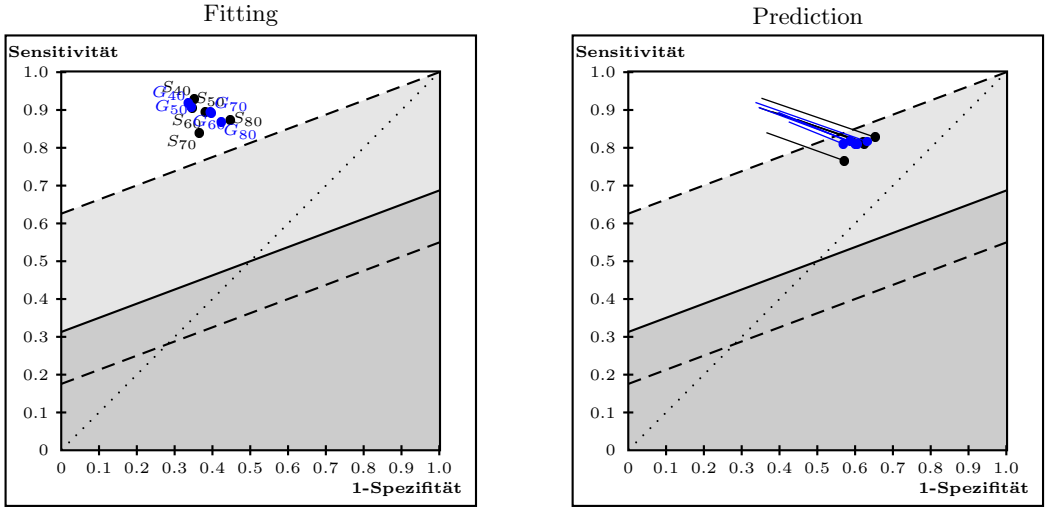


Abbildung 6.35: ROCS-Diagramm für die Ergebnisse von gier.ga und gier.swarm für den Datensatz *Beratungsunternehmen*: Im Prediction-Diagramm ist jeweils die Linie zwischen Fitting- und Prediction-Ergebnis eingezeichnet, die Beschriftung wird aus Gründen der Übersichtlichkeit auf eine Diagrammhälfte begrenzt.

6.3.3.3 Streudiagramme der Trials für den Datensatz Hochschule

Eine nähere Analyse einzelner Trials für den Datensatz *Hochschule* soll einige Systematiken verdeutlichen. In Abbildung 6.36 sind die Ergebnisse von jeweils 50 Trials der vier $\text{GIER}_{[1.0]}$.ga-Varianten als Punkte in einem Streudiagramm eingetragen. Die Kreuzungen der farbigen Linien markieren den Schwerpunkt der vier farbigen Punktewolken: Das Zusammenrücken der Ergebnisse in der Prediction-Bedingung wird auch in dieser Visualisierungsform deutlich.

Eine Reihe von Ergebnissen von $\text{GIER}_{[1.0]}$.ga 50 fällt scheinbar aus dem Rahmen: Diese Besonderheit kann erklärt werden, wenn die Menge der möglichen Ergebnisse bei Vorhersage der Majoritätsklasse berücksichtigt wird. Von den 96 Fällen im Trainingsdatensatz können maximal 53 der negativen Klasse angehören. Das schlechteste Fitting-Ergebnis bei Vorhersage der positiven Klasse wären dementsprechend $A_{f_1} = \frac{96-53}{96} \approx 0,448$ bei einem *Prediction*-Ergebnis von $A_{p_1} = 1$. Das maximale *Fitting*-Ergebnis wäre bei 96 positiven Fällen in der Trainingsstichprobe mit $A_{f_2} = 1$ erreicht, bei einem entsprechenden *Prediction*-Resultat von $A_{p_2} = \frac{95-53}{95} \approx 0,442$. Dementsprechend sind $P_1(0,448;1)$ und $P_2(1;0,442)$ mögliche Ergebnisse bei exklusiver Vorhersage der Majoritätsklasse. Alle anderen möglichen Ergebnisse liegen dabei in der Abbildung auf der Strecke $\overline{P_1P_2}$. In dieser Weise sind die scheinbaren Ausreißer des Diagramms zu erklären: Wenn beide Regeln in einem Trial von $\text{GIER}_{[1.0]}$.ga 50 eine größere Zahl von positiven Fällen abdecken, wird die negative Klasse nicht vorhergesagt. Die Ergebnisse für $\text{GIER}_{[1.0]}$.swarm sind in dieser Bedingung sehr ähnlich, die größere Trialanzahl führt hier allerdings zu deutlicheren Konturen, was in Abbildung 6.37 ersichtlich werden kann. Auch hier liegen einige Ergebnisse von $\text{GIER}_{[1.0]}$.swarm 50 auf der Majoritätsklassen-Strecke.

Die Auswirkungen eines unterschiedlichen Trainingsset-Testset-Splits können in Abbildung 6.38 nachvollzogen werden. Hier verändert sich auch die Lage der Majoritätsstrecke dementsprechend. Aufgrund der

Hochschule (50%), GIER$_{[1.0]}$.ga

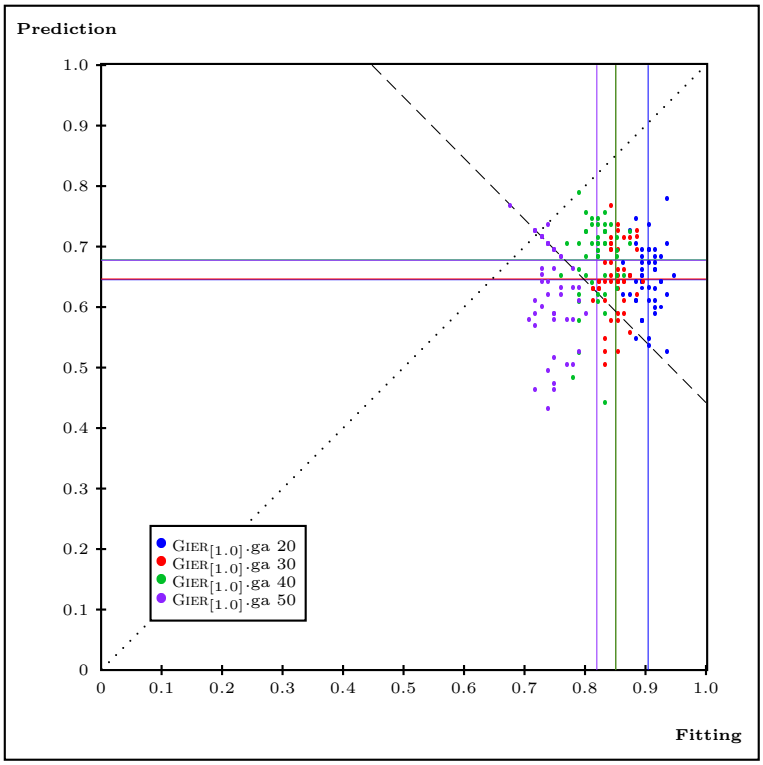


Abbildung 6.36: Streudiagramm der Ergebnisse von gier.ga für den Datensatz *Hochschule* (50%): Für jede Variante sind Fitting- und Prediction-Ergebnisse von 50 Trials als Punkt im Diagramm markiert. Die farbigen Linien entsprechen den Mittelwerten der Varianten für Fitting und Prediction. Die gestrichelte Linie markiert die möglichen Ergebnisse bei Vorhersage der Majoritätsklasse.

Hochschule (50%), GIER$_{[1.0]}$.swarm

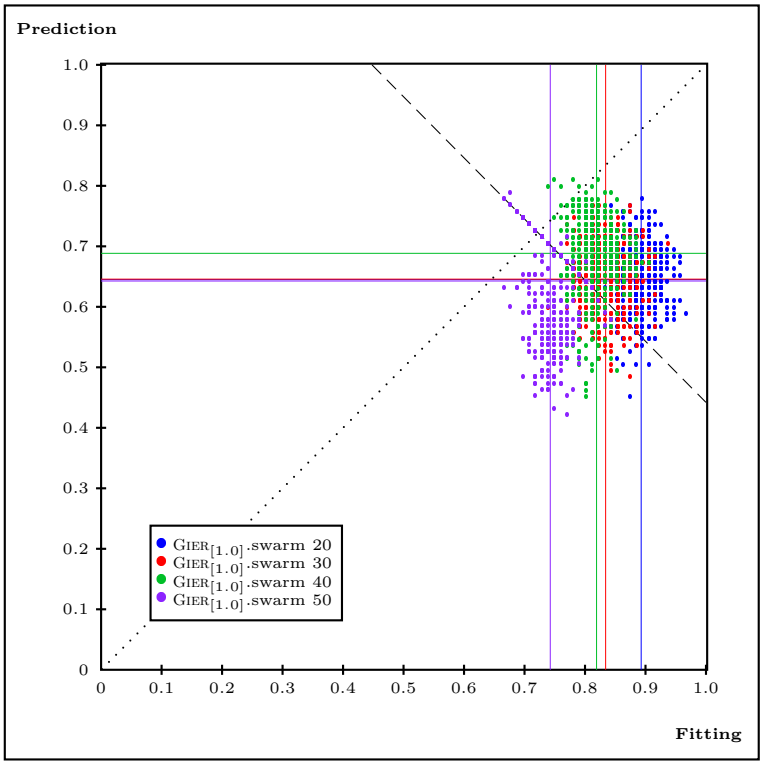


Abbildung 6.37: Streudiagramm der Ergebnisse von gier.swarm für den Datensatz *Hochschule* (50%): Für jede Variante sind Fitting- und Prediction-Ergebnisse von 500 Trials als Punkt im Diagramm markiert. Die farbigen Linien entsprechen den Mittelwerten der Varianten für Fitting und Prediction. Die gestrichelte Linie markiert die möglichen Ergebnisse bei Vorhersage der Majoritätsklasse.

kleineren Teststichprobe ist die Anzahl der möglichen Generalisierungsergebnisse kleiner und die Punktwolke wirkt spärlicher. Gleichzeitig ist die Streuung der Prediction-Ergebnisse aufgrund der Stichprobenzahl größer, die der Fitting-Ergebnisse aus gleichem Grund kleiner. Insgesamt weist hier GIER$_{[1.0]}$.swarm 60 einen leichten Vorteil gegenüber den Konkurrenten auf.

Hochschule (80%), GIER$_{[1.0]}$.swarm

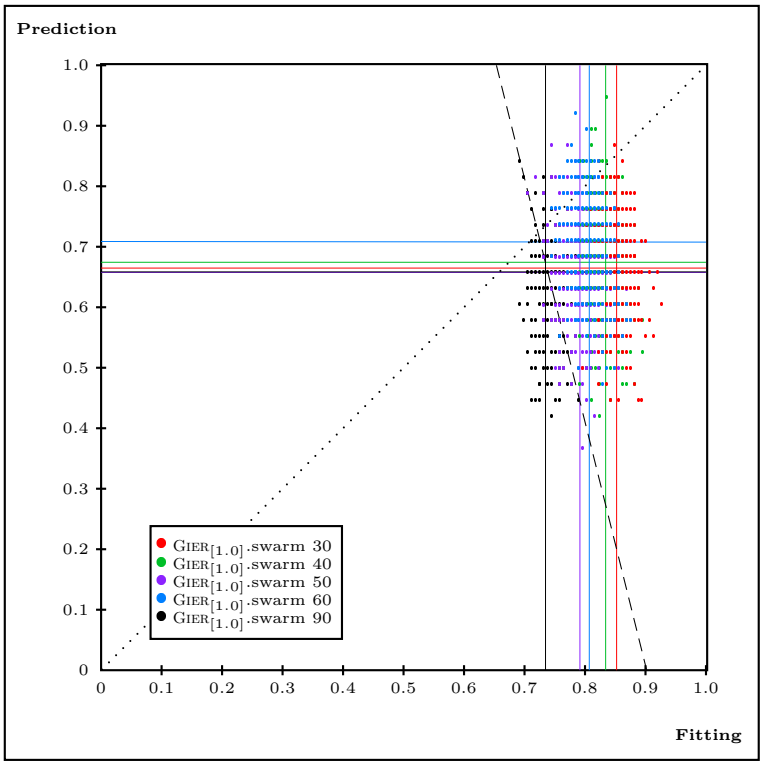


Abbildung 6.38: Streudiagramm der Ergebnisse von gier.swarm für den Datensatz *Hochschule* (80%): Für jede Variante sind Fitting- und Prediction-Ergebnisse von 500 Trials als Punkt im Diagramm markiert. Die farbigen Linien entsprechen den Mittelwerten der Varianten für Fitting und Prediction. Die gestrichelte Linie markiert die möglichen Ergebnisse bei Vorhersage der Majoritätsklasse.

6.3.3.4 Vergleich mit konkurrierenden Algorithmen

Zuletzt sollen die Algorithmen aber auch mit den übrigen Verfahren der Konkurrenz verglichen werden. Hierzu sind die Ergebnisse aller Algorithmenvarianten für die drei Datensätze hinsichtlich des *Fitting* in Tabelle 6.20 zusammengetragen, hinsichtlich der Generalisierung in Tabelle 6.21.

Sollte tatsächlich ein Gewinner der Konkurrenz gekürt werden müssen, würde die Wahl auf den Algorithmus FaF MaxVal fallen, der seinen Sieg aber allein aufgrund der Tendenz, nahezu ausschließlich die Majoritätsklasse zu wählen, in zweifelhafter Weise verdienen würde. Dementsprechend wäre der eigentliche Titel hier wohl an die logistische Regression abzutreten. FaF ZigVal Fold und CART zeigen vergleichbare Ergebnisse in der Generalisierung, wobei CART im *Fitting* dominiert, was auf Overfitting hindeutet.

Die Ergebnisse von GIER$_{[1.0]}$.ga und GIER$_{[1.0]}$.swarm fallen im Allgemeinen hinter den Ergebnissen der Konkurrenzverfahren zurück. Einzelne Varianten können hingegen im Datensatz *Hochschule* ähnliche Generalisierungsleistungen anbieten wie die logistische Regression, was insofern bemerkenswert ist, als dass es sich in den Ergebnissen um hierarchische Regelmengen handelt. In jedem Fall sind die gewonnenen Bäume bzw. Regelmengen in der Vorhersage leistungsstärker als CART mit dem Problem der Wahl des Regelgrößenparameters. Würde diese Wahl aufgrund der Generalisierungsergebnisse erfolgen, dann wäre die Konkurrenz von einem Downhill-Algorithmus gewonnen worden: Dies würde aber eine Verletzung des Prinzips darstellen, ausschließlich die Trainingsstichprobe für die Gewinnung von Ergebnissen zu verwenden. Sollte die Parameterwahl also mit Hilfe der Trainingsstichprobe erfolgen, müsste ein Teil

der Trainingsstichprobe als Testdatensatz zur Seite gelegt werden. Die Schwankungen der Ergebnisse demonstrieren aber, dass die Wahl des Parameters einen deutlichen Einfluss haben kann.

Tabelle 6.20: Fitting-Ergebnisse aller Algorithmen

		Datensatz		
		Hochschule		*Beratungsunternehmen*
Algorithmus	Mindestfallzahl	50%:50%	80%:20%	50%:50%
GIER$_{[1.0]}$.GA	20	90	-	-
	30	85	-	-
	40	85	-	85
	50	82	-	84
	60	-	-	81
	70	-	-	81
	80	-	-	79
GIER$_{[1.0]}$.swarm	20	89	-	-
	30	83	85	-
	40	82	83	85
	50	74	79	84
	60	-	81	82
	70	-	-	79
	80	-	-	79
	90	-	73	-
LR	-	78	76	78
CART	-	92	92	92
FaF MaxVal	-	72	72	73
FaF ZigVal Fold	-	71	70	71

6.4 Diskussion der Ergebnisse

Abschließend soll in der Überschau versucht werden, ein Fazit zu ziehen bezüglich der vorgenommenen methodischen Überprüfungen und insbesondere die methodisch betonte Klasse der *Fast and Frugal Trees* als Klassifikationsmethode für die Psychologie zu positionieren.

6.4.1 Fazit des Methodenvergleichs

Mit den Betrachtungen der Ergebnisse wird deutlich, warum selbst die Ergebnisse über insgesamt 33 Datensätze vorsichtig interpretiert werden müssen. Die Auswahl der Datensätze erfolgte im Rahmen dieser Arbeit ohne vorherige Prüfung der Ergebnisse der Algorithmen. Die große Streubreite möglicher Ergebnisse würde aber eine selektive Auswahl von Datensätzen für einen Erfolg beinahe jedes Verfahren der Konkurrenz ermöglichen, so dass die Qualität der Ergebnisse durch Auswahl entsprechender Probleme nahezu frei manipuliert werden könnte.

Dementsprechend sind die Ergebnisse eher als eine Art Existenzbeweis zu betrachten: Es gibt Datensätze und Problemstellungen, bei denen sehr einfache Heuristiken in Form der *Fast and Frugal Trees* bessere Ergebnisse erbringen als die klassischen Algorithmen, die als Default oft eingesetzt werden. Weiterhin ist an den Ergebnissen bemerkenswert, dass weder aus *Fitting*-Ergebnissen noch aus beliebig ausgewählten *Prediction*-Ergebnissen direkte Auswahlempfehlungen für einen Algorithmus abgeleitet werden können. Im Gegenteil, das differenzielle *Overfitting* der Verfahren würde in diesem Fall sogar bei einer Nutzung der Fitting-Informationen zur Wahl eines schlechteren Verfahrens führen. Die *Leave-One-Out*-Kreuzvalidierung ist hingegen für solche Fälle eine geeignete Auswahlmethode, in der die vorhandene Stichprobe auch der Größe der Stichprobe entspricht, auf deren Basis vorhergesagt werden soll[9]. *Fit-*

[9]Wobei auch die Leave-One-Out-Kreuzvalidierung in bestimmten Anwendungsfällen irreführende Ergebnisse erbringen

Tabelle 6.21: Prediction-Ergebnisse aller Algorithmen

		Datensatz		
		Hochschule		*Beratungsunternehmen*
Algorithmus	Mindestfallzahl	50%:50%	80%:20%	50%:50%
GIER$_{[1.0]}$.GA	20	65	-	-
	30	65	-	-
	40	68	-	69
	50	68	-	71
	60	-	-	70
	70	-	-	70
	80	-	-	71
GIER$_{[1.0]}$.swarm	20	65	-	-
	30	65	66	-
	40	69	67	70
	50	64	66	69
	60	-	71	70
	70	-	-	67
	80	-	-	69
	90	-	66	-
LR	-	69	71	73
CART	-	66	67	66
FaF MaxVal	-	72	72	73
FaF ZigVal Fold	-	66	65	70

ting-Ergebnisse sind vielleicht für Fragen einer ökonomischen Datenzusammenfassung interessant, ohne eine Absicherung durch *Prediction*-Ergebnisse sind sie aber schwer zu interpretieren. Gleiche Argumente können in vielen Fällen gegen den unkritischen Einsatz linearer Regressionsmodelle eingebracht werden.

Die Analyse der ökologischen Rationalität von Klassifikationsalgorithmen ist allerdings mit einer einzelnen Simulationsstudie nicht zu leisten. Die analytische Aufbearbeitung der Fragestellung und die Entwicklung formaler Grundlagen für die Entscheidungsfindung hinsichtlich der Algorithmenwahl bleibt ein Desideratum der Forschung. Hier bieten sich Studien zu kleinen kontrollierbaren Datensätzen an sowie die formale Suche nach Vergleichbarkeiten und Differenzen zwischen Algorithmenklassen. Was allerdings geleistet werden kann, ist das Aussprechen einer Empfehlung, in die Auswahlmenge der Algorithmen einfache Heuristiken aufzunehmen. Auch der GIER$_{[1.0]}$.downhill-Algorithmus zeigt ein Potenzial für die Exploration der Struktur von Datensätzen. Hier könnte zudem erwogen werden, ob sich die Downhill-Ergebnisse als Grundlage für die Parametersetzung bei genetischen Algorithmen oder Schwarm-Algorithmen anbieten könnten.

6.4.2 Anwendungsempfehlungen und Modifikationsmöglichkeiten für Fast and Frugal Trees

Die relative Leistungsstärke von *Fast and Frugal Trees* gegenüber klassischen Verfahren konnte vor allem für kleinere Datensätze demonstriert werden. Die Größe der untersuchten Stichproben ist für diagnostische Fragestellungn als durchaus typisch zu beurteilen. Insbesondere für den Fall, dass sich die Größe der Informationsbasis im Laufe der Zeit ändert (was bei diagnostischen Verfahren vorkommen sollte), kann der Wechsel des Prädiktionsverfahrens ab einer gewissen Stichprobengröße vielversprechend sein. Für diesen Anwendungsfall und bei kleineren Stichproben können *Fast and Frugal Trees* eine gute Klassifikationsmethode darstellen. Zwei Variationsmöglichkeiten für *Fast and Frugal Trees* sollen abschließend diskutiert werden.

kann.

6.4.2.1 Erste Variation: Binarisierung und mögliche Erweiterungen

Es soll auf den Umstand zurückgegangen werden, dass *Fast and Frugal Trees* binäre Prädiktoren voraussetzen. Kontinuierliche Prädiktoren wurden in dieser Studie generell mediandichotomisiert. Um damit nicht einen ungerechtfertigten Vorteil zu erzeugen, wurde die Mediandichotomisierung immer anhand der jeweiligen Trainingsstichprobe vorgenommen. Um natürlich binäre Prädiktoren aber als solche behandeln zu können, wurde für Variablen mit den zwei Werten „0“ und „1“ der Median automatisch auf den Mittelwert der Werte gesetzt [10]. Es mag auch für kontinuierliche Prädiktoren „natürlichere“ Arten der Dichotomisierung geben, sie erfordern aber weiteres (fallibles) Wissen. Die Mediandichotomisierung führte zudem zu ungefähr gleich großen Gruppen mit divergierenden Prädiktorwerten. Das hat zum einen zur Folge, dass bereits im ersten Schritt über die Hälfte der Fälle entschieden wird, zum anderen aber auch, dass die Validitätswerte auf ähnlicher Grundlage berechnet werden.

Theoretisch könnten *Fast and Frugal Trees* aber auch für mehrstufige kategoriale Prädiktoren konstruiert werden, die notwendigen Veränderungen würden ausschließlich das Konstruktionsverfahren, nicht aber die Baumstruktur verändern: Für jede Kategorie des Prädiktors müssten zwei Validitäten berechnet werden. Bei zwei Stufen wurde so vorgegangen, dass positive und negative Validität eindeutig jeweils einer Ausprägung zugeordnet waren, indem die Richtung des Cues festgelegt wurde. In diesem Fall würden bei drei Ausprägungen sechs Validitäten zu berücksichtigen sein, drei positive und drei negative. Ein MaxVal-Verfahren würde zuerst nach dem höchsten gefundenen Wert suchen und damit den Cue entsprechend der Art der Validität polen. Die jeweils höchsten Werte der Cues werden in eine Rangreihe gebrachtm und der Baum wäre konstruiert. Das ZigVal-Verfahren würde hingegen alternierend die höchste der drei positiven Validitäten oder die höchste der drei negativen Validitäten berücksichtigen. Die damit gewählten Kategorien werden genutzt, die Cues zu binarisieren, so dass zusammen mit der Polung die Daten so transformiert werden können, dass in der Anwendung keine Unterschiede zum Standardvorgehen anfallen. Differenziertere Varianten würden positive und negative Validitäten auch für Mengen von Kategorien mit mehr als einem Element berechnen, allerdings ist aufgrund der kombinatorischen Explosion hier wieder eine Begrenzung der Kategorienzahl oder Suchtiefe sinnvoll. Die Cueabfrage hätte hingegen weiterhin die Form „Ist die Ausprägung von Cue k gleich x?“ oder „Ist die Ausprägung von Cue k ungleich x?“ je nach Polung des Prädiktors[11]. Ein solches Verfahren würde zu einem kategorienbasierten *Fast and Frugal Tree* führen, dessen Konstruktionsverfahren aufwändiger geworden ist. Die Anwendung ist von dieser Änderung aber nicht betroffen, wenn berücksichtigt wird, dass die Teilmengen, die pro Schritt klassifiziert werden, unter Umständen kleiner werden, so dass mehr Prädiktorwerte abgefragt werden müssen.

Letztendlich ließe sich auf einem ähnlichen Weg auch ein *FaF Tree* für mehr als zwei Kategorien des Kriteriums entwickeln. Dementsprechend müssten so viele Validitätsarten wie Kategorien eingeführt werden. Für jede Kategorie des Prädiktors werden alle Validitätsarten berechnet. Ein Max-Algorithmus würde die höchste Validität für eine Entscheidung in Richtung der durch die Validitätsart angegebenen Kategorie auswählen, ein Zig-Algorithmus würde zwischen Kategorien alternieren. Es wäre hier auch ein ZigMax-Algorithmus denkbar, der unter den Entscheidungsrichtungen, in deren Richtungen am wenigsten oft entschieden wurde, die mit der maximalen Validität auswählt, welche durch einen beliebigen verbleibenden Prädiktor demonstriert wird. Andererseits müsste die Verwendung eines Prädiktors auf einer Stufe für eine bestimmte Entscheidungsart seine Verwendung für andere Entscheidungsarten nicht ausschließen. Somit sind auch mehrkategorielle Klassifikatoren als *Fast and Frugal Tree* mit einfachen Konstruktionsverfahren konstruierbar. Diese Bäume würden weiterhin *One-Reason Decision Making* exemplifizieren, allerdings ginge die Eleganz des eindimensionalen lexikografischen Vergleichs mit einem Splittingprofil verloren. Der Baum wäre noch als binärer Entscheidungsbaum darstellbar, aber nur, wenn die Bindung der Richtungen von Entscheidungsknoten an Entscheidungsarten aufgegeben würde und die Endknoten die Entscheidungsart auf andere Weise spezifizieren könnten.

[10] In wenigen Datensätzen (*Hepatitis*, *Education*) wurden einige binäre Variablen aufgrund anderer Kodierung als kategoriale behandelt, dies allerdings bei allen Bäumen in gleicher Weise. Ein Vergleich mit rekodierten Ergebnissen zeigt für einige Verfahren Ergebnisschwankungen, allerdings nicht bei den beiden näher analysierten Varianten. Für binäre Variablen ist eine Mediandichotomisierung bspw. wenig zielführend. Allerdings sollen auch nicht fälschlicherweise Variablen als binär deklariert werden, wenn zufällig nur zwei Werte in der Trainingsstichprobe auftreten.

[11] Die positive Antwort auf die erste und die negative Antwort auf die zweite Frage implizieren, dass die Prädiktorvariable den ausgewählten Wert aufweist.

6.4.2.2 Zweite Variation: Cuereihenfolge versus Entscheidungsreihenfolge

Die Überlegungen für den mehrkategoriellen Fall bezüglich der Mehrfachverwendung von Cues sind auch für den zweikategoriellen Fall relevant: Angenommen, es gibt einen Cue mit $V_1^+ = V_1^- = 1$ und einen zweiten Cue mit $V_2^+ = V_2^- = 0,6$: Jedes der Baumverfahren würde auf der zweiten Ebene den zweiten Cue verwenden, auch wenn der erste Cue allein zu einer perfekten Klassifikation der Daten führt. Sobald ein Cue im binären Fall für beide Entscheidungsrichtungen verwendet wird, sind alle späteren Ebenen des Baums für die Entscheidungen irrelevant, da *für jeden Fall entweder ein positiver oder ein negativer Cuewert existieren muss.* Wenn ich aber im ersten Schritt eine positive Entscheidung allein auf den positiven Cuewert stütze, kann ich mit der gleichen Berechtigung alle Entscheidungen auf diesen Cue stützen. Ein Konstruktionsverfahren, das diesen Fall zulassen soll, würde bei einer Verwendung eines Cues für eine Entscheidungsrichtung den für die erste Verwendung des Cues irrelevanten Validitätswert nicht automatisch aus der Auswahlmenge nehmen. Ein MaxVal-Algorithmus würde damit nicht die Maxima aus V_i^+ und V_i^- in eine Reihenfolge bringen, sondern alle Einzelwerte. Damit wird die Anzahl der Cues nicht automatisch zur Ebenenzahl des Baumes unterhalb der Wurzel, sondern die Cueanzahl stellt nur noch den maximal möglichen Wert dar. Die zweite Verwendung des gleichen Cues führt zu einem Abbruch der Suchprozedur. Die entsprechende Modifikation der ZigZag-Verfahren ist naheliegend analog. Im Extremfall würde nur noch eine Ebene unterhalb des Wurzelknotens existieren und ein Cue allein entscheidet für jeden Fall definitiv über die Klassifikation.

Eine derartige Variante ist - wie auch die zuvor beschriebenen Varianten - derzeit noch nicht Teil der Software, die Modifikationen konnten damit noch nicht geprüft werden, sie stellen aber sicherlich den Inhalt einer zukünftigen Forschungsstudie dar. Dies soll die Darstellung der Ergebnisse der Simulationsstudie abschließen. Im letzten Kapitel soll die Rationalität einfacher Heuristiken für die Regelinduktion diskutiert werden.

- Ah! Voilà un sujet, s'écria il aperçut le petit prince.
Et le petit prince se demanda:
-Comment peut-il me connaître puisqu'il ne m'a encore jamais vu!
Il ne savait pas que, pour le rois, le monde est très simplifié. Tous les hommes sont des sujets.

Antoine de Saint-Exupery, Le Petit Prince

7 Die paradoxe Rationalität einfacher Klassifikationsverfahren

In diesem Kapitel soll die Beziehung zwischen einfachen Heuristiken und dem *Frame*-Problem herausgearbeitet werden. Im Anschluss wird mit Hilfe einer Modellbildung reflektiert, wie weit die in dieser Arbeit entwickelten Lösungsansätze reichen können. Abschließend sollen epistemologische Implikationen der Ergebnisse diskutiert werden.

7.1 Einfache Heuristiken und das Frame-Problem

7.1.1 Roboter, die Kognitionswissenschaft und das Frame-Problem

Dennett (1998) erläutert eine Variante des *Frame*-Problems anhand einer Robotergeschichte: Der Roboter $R1$ lernt, dass seine Ersatzbatterie auf einem kleinen Wagen in einem Raum eingeschlossen worden ist, in dem eine Zeitbombe aktiviert wurde, die bald explodieren wird. $R1$ hat Raum und Schlüssel lokalisiert und formuliert nun einen Plan zur Rettung der Batterie, der basal aus dem Herausziehen des Wagens bestehen soll. Er setzt den Plan in die Tat um, um von der Zeitbombe, die sich ebenfalls auf dem Wagen befand, in seine Einzelteile zerlegt zu werden. Ein neues Modell - $R1D1$[1] - wird deshalb mit einem Deduktionsprogramm ausgestattet und lernt, über Nebeneffekte nachzudenken. Aber auch dieser Roboter wird von der Explosion erfasst, während er deduziert, dass das Herausziehen des Wagens nicht die Farbe des Raumes verändern wird. $R2D1$ - das Folgemodell - wird deshalb programmiert, irrelevante Implikationen zu ignorieren. Seine Fortschritte sind allerdings noch unbefriedigender, da er Tausende Implikationen als irrelevant erachtet und in die Liste der zu ignorierenden Implikationen überführt hat, als die Explosion die Liste und den Roboter zerstört. Für Dennett (1998) werden alle Roboter ein Opfer des *Frame*-Problems. Die Bemessung von Relevanz fällt Menschen leichter als den Maschinen und ist auch unabdingbar für wissenschaftliches Arbeiten:

> ...the number of satellites orbiting the earth is clearly irrelevant in assessing the rational credibility of the hypothesis that certain forms of social conflict enhance social cohesion. (McLaughlin, 1970, S. 274)

Damit argumentiert McLaughlin (1970) bspw. gegen das Rationalitätskriterium, eine vollständige Informationssammlung vor einer Entscheidung abgeschlossen haben zu müssen. Das Dilemma entsteht, wenn berücksichtigt wird, dass die Einschätzung von Relevanz nicht auf rationalem Wege erfolgen kann. Versteht man den Kognitivismus als die Vorstellung interner kognitiver Prozesse in Begriffen von Regeln und Repräsentationen, die in symbolischer und propositionaler Form definiert sind, ergibt sich für Costall und Still (1991) das Frameproblem als eine der grundlegenden Klippen des Ansatzes:

[1] Dennett (1998) spielt hier eindeutig auf den *Star Wars*-Roboter $R2D2$ an.

> There is the problem of *solipsism*: how can the knower ever reach beyond internal representations to the reality they are supposed to represent? There is the problem of *development*: how can a system of formal rules ever be flexible enough to capture the mutuality between a growing organism and its richly structured and changing environment? And there ist the problem of *relevance*: how does anyone following a rule know when to apply that rule? (The so-called #frame# problem.) Most generally, there is the problem of *meaning*: how do symbolic representations attain their semantic status? (Costall & Still, 1991, S. 2)

Von kognitiver Seite scheint sich zunächst keine Lösung für die Herausforderung durch das *Frame*-Problem abzuzeichnen, was Costall und Still (1991) auf ein fehlendes Problembewusstsein zurückführen:

> Of the possible defensive options available, the main one chosen has been the 'no' alternative option, the insistence that cognitivism is the only game in town. With this goes an airy optimism, a confidence that these problems will eventually be solved, and hence are not *really* problems at all[...]. (Costall & Still, 1991, S. 2)

Eine dieser Einschätzung entgegenlaufende Haltung demonstriert hingegen Fodor. Während Fodor (2000) lokale mentale Prozesse mit der Vorstellung von Denken als Komputation kompatibel sieht und diese zudem für modular strukturiert und angeboren hält, kann er diese Eigenschaften nicht für globale Kognition finden, was für ihn Anlass zu einer pessimistischen Prognose für die Kognitionswissenschaft gibt:

> Since the mental processes thus afflicted with globality apparently include some of the ones that are most characteristic of human cognition, I'm on balance not inclined to celebrate how much we have so far learned about how our minds work. The bottom line is that the current situation in cognitive science is light years from being satisfactory. Perhaps somebody will fix it eventually; but not, I should think, in the foreseeable future, and not with the tools that we currently have in hand. (Fodor, 2000, S. 5)

Das *Frame*-Problem diskutiert Fodor (2000) in Bezug auf abduktives Schließen, also das Schließen auf die beste Erklärung. Im abduktiven Schließen entsteht ein Konflikt aus der Notwendigkeit, den gesamten Hintergrund epistemischer Überzeugungen potenziell berücksichtigen zu müssen, aber gleichzeitig der pragmatischen Notwendigkeit, dass nur eine kleine Teilmenge *relevanter* Hintergrundüberzeugungen konsultiert wird:

> How to make abductive inferences that are both reliable and feasible is what they call in AI frame problem. No doubt the claim is tendentious [...], but I think it's plausibly *because* of the frame problem that our robots don't work. After all robots are mostly computing machines. So if a lot of quotidian cognition is abductive, and if there are intrinsic tensions between abduction and computation, *why would you even expect that our robots would work?* (Fodor, 2000, S. 38)

Damit sieht auch Fodor das Gebiet der komputationalen Theorie geprägt von einer Unterdrückung des Gedankens, dass die Begrenztheit der Theorie zu einem Misserfolg des Programms führen konnte, denn er sieht die Lösung des *Frame*-Problems als zentrale Aufgabe:

> Cognitive science minus the syntactic theory of computation is Hamlet without the Prince. But cognitive science minus the frame problem is Hamlet without anybody much except Polonius. (Fodor, 2000, S. 42)

Den naheliegenden Schritt zu Hierarchie, mit dem etwa Russell erfolgreich das Mengenparadox durch das Klassenkonzept bekämpfen konnte, scheint für Fodor die Kognitionswissenschaft nur einen begrenzten Erfolg erwarten zu lassen, ebenso sieht er die Möglichkeit von Beratung begrenzt, wenn die Einschätzung der Qualität des Beraters und seiner Ratschläge der Lösung des Problems vorangestellt werden muss.

Wenn Pinker (2005) versucht, individuelles und wissenschaftliches Denken zu trennen, kontert Fodor (2005), indem er auf die Inkonsequenz einer Eliminierung des Individuellen aus der Wissenschaft verweist. Fodor plädiert dafür, die Forschungsbemühungen auf Bereiche zu konzentrieren, in denen der Einfluss

der Globalität minimal ist, so dass sie aufgrund eines vernünftigen Standards an wissenschaftlicher Einsicht ignoriert werden dürfen. Oaksford und Chater (1998) betrachten aber diese Einschränkung dann als problematisch, wenn sie zu einer künstlichen Vereinfachung des Untersuchungsgegenstandes hin zu „Spielzeugproblemen“ - also fiktiven Mikrowelten mit geringen Vorannahmen - einlädt:

> [...] current theories in the psychology of reasoning may be successful only because they deal directly just with data from highly simplified laboratory tasks. [...] Another way of expressing this concern is that current theories of reasoning lack „ecological validity“. (Oaksford & Chater, 1998, S. 98)

Die Autoren sind skeptisch gegenüber der Annahme, dass mit Hilfe von Erkenntnissen aus Spielzeugbereichen die alltägliche Praxis menschlicher Inferenz auf dem Weg der Skalierung erreicht werden könnte:

> Theories of inference that are adequate in such domains [...,] tend to fail disastrously when they are *scaled up* to deal with real world inferential problems involving more information [...]. This is because they are generally computationally intractable. (Oaksford & Chater, 1998, S. 125)

Fodor (2005) diagnostiziert das Versagen künstlicher Intelligenz bei Aufgaben wie dem Überqueren einer Straße, der Entscheidung über den Ort der Mittagspause, dem Verfolgen und erst recht der Konstruktion von Argumenten. Die Simulationen des Alltagsdenkens haben bisher für Fodor eine ähnliche Erfolgsquote zu verbuchen wie jene anderen Versuche, und seine abschließende Herausforderung an Pinker kann auch als Aufforderung an die ABC-Gruppe gedeutet werden:

> So far, when the architectural assumptions of Classical cog. sci. have been applied to the problems of commonsense cognition, the experiments haven't come out well; adjectives like 'appalling' spring to mind. Having said that, I should add that if Pinker really doubts that 'human beings have a power of reliable abductive inference that is better than heuristic' then he should, by all means, continue his search for the heuristics that they aren't better than. Better him than me. (Fodor, 2005, S. 27 f.)

7.1.2 Forsters Komplexitätserhaltungssatz

Das *Frame*-Problem soll hier aus der Perspektive des ABC-Ansatzes diskutiert werden. Dazu soll das Konzept der Einfachheit einer Prüfung unterzogen werden, dessen Verbindung zu versteckter Komplexität im Kontext von *Frame*-Problem und *simple heuristics* betrachtet werden soll. Weiterhin wird die Frage der Heuristikauswahl im ABC-Programm im Hinblick auf diese Kopplung von Einfachheit und Komplexität geprüft, die abschließend im weiteren Kontext von Wissenschaft und Forschung thematisiert werden soll. Forster (1999) geht von einem Erhaltungssatz der Komplexität in folgender Form aus:

> ... there is a conservation law for complexity that says that simplicity achieved at the higher level is at the expense of complexity at a lower level of implementation. (Forster, 1999, S. 545)

Aus diesem Grund interpretiert er die Einfachheit einfacher Heuristiken im Fall ihres Erfolges nicht als voraussetzungslos:

> ... fast and frugal heuristics emerge from an underlying complexity in a process of evolution that is anything but fast and frugal. Simplicity is just the tip of a very large iceberg. The underlying complexity is important in all these fields, but especially in machine intelligence and AI. (Forster, 1999, S. 543)

Um zum Eisberg durchzustoßen, untersucht Forster die Rationalität von Überzeugungen in der Rekonstruktion des ABC-Ansatzes und kommt spekulativ auf eine Variation eines reliabilistischen Kriteriums zurück, wie es in ähnlicher Form von Nozick (1993) eingesetzt wird:

> How do we characterize the rationality of our beliefs? Perhaps Gigerenzer and colleagues would give the [...] answer, that rational beliefs are those that are formed on the basis of learning rules that are fast, frugal, and ecologically valid. (Forster, 1999, S. 544)

Eine Überzeugung, die aufgrund eines zeitraubenden und informationsintensiven Such- und Entscheidungsvorgangs gewonnen wurde, der im Allgemeinen nützliche Überzeugungen produziert, kann allerdings *für sich betrachtet* nicht weniger rational sein als die gleiche Überzeugung, die durch eine schnelle und frugale Heuristik gewonnen wurde. Zu den relevanten Merkmalen des Prozesses bei der Beurteilung der Rationalität von gewonnenen Überzeugungen sind letztendlich nur Qualitätsmaße der Überzeugungen selbst relevant - anders als in der Bewertung des Verhaltens zur Gewinnung dieser Überzeugungen, deswegen erscheinen „fast“ und „frugal“ in dieser Definition problematisch. Interessanter ist aber Forsters Antwort auf die Frage, ob sich durch den Erfolg alltäglicher menschlicher Entscheidungsfindung durch einfache *fast and frugal heuristics*, die zudem als Computerprogramm formalisierbar sind, nicht die Automatisierung dieser Entscheidungen durch Rechner anbieten kann: Auch hier bleibt Forster skeptisch, da er Entscheidungsregeln erneut als oberste Ebene einer riesigen kognitiven Infrastruktur betrachtet, zu der sowohl die Eigenschaften des komplexen menschlichen Gehirns als auch die Lernerfahrung des Individuums gehören:

> Witness the complexity of the human brain. It is a hugely complex apparatus whose job it is to prepare for rational action in real time. A decision rule may be written simply in a natural language, but to implement the rule one at least needs to know the language. Thus simple rules may be an essential feature of our intelligence (simple heuristics that make us smart), but before we can act on the rule, we go through decades of learning and education. (Forster, 1999, S. 547)

Es wird dabei allerdings nicht vollständig deutlich, was als eine „natürliche Sprache“ verstanden werden kann und in welchem Verhältnis phylogenetisch erworbene Hirnstrukturen zu ontogenetisch erworbenen stehen sollen. Für den Fall der Rekognitionsheuristik formuliert Forster etwa:

> While such recognition-based decision making is simple at one level, it requires complex recogniton machinery underneath and many years of accumulated experience to work. (Forster, 1999, S. 547)

Chater, Oaksford, Nakisa und Redington (2003) wenden verstärkend ein, dass eine Heuristik wie TTB nur in einer seriellen Architektur als schneller und einfacher verstanden werden kann als konkurrierende Heuristiken. In anderen Bereichen scheinen für die Autoren dementsprechend parallel verarbeitende Heuristiken wahrscheinlicher, wenn die schnelle Integration großer Informationsmengen gefragt ist wie in der Sprachverarbeitung, Wahrnehmung, der Motorik oder dem Alltagsdenken. Forster (1999) bindet diese Überlegung direkt an eine Formulierung des *Frame*-Problems an, indem er sich auf Dennetts Robotergeschichten bezieht und erneut die Bedeutsamkeit von Kontingenzen unterstreicht, mit denen Menschen - im Gegensatz zu Maschinen - geringe Schwierigkeiten haben: Menschen generalisieren auf Situationen, die sie nie erlebt haben, obwohl mehr Kontingenzen berücksichtigt werden könnten als sich zählen ließen. Forster verbindet diesen Gedanken mit dem Konzept des Bias im Maschinenlernen, mit dem die notwendige Voreingenommenheit von Lernalgorithmen beschrieben wird:

> In machine learning, the slogan is that every machine has a built-in bias, which allows it to succeed in one kind of learning problem at the expense of failing in others. The idea is that there are no free lunches in learning - if learning succeeds, then it is because of design features built into the machine. (Forster, 1999, S. 551 f.)

Insbesondere bezieht sich Forster auf die *No-Free-Lunch*-Theoreme (Wolpert, 1992, 1995, 1996), in denen bewiesen wird, dass bei der Vergabe von gleichen Wahrscheinlichkeiten für alle möglichen Universen jede prädiktive Hypothese die gleiche A-Priori-Wahrscheinlichkeit für Erfolg aufweist:

> The importance of the no-free-lunch theorems is that they show what epistemology is not. One cannot begin from a blank slate (with nothing more than a knowledge of the space of possible universes). One must begin with an a priori prejudice (a „bias“ as the machine learning theorists call it) before the knowledge seeking process can get off the ground. In that sense, there is no free lunch, which is where the theorems get their name. (Forster, 1999, S. 552 f.)

Dementsprechend kann nicht zwischen prädiktiven Hypothesen differenziert werden, wenn man nicht mit der Differenzierung beginnt, etwa in der Form von Annahmen wie dem Markov-Prinzip[2] oder der Stationarität[3]. Damit wird es nun aber unmöglich, die ökologische Validität von Selektionsregeln durch A-Priori-Prinzipien zu beweisen (vgl. Putnam, 1983). Freitas (2002) nennt die resultierende Voreingenommenheit einen *induktiven Bias*:

> An inductive bias can be defined as any criterion (explicit or implicit), other than strict consistency with the data, used to favor one hypothesis over another [...]. (Freitas, 2002, S. 36)

Diesen Bias sieht er aber wie Forster als notwendig für die Verwendung einer Induktionsregel an, d.h. es handelt sich dabei nicht um einen vermeidbaren Fehler oder den Hinweis auf eine zu ersetzende Induktionsregel (vgl. Abschnitt 2.1.3):

> The need for bias becomes clear once one realizes that, given a set of data instances (training set), the number of hypothesis (e.g., candidate rules) implying these instances is potentially infinite [...]. Hence, any algorithm that discovers prediction rules must have an inductive bias. Without inductive bias there is no induction [...], in the sense that the algorithm would be unable to prefer one hypothesis over many other hypotheses that are equally consistent with the data. (Freitas, 2002, S. 36)

Er unterscheidet weiterhin zwei Untergruppen des induktiven Bias: den *Repräsentations-Bias* und den *Präferenz-Bias* bzw. *Such-Bias*. Der Repräsentations-Bias ist mit der Wahl einer Repräsentationssprache verbunden, durch die vorbestimmt wird, welche Art von Wissen durch einen Algorithmus entdeckt werden kann. Enthält eine Sprache keine Möglichkeit der Selbstreferenz, scheiden bspw. dementsprechend Phänomene, die auf Selbstreferenz aufbauen, aus dem Raum der untersuchbaren Phänomene heraus. Ein Präferenz-Bias wird aus der Präferenz bestimmter Hypothesen gegenüber anderen möglichen Hypothesen gebildet, entweder in der Bewertung von Hypothesen oder der Generierung von Hypothesen. Im Allgemeinen ist es deshalb nicht möglich, von der prinzipiellen Überlegenheit eines Data-Mining-Algorithmus gegenüber einem anderen zu sprechen: Eine größere relative prädiktive Genauigkeit kann einem Algorithmus gegenüber einem Algorithmus nur in Bezug auf einen oder wenige Anwendungsbereiche zugesprochen werden oder für eine Menge von Datensätzen. In diesem Sinne erweisen sich Entscheidungsbäume als vorteilhaft für Datensätze, die im Bereich der Kreditwürdigkeitsüberprüfung gesammelt werden, nicht aber für Probleme, in denen eine parallele Klassifikation notwendig wird:

> [...] decision tree-building algorithms tend to do well on credit data sets. One of the reasons seems to be that this kind of data set is usually produced by a human-decision maker who classifies the data considering the value of one predictor attribute at a time. Since decision trees classify the data in basically the same way [...], the bias of decision tree-building algorithms is naturally suitable for this kind of data. (Freitas, 2002, S. 37)

Als zweites Beispiel für das *Frame*-Problem bei Heuristiken führt Forster das *Grue*-Konzept von Nelson Goodman an (Goodman, 1965), durch das die Abhängigkeit von Beobachtungen *und* Gesetzmäßigkeiten von sprachlichen Konventionen demonstriert wird. Die eigentliche Schwierigkeit von Entscheidungsregeln identifiziert Forster in dem notwendigen Trade-Off zwischen der Berücksichtigung allgemeiner Invarianten der Natur[4] - und einer notwendigen Anpassung an spezifische Kontexte:

> On the one hand, it is clear that some degree of specificity is required if rules are tailored to a particular context. On the other hand, the general regularities of nature have to be exploited if predictive accuracy is to be achieved. [...] I have argued for the importance of general principles, such as the Markov condition, stationarity, and spatial symmetry. The key is to exploit these generalities and to utilize information pertaining to a particular environment

[2] Der Zustand zum Zeitpunkt $t + 1$ hängt allein vom Zustand zum Zeitpunkt t ab.

[3] Die Vorhersage von i auf $i + 1$ ist immer gleich.

[4] Gemeint sind wohl eher Invarianten der Umgebung als Invarianten jener diffusen Entität, auf die mit dem Begriff „Natur“ verwiesen wird.

> at the same time. These two elements must work in concert, and this is clearly a highly complicated task even if the implementation of the decision rule is very simple at the surface. (Forster, 1999, S. 558)

Dabei wird es zu einem entscheidenden Problem, festzulegen, welches die generellen und welches die spezifischen Prinzipien sein sollen und wie damit auf eine Umgebung reagiert werden kann. Dabei erfordert eine Bestimmung des Erfolgs der Anpassung eine Brücke zwischen Reaktionen und Umgebung, die Forster unter dem Begriff *fit to reality* untersucht. Forster (1999) hinterfragt weiterhin das Kriterium, das eingesetzt werden soll, um einen *fit to reality* bestimmen zu können. Als möglichen Kontrast setzt er die Divergenz zwischen realistischen und instrumentalistischen Wissenschaftsauffassungen:

> Instrumentalists view scientific hypotheses as *merely* instruments for the control and prediction of nature. I say „merely“ because realists also agree that theories are instruments for prediction. However, realists think that science aspires to the more lofty goal of truth, which includes not only the truth of prediction, but also the truth of theoretical statements about the reality behind the observable phenomena. (Forster, 1999, S. 558 f.)

Damit bleibt es eine Aufgabe, für die ökologische Rationalität zu definieren, was als Anpassung verstanden werden soll. Zur Klärung dieser Frage gehören die Bestimmung der zeitlichen Perspektive, die Wahl statistischer Anpassungsmaße etc. Prinzipiell sieht Forster allerdings keinen Grund, das Programm deshalb zu diskreditieren:

> If ecological rationality is to be a precise notion, then the variety of ways in which „fit to reality“ is defined might be an embarassment. However, I think that any of these kinds of fit would form a valid explication of the idea. (Forster, 1999, S. 559)

Man könnte einwenden, dass auch Kriterien kontextspezifisch gewählt werden müssen, allerdings würde ohne Bestimmung der Prinzipien, nach denen diese Wahl funktionieren könnte, der Ansatz ähnlich kritisierbar wie das Einzeichnen einer Zielscheibe nach dem Pfeilschuss. In der prinzipiellen Entscheidung zwischen realistischen und instrumentalistischen Kriterien entscheidet sich Forster aber zu Gunsten eines Maßes der prädiktiven Genauigkeit:

> ... the reasons why predictive accuracy is a more interesting alternative to empirical adequacy is that (1) in cases where empirical adequacy is achievable the achievement is equally well described in terms of predictive accuracy, and (2) there are many situations in which empirical adequacy is not achieved (it is a black-and-white notion), but at least some *degree* of predictive accuracy is achieved. (Forster, 1999, S. 559)

Auch die Bereichsspezifität von Heuristiken trägt für Forster zur Vereinfachung ihrer Aufgabe auf Kosten einer kritischen Vorleistung bei, da die richtige Vorauswahl von Handlungen, die durch einfache Entscheidungsregeln gesteuert werden, für ihre verlässliche Operation und ökologische Validität verantwortlich ist. Dementsprechend ist es für Forster essenziell von Bedeutung, dass für Heuristiken nicht nur ökologische Adaptivität demonstriert wird, sondern auch zum einen erklärt werden kann, *warum* Heuristiken adaptiv sind, und zum anderen, wie sie evolviert oder gelernt werden konnten. Das Konservierungsgesetz der Komplexität trifft für Forster auch auf einfache Heuristiken zu. Ähnlich wie es als schwierig einzustufen ist, eine Benutzerschnittstelle zu programmieren, die die einfache und fehlerfreie Bedienung einer Software ermöglicht, ist es für Forster ebenso komplex, Heuristiken zu konstruieren und Anwendungsregeln für sie zu konzipieren, damit die Heuristiken *fast and frugal* sein können:

> [Gigerenzer's & Todd's] „apparently paradoxical thesis: *Higher order cognitive mechanisms can often be modeled by simpler algorithms than can lower order mechanisms*“ [...] seems to me exactly as one would expect. (Forster, 1999, S. 562)

Damit kann der Gedanke der Komplexitätserhaltung auch genutzt werden, um Lösungen für das *Frame*-Problem im ABC-Ansatz zu überprüfen.

7.1.3 Der Mechaniker in der Werkzeugkiste

Für Simon (1978) sind die Probleme der realen Welt so enorm, dass selbst Computerprogramme - diese Betonung setzt Simon - nur als grobe Annäherungen verstanden werden können und Hoffnung die angemessene Haltung ist, wenn es um die Beurteilung der Angemessenheit von Modellen geht. Trotzdem sind Menschen trotz geringer komputationaler Fähigkeiten oft in der Lage, gerade jene Probleme zu lösen, die Rechnern große Probleme bereiten, also Probleme mit schlecht definierten Zielen oder Problemräumen und Probleme ohne eine reguläre mathematische Struktur. Dementsprechend geht Simon davon aus, dass Menschen Heuristiken verwenden, um diese Probleme zu lösen:

> People solve such problems by the shrewd use of heuristics and at the expense of giving up guaranteed completeness of search and optimality of the solutions attained. (Simon, 1995, S. 98)

Für Goodie, Ortmann, Nerissa Davis, Bullock und Werner (1999) führte die anfängliche Begeisterung für Heuristiken als Alternative zu missglückenden Versuchen der Modellierung von unbegrenzter Rationalität allerdings zu einem neuen Problem:

> [...] the initial success of the heuristics approach to creating artificial intelligence has led to the tacit assumption that certain heuristics are candidates for nearly universal application. [...] This pespective views the best heuristics as those that are able to deal with the widest range of problems. In this respect, AI treats heuristics as general-purpose demons. (Goodie et al., 1999, S. 328)

Die Autoren sehen das *Frame*-Problem als Motivation für die Forderung nach Bereichsspezifität von Heuristiken (vgl. Ketelaar & Todd, 2001), da auch schnelle und frugale Heuristiken in einem schlecht gewählten Anwendungsbereich aufgrund fehlender ökologischer Validität versagen müssen. Damit wird gleichzeitig deutlich, dass die Bereichsspezifität von Heuristiken nicht risikofrei ist, eine fehlerhafte Auswahl einer Heuristik in unpassender Umgebung kann sogar die Möglichkeit akzeptabler Resultate verhindern. Es ist eine wesentliche Fehlerquelle in dem Entscheidungsschritt verborgen, die richtige Heuristik für eine Umgebung auszuwählen: Im Falle einer wirklichen Werkzeugkiste wählt der Mechaniker das Werkzeug für einen bestimmten Zweck aus, in der *Adaptive Toolbox* muss sich der Mechaniker allerdings *in* der Werkzeugkiste befinden.

Es wäre ein idealer Zustand, wenn Heuristiken sich förmlich selbst auswählen könnten, wenn etwas in der Umgebung die Aktivierung genau eines Werkzeugs veranlasst. Eine solche Bestimmung von Umgebungsklassen ist allerdings äußerst problematisch, und für die Diskriminierung von Umgebungsarten sind Prozesse notwendig, die nicht in gleicher Weise spezifisch sein können wie die Heuristiken. Es muss eine Stelle geben, an der Informationen integriert werden, damit nicht gleichzeitig mehrere Werkzeuge zum Einsatz kommen oder kein Werkzeug ausgewählt wird. Für die Auswahl von Heuristiken stehen prinzipiell eine ganze Reihe von Mechanismen zur Verfügung, von denen hier eine Auswahl genannt werden kann:

- Bewusste Entscheidung durch einen allgemeinen Prozess (Meta-Kognition): Heuristiken werden nach Überlegung bewusst ausgewählt und eingesetzt.
- Lernen und Erfahrung: Aufgrund von Erfahrungen mit spezifischen Umgebungen und spezifischen Heuristiken wird eine Heuristik ausgewählt, die der Umgebung angepasst ist.
- Meta-Heuristiken: Eine Heuristik für die Entscheidung zwischen Heuristiken wird ausgewählt, die selbst keiner Entscheidung bedarf.
- Vordefinierte Auslöser: Die Heuristik reagiert auf bestimmte Schlüsselreize der Umgebung in einer Art Schlüssel-Schloss-Prinzip.
- Vorbereitete Umgebungen: Umgebungen werden automatisch und unwillkürlich so wahrgenommen, dass die Anwendung einer Heuristik dem Anwender natürlich erscheint.
- Angeeignete Umgebungen: Der Anwender lernt aufgrund von Erfahrungen, Umwelten so wahrzunehmen, dass die Anwendung einer Heuristik natürlich erscheint.

- Auflösung der Interface-Idee: Im Sinne von Dreyfus (2007) wird auf das Konzept der Umweltrepräsentation verzichtet, die Heuristiken ergeben sich aus der Interaktion mit der Welt.
- Konsequente Auflösung der Interface-Idee: Die Heuristiken selbst befinden sich teilweise nicht im Gehirn, sondern in der Welt selbst.

Für Cooper (2000) gefährdet das Fehlen einer Spezifizierung der Entscheidungsmechanismen die Integrität des Programms, da die Falsifizierbarkeit des Ansatzes gefährdet ist. Einhorn und Hogarth (1981) schlägt Strategien der Entscheidung auf mehreren Ebenen vor, wenn nämlich Strategien selbst als multidimensionale Objekte in einem Entscheidungsraum konzipiert werden, wobei die Qualitätseigenschaften der Strategien diesen Raum aufspannen könnten. In dieser Weise könnte bspw. *elimination by aspects* (Tversky, 1972) eingesetzt werden, um Strategien aus der Auswahl zu eliminieren. Einhorn warnt aber davor, dass die Auswahl von Metastrategien zu einer Schleife ohne Ausgang führen könnte. Auch für Feeney (2000) droht ein infiniter Regress, wenn Entscheidungen über das Entscheiden getroffen werden müssen, weshalb sie die Notwendigkeit meta-kognitiver Prinzipien für den Ansatz fordern; ähnlich argumentiert Fuller (2000). Für Luce (2000) sollte zudem die Rolle der Evolution in der Auswahl von Heuristiken überprüft werden, während für Morton (2000) eine Auswahlregel nicht in exakter Parallele zu den Heuristiken erster Ordnung stehen kann, obwohl er einräumt:

> [...] some features of first order heuristics are very attractive as attributes of meta-choice, in particular the use of a simple stopping rule and the determination of some basic parameters by the environment. (Morton, 2000, S. 759)

Wird zudem die Informationsaufnahme berücksichtigt, die für die Entscheidung zwischen Heuristiken notwendig erscheint, kann ebenfalls eine Instanz des Komplexitätserhaltungssatzes attestiert werden:

> ... meta-level decisions will typically require some prior knowledge about the structure of the environment (e.g., whether it is non-compensatory, J-shaped etc.), which may add substantially to the overall processing cost of a fast and frugal model. (Shanks & Lagnado, 2000, S. 762)

Dem Vorwurf von Cooper (2000), dass vielleicht nicht die optimale Heuristik ausgewählt wird, begegnen Todd et al. (2000) mit dem Hinweis, dass eine Optimierung auch nicht auf der Meta-Ebene angestrebt wird. In Bezugnahme auf Dawes wird zudem ein flaches Maximum postuliert, als dass verschiedene Heuristiken ähnlich gut abschneiden können[5]. Es wird betont, dass ein zu langes Nachdenken über Entscheidungen mit Kosten verbunden sein kann. Todd et al. (2000) tendieren zu einer Lösung des Entscheidungsproblems entweder durch Meta-Heuristiken oder Prozeduren, die gerade keine Intelligenz benötigen. Die Rolle der Meta-Kognition (vgl. Cox, 2005) in Bezug auf das Entscheidungsproblem zwischen Strategien - dieser Text geht der adaptiven Werkzeugkiste voraus - wird hingegen bspw. bei Payne, Bettman und Johnson (1993) betont:

> Our hypothesis is that the contingent use of strategies (heuristics) represents an intelligent response to decision problems by an individual willing to trade off accuracy and effort. That is, we propose that observed decision processes often reflect a reasonable compromise between the desire to make a good decision and the desire to minimize the cognitive resources used in making the decision. In short, we believe that individuals are adaptive decision makers. (Payne et al., 1993, S. xi f.)

Over (2000) vermisst diese Konzeption einer intelligenten Meta-Kognition im ABC-Ansatz. Er betrachtet etwa *Fast and Frugal Trees* als für sich genommen unzureichend für Entscheidungen in Krankenhäusern:

> This process looks nicely fast and frugal, but of course its actual use in a hospital would not always be so simple. The medical people applying it would be expected to use their expert judgement sometimes in default reasoning. An unlimited number and variety of possible

[5] Allerdings wird diese These schlechter unterstützt als die auf Regressionsgewichte bezogene These von Dawes.

> complications might mean that a low-risk patient, according to the heuristic, should be reclassified as a high-risk one. Even Gigerenzer and Todd describe the object of the heuristic as deciding which treatment a patient „should“ have, and deontic conclusions like this are very much open to withdrawal in default or defeasible reasoning. But the book does not study any higher-level reasoning, whether deontic or any other, which can override heuristics in novel or unusual cases. (Over, 2000, S. 187)

Letztendlich weist Over damit auf die Unmöglichkeit der Bestimmung von Relevanz von Ereignissen für den Entscheidungsprozess durch algorithmische Heuristiken hin, das *Frame*-Problem kann durch Bereichsspezifität eben dann nicht gelöst werden, wenn Bereiche sich nicht eindeutig trennen lassen und potenziell erneut alles mit allem in Berührung treten kann.

Im Folgenden soll ein Versuch dargestellt werden, die Auswahl von Strategien über einen Mechanismus des Verstärkungslernens zu modellieren, den Rieskamp und Otto (2006) zur Lösung des von den Autoren so benannten *Strategy Selection Problem* vorstellen. Mechanismen der Nutzenmaximierung unter Berücksichtigung von Strategiekosten und Strategieergebnissen werden abgelehnt, und bei der Verwendung von Meta-Strategien wird vor Rekursionen gewarnt.

> [...] applying a meta-strategy to strategy selection could run into a recursive homunculi problem of deciding how to decide. (Rieskamp & Otto, 2006, S. 207)

Als Alternative schlagen sie ein Modell vor, das sie als Theorie des *Strategy Selection Learning (SSL)* bezeichnen, das von Menschen annimmt, dass sie ein Repertoire von Strategien für übliche Inferenzprobleme besitzen. Über ein Verstärkungsprinzip, das die Leistungen von Strategien in der Vergangenheit berücksichtigt, werden nach der Theorie in einer Situation passende Einzelstrategien von Menschen mit höherer Wahrscheinlichkeit ausgewählt als unpassende. Die Annahmen der SSL lassen sich in folgender Weise formulieren: Individuen besitzen eine Menge S (Repertoire) aus N kognitiven Strategien. Die Präferenzen sind für jedes Individuum für jede Strategie i als positive Erwartungen $q(i)$ definiert, deren Wert für jeden Zeitpunkt t als $q_t(i)$ bezeichnet werden soll. Es werden diskrete Zeitintervalle angenommen (zum Beispiel Trials) und die Wahrscheinlichkeit $p_t(i)$, dass zu einem Zeitpunkt t Strategie i gewählt wird, wird bestimmt als:

$$p_t(i) = \frac{q_t(i)}{\sum_{j=1}^{N} q_t(j)}. \tag{7.1}$$

Diese Konzeption ist nahezu identisch mit dem Prinzip der *Roulette-Wheel*-Selektion, das für genetische Algorithmen diskutiert wurde. Die anfänglichen Erwartungen $q_i(i)$ werden als Produkt von maximaler Auszahlung durch eine korrekte Entscheidung (r_{max}), einem anfänglichen Assoziationsparameter ($w > 0$) und einem anfänglichen Präferenzparameter ($0 < \beta_i < 1$ mit $\sum_{i=1}^{N} \beta_i = 1$, so dass N-1 Freiheitsgrade bestehen) bestimmt, also als

$$q_t(i) = r_{max} \cdot w \cdot \beta_i. \tag{7.2}$$

$S(t)$ soll die Menge der Indizes der Strategien bezeichnen, die zum Zeitpunkt t ausgewählt worden sein können. Das heißt, es gilt $i \in S(t)$ genau dann, wenn zum Zeitpunkt t sowohl die für die Anwendung einer Strategie notwendigen Informationen dem Individuum bekannt sind als auch seine Entscheidung der durch die Strategie angezeigten Entscheidung entspricht. Die Auszahlung einer Strategie zum Zeitpunkt t sei $r_t(i)$. Nach jedem Zeitintervall werden die Erwartungen der Strategien abhängig von Auswahl und Erfolg angepasst in der Form

$$q_{t+1}(i) = \begin{cases} q_t(i) & i \notin S(t) \\ max(\rho, q_t(i) + r_t(i)) & i \in S(t) \wedge (\forall j)(j \in S(t) \rightarrow j = i) \\ max(\rho, q_t(i) + p_t(i) \cdot r_t(i)) & i \in S(t) \wedge (\exists j)(j \in S(t) \wedge j \neq i). \end{cases} \tag{7.3}$$

Dabei stellt ρ den minimalen Erwartungswert dar - Rieskamp und Otto (2006) schlagen $\rho = 0,0001$ vor. Zusätzlich wird noch die Möglichkeit einer (technischen) Fehlentscheidung mit der Gesamtwahrscheinlichkeit ϵ eingeräumt, bei gleicher Wahrscheinlichkeit für jede Fehlermöglichkeit, bei insgesamt k Alternativen also eine Fehlerwahrscheinlichkeit von $\frac{\epsilon}{k-1}$ für jede der $k-1$ falschen Alternativen. Falsch bezieht sich

dabei allerdings auf die technische Anwendung der Strategien, also auf einen Rechen- oder Analysefehler. Bei perfekter Anwendung würde sich die Wahrscheinlichkeit für die Auswahl einer Alternative a aus der Alternativenmenge mit k Alternativen durch Strategie i als $p(a|i)$ ergeben[6], die Wahl einer beliebigen anderen Alternative als $p(\neg a|i)$ mit $p(a|i) + p(\neg a|i) = 1$. Rieskamp und Otto (2006) schlagen vor, dass bei der Vorgabe einer Fehlermöglichkeit sich die Wahrscheinlichkeit $p(a|i, \epsilon)$ als

$$p(a|i, \epsilon) = (1 - \epsilon) \cdot p(a|i) + \frac{\epsilon}{k - 1} \cdot p(\neg a|i) \tag{7.4}$$

ergeben soll[7]. Damit ergibt sich schließlich $p_t(a)$ als

$$p_t(a) = \sum_{i=1}^{N} p_t(i) \cdot p(a|i, \epsilon). \tag{7.5}$$

Die Analyse eines experimentellen Inferenzproblems zeigt in der Sicht der Autoren eine gute Vorhersageleistung der SSL-Theorie, und auch Rieskamp (2006) bewertet die Theorie positiv im Vergleich mit alternativen Erklärungsansätzen. Rieskamp und Otto (2006) resümieren:

> People appear to select their strategies adaptively, such that strategies that perform well become more likely to be selected. SSL provides a computational theory that describes how this strategy selection process could take place. By following the traditional roots of psychology in learning, the strategy selection problem receives a promising solution. (S. 232)

Die Anbindung an die behavioristische Lerntheorie kann allerdings auch als Erblast des Ansatzes bewertet werden, je nachdem, wie die Ansprüche des Modells auf Beschreibungsgenauigkeit bewertet werden. Die zuletzt zitierte Passage suggeriert, dass nicht nur die Ergebnisse des menschlichen Lernprozesses, sondern der Lernprozess selbst modelliert werden soll. Eine Reduktion von Strategieselektion auf das skizzierte Verhalten kann aber nicht in allen Entscheidungssituationen als plausibel betrachtet werden. Auch wenn es genauso wenig plausibel erscheint, dass Menschen durch Strategieentscheidung ihren Nutzen maximieren, ist es nicht ausgeschlossen, dass einige Menschen eine Maximierung versuchen bzw. dass manche Menschen versuchen, zumindest nach einem individuell gewählten Kriterium zu entscheiden. Die Berücksichtigung von Sonderfällen stellt zudem auch für den SSL-Ansatz ein Problem dar. Zur Bestimmung, ob eine Meta-Wahl in den Anwendungsbereich der SSL-Theorie fällt, wäre also eine Meta-Meta-Entscheidung zu treffen, und der rekursive Graben klafft erneut auf.

Es muss resümiert werden, dass das Problem der Auswahl von Werkzeugen aus der Adaptiven Werkzeugkiste zu den virulenten Problemen des Ansatzes gehört, dessen besonderes Gewicht sich aus der Nähe zum *Frame*-Problem und Schwierigkeiten der Multimodularitätskonzeption(vgl. Fodor, 2000; Woodward & Cowie, 2004) ergibt. Aus der Hypothese Forsters, dass Komplexität erhalten bleiben muss, ist dies allerdings ein zu erwartendes und damit einfaches Ergebnis.

7.2 Ein Rückblick auf das Ausgangsproblem

7.2.1 Ein Modellierungsvorschlag für die Regelinduktion

An dieser Stelle soll eine Überführung der Ausgangsproblemstellung dieser Arbeit in eine Modellsprache genutzt werden, um visualisieren zu können, wie die Aufgabenstellung von angrenzenden Fragen separiert

[6] Die konditionale Wahrscheinlichkeit für Wahlen bei gegebener Anwendung einer Strategie sollte unabhängig vom Zeitpunkt sein, deswegen wird hier $p_t(a|i) = p(a|i)$ angenommen und auf den Index von Rieskamp und Otto (2006) verzichtet.

[7] Allerdings wird nicht deutlich, wie das Vorliegen der Anwendungsbedingungen für eine Strategie geprüft werden kann, wenn diese die Identität von individueller Wahl und Strategiewahl voraussetzen. Angenommen, eine Strategie sei deterministisch und ein Proband entscheidet sich gegen die Voraussage einer Strategie, obwohl er die Informationen für eine Strategie gesammelt hat. Wird diese Strategie nun trotzdem verstärkt, da der Proband einen technischen Anwendungsfehler der Strategie hätte begehen können und umgekehrt, muss die Verstärkung bei Übereinstimmung nun mit einem Faktor multipliziert werden, da die Übereinstimmung zufällig entstanden sein könnte? Prinzipiell kann jede Strategie in diesem Modell jede mögliche Alternative wählen, dementsprechend ist das Kriterium der Übereinstimmung nicht mehr eindeutig definiert:

> It is assumed that a strategy was selected if the necessary information for applying the strategy was acquired and the choice coincides with the strategy's prediction. (Rieskamp & Otto, 2006, S. 209)

werden muss, die im folgenden Abschnitt umrissen werden sollen. In dieser Weise kann deutlich werden, durch welche Vorannahmen die Leichtigkeit der Problemformulierung erkauft worden ist. Das Regelinduktionsmodell für die Klassifikation ist angelehnt an die Modellierung der Veränderung von Weltzuständen von Holland et al. (1986), die sich der Sprache von Morphismen bedient[8].

7.2.1.1 Modellierung von Klassifikation als q-Morphismus

Im vorgestellten Modell sollen Weltzustände mit Objekten oder Personen gleichgesetzt werden, die es zu einem bestimmten Zeitpunkt zu klassifizieren gilt. Klassifikationsprozeduren sind auf Vereinfachungen angewiesen, um praktikabel zu sein, es kann kein Regelsystem erstellt werden, das jede erdenkliche Unterschiedlichkeit zwischen Objekten berücksichtigen kann. Analog schreibt Langton (1989a) über Systemregeln:

> Systems that attempt to supply *global* rules for *global* behavior simply *cannot* povide a different rule for every global state. Thus, the global states must be classified in some manner; categorized using a coarse-grained scheme according to which the global states within a category are indistinguishable. The rules of the system can only be applied at the level of the resolution of these categories. (S. 42)

Morphismen sind strukturverträgliche Abbildungen von Mengen auf Mengen. Würde jeder Zustand der Welt auf einen spezifischen Modellzustand eineindeutig abgebildet, wobei sich die Struktur der Welt in der Struktur des Modells wiederfinden würde, so wäre das Modell ein Isomorphismus. Dazu schreiben Holland et al. (1986), dass aufgrund der Komplexität realistischer Umgebungen und der Begrenztheit kognitiver Systeme mentale Modelle nicht als Isomorphismen beschrieben werden können, bei denen jeder Weltzustand auf einen einzigartigen Modellzustand abgebildet wird. Als Alternative schlagen die Autoren die Verwendung eines homomorphen Modells vor: Mengen von Umweltzuständen werden hier zu Kategorien gebündelt, wobei alle Mitglieder einer Kategorie auf den gleichen Modellzustand abgebildet werden:

> A faithful model based on categories, in which the mapping from elements of the world to elements of the mental model is many-to-one, is called a *homomorphism*. (Holland et al., 1986, S. 31)

Holland et al. (1986) beschreiben die Induktion von Äquivalenzklassen aus den Weltzuständen als einen Schritt, in dem einfache *Detektoren* Eigenschaften der Ausgangszustände enkodieren, so dass jedes Objekt ein Detektorprofil erhalten kann[9]. Objekte mit gleichem Profil werden auf den gleichen Modellzustand abgebildet.

Bei der Übertragung auf das allgemeine Problem der Klassifikation werden folgende Veränderungen vorgenommen. Die Detektoren werden konzipiert als Prüfverfahren, die bestimmte Eigenschaften der Untersuchungseinheiten messen. Werden Objekten oder Personen identische Messwerte zugeteilt, so sind sie für das Modell ununterscheidbar. Im Beispiel (s. Abbildung 7.1) werden elf Personen mit zwei Testverfahren untersucht, so dass mit der Abbildung M_1 Personenkategorien auf Testprofile abgebildet werden[10]. Die Messung auf einer Skala stellt ihrerseits eine homomorphe Abbildung eines empirischen Relativs auf ein numerisches dar [11]. Zwei Personen bilden jeweils eine eigene Kategorie für sich.

[8] Eine präzise Behandlung leistet bspw. Pfanzagl (1968)

[9] Holland et al. (1986) berücksichtigen zunächst nur binäre Detektoren wie „klein (ja/nein)“, hier soll ein weiterer Begriff verwendet werden.

[10] Die Abbildung einer Menge $\mathcal{A}$ auf die Menge $\mathcal{B}$ ist eine linkstotale, rechtseindeutige Relation $m \subseteq \mathcal{A} \times \mathcal{B}$, so dass jedem Element a der Menge $\mathcal{A}$ ein Element $m(a) = b$ der Menge $\mathcal{B}$ zugeordnet wird, also $m : \mathcal{A} \rightarrow \mathcal{B}$ (Reinhardt & Soeder, 1998).

[11] Ein empirisches System oder Relativ kann als eine Menge von Objekten $\mathcal{A}$ zusammen mit einer Menge von Relationen definiert werden, die in Bezug auf diese Menge definiert sind, also als $E = \langle \mathcal{A}, Q_1, Q_2, \ldots, Q_k \rangle$. Für eine Messung ist zunächst formal ein numerisches Relativ vom gleichen Typ erforderlich, d.h. ein Relativ, das aus einer Menge an Symbolen oder Zahlen besteht sowie aus einer Menge von Relationen, die in Anzahl und jeweils hinsichtlich ihrer Stellen den Relationen im empirischen Relativ entsprechen (z.B. $N = \langle \mathbb{R}, R_1, R_2, \ldots, R_k \rangle$). Von einer homomorphen Abbildung des empirischen Relativs auf das numerische - also von einer Messung - kann genau dann gesprochen werden, wenn jede Relation Q_i für eine geordnete Menge von Objekten genau dann erfüllt ist, wenn die Relation R_i für die geordnete Menge der Abbildungswerte dieser Objekte erfüllt ist (Gigerenzer, 1981). Je nach Relationen, die im empirischen Relativ gelten und die zudem im numerischen Relativ berücksichtigt werden, kann der Skala ein Niveau zugeordnet werden, auch wenn Benennung und Stellenwert dieser Skalenniveaus umstritten sind (vgl. Michell, 1999)

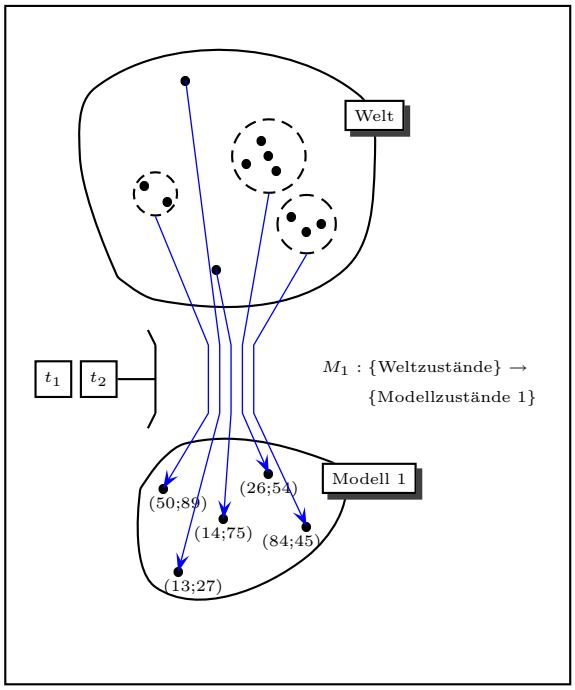


Abbildung 7.1: Kategorisierung im Modell: Durch die Abbildung M_1 werden Kategorien von Weltzuständen bzw. Personen aufgrund ihres durch die Detektoren ermittelten Profils auf Modellzustände abgebildet.

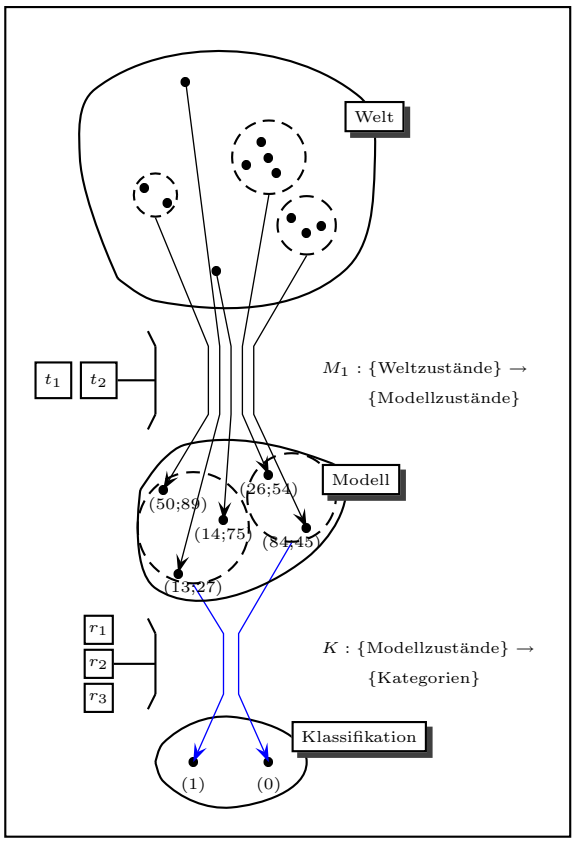


Abbildung 7.2: Klassifikation im Modell: Durch die Abbildung K werden Kategorien auf Klassifikationen abgebildet. Eine Liste von Entscheidungsregeln bestimmt über die Zuweisung.

Das Modell muss für die Klassifikationsaufgabe um eine weitere Komponente ergänzt werden. Im Allgemeinen gibt es weniger Klassifikationskategorien als Testprofile. Deshalb ist eine zweite Reduktion der Kategorien erforderlich, die durch das hierarchische Regelsystem übernommen wird. Auch dieser Schritt lässt sich als Homomorphismus auffassen, wie in Abbildung 7.2 gezeigt. Durch K werden Modellzustände auf Klassifikationskategorien abgebildet. Durch die Verkettung der beiden Abbildungen, also $M_1 \circ K$ wird jeder Beobachtungseinheit damit eine Klassifikationskategorie zugeordnet. Für eine Person p im Beispiel stellt $k = K(M_1(p))$ damit die zugeordnete Kategorie dar.

In der Modellierung von Holland et al. (1986) steht die Vorhersage der Weltzustände im Folgeschritt im Vordergrund. Für diesen Schritt wird mit Hilfe der gleichen Abbildung M_1 der neue Weltzustand auf den neuen Modellzustand abgebildet. Im Falle der Klassifikation steht allerdings nicht die Vorhersage der Zukunft im Vordergrund, sondern die Korrektheit der Klassifikation bezüglich eines Kriteriums. Allerdings ist das Kriterium selbst eine Kategorisierung, die nicht auf die gleiche Art zustande kommt wie die Klassifikation. In das Modell werden deshalb zwei weitere Abbildungen eingeführt.

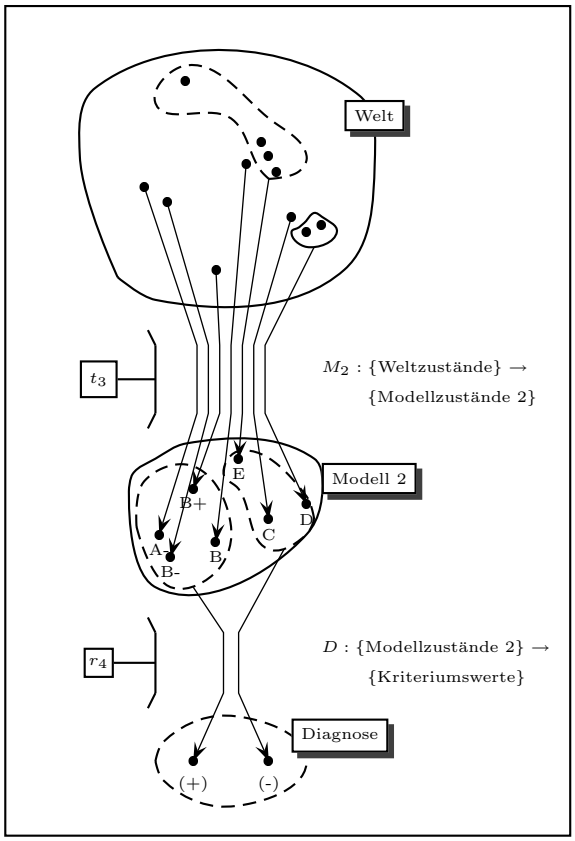


Abbildung 7.3: Diagnose im Modell: Zum Zeipunkt der Kriterienerhebung werden durch die Abbildung M_2 Weltzustände auf Beschaffenheitkategorien abgebildet, die ihrerseits durch die Abbildung D auf Kriteriumswerte abgebildet werden.

In einem ersten Schritt werden die Untersuchungseinheiten bewertet hinsichtlich einer oder mehrerer Eigenschaften. Dies entspricht einer zweiten Messung, die üblicherweise im geplanten Entscheidungsfall nicht möglich ist, nicht vor der Klassifikation möglich ist oder zu kostspielig erscheint. Um Entscheidungsregeln zu erstellen, sind Kriteriumsinformationen notwendig. Im Beispiel soll zunächst in Form einer Benotung die zur Debatte stehende Leistung gemessen werden [12]. Aufgrund dieser Messung wird durch Abbildung M_2 jeder Untersuchungseinheit ein Modellzustand in diesem zweiten Modell zugeordnet; auch hier gibt es Kategorien, so dass Gruppen von Personen die gleiche Note erhalten. In einem zweiten Schritt wird jedem Modellzustand ein Kriteriumswert im Sinne einer Diagnose zugeordnet[13]. Im Beispiel wird die simple Entscheidungsregel angewandt: Ordne einen Modellzustand, der mit einer besseren Note als $C+$ verbunden ist, einen positiven Kriteriumswert zu und jedem anderen Modellzustand einen nega-

[12] Dabei kann hier in anderen Fällen durchaus auch eine mehrdimensionale Messung erfolgen. Unter Umständen ist in diesen Fällen eine komplexer strukturierte Diagnose notwendig als im Beispiel.

[13] Im Folgenden wird von einem Problem mit binärem Kriterium ausgegangen. Eine Erweiterung auf mehr als zwei Kriteriumskategorien und dementsprechend mehr als zwei Vorhersagekategorien ist möglich.

tiven Kriteriumswert. Diese zweite Abbildung D bildet also Modellzustände auf Diagnosen bestimmter Kriteriumswerte ab, so dass durch $M_2 \circ D$ jeder Beobachtungseinheit eine Kriteriumskategorie zugeordnet wird, so dass für eine Person p im Beispiel $d = D(M_2(p))$ die zugehörige Diagnose darstellt.

Zu einem sinnvollen Modell wird die Formalisierung von Holland et al. (1986) aber erst dann, wenn nicht nur Zustände berücksichtigt werden, sondern auch die Transitionen zwischen Zuständen vorhergesagt werden können. In ihrem Ansatz verwenden sie deshalb eine Transitionsfunktion T, die einen Weltzustand zum Zeitpunkt t auf den dazugehörigen Weltzustand zum Zeitpunkt $t + 1$ abbildet. Analog wird eine Transitionsfunktion T' eingeführt, die die Modellzustände aufeinander bezieht:

> The next step in describing a mental model is to introduce a *model* transition function T', which is intended to mimic the transition function T operating in the world. The function T' of the model will then describe the manner in which *categories* of environmental states, coupled with *categories* of actions, lead to *categories* of subsequent states. (Holland et al., 1986, S. 32)

Bei der Anpassung an das Klassifikationsmodell müssen einige Veränderungen vorgenommen werden:

1. Die Beziehung zwischen Vorhersagen und Diagnosen ist normativ vorgegeben: Es ist das Ziel des Modells, dass eine festgelegte Transition in dem Sinne gilt, dass positive Vorhersagen auf positive Diagnosen und negative Vorhersagen auf negative Diagnosen abgebildet werden können.

2. Zwischen Testprofilen und Leistungsprofilen muss kein Zusammenhang modelliert werden, der interessierende Zusammenhang ist durch die Abbildung von Vorhersagen auf Diagnosen ($\hat{I}$) repräsentiert.

3. Für den Fall, dass Testwerte und Leistungswerte zum gleichen Zeitpunkt erhoben werden, ist die Transitionsfunktion (I) die Abbildung der Weltzustände auf sich selbst, da beide Messungen bei gleichem Personenzustand vorgenommen werden. Gibt es ein zeitliches Intervall zwischen den Messungen, dann entspricht I der üblichen Transitionsfunktion.

Die Verbindung der beiden Teilabbildungen durch die Transitionsfunktionen zeigt Abbildung 7.4. Das Modell ist in der vorliegenden Form allerdings irreführend. Holland et al. (1986) nennen ein Modell genau dann valide, wenn es einen Homomorphismus darstellt und im Diagramm Kommutatitvität im folgenden Sinne besteht:

> [...] carrying out a transition in the environment and then determining the equivalence class of the resulting state will have the same outcome as determining the equivalence class of the intial environmental state and then carrying out the transition in the model. (S. 33)

Übersetzt in die hier gewählte Modellierung entspricht dies der Forderung:

$$M_1 \circ K \circ \hat{I} = I \circ M_2 \circ D \tag{7.6}$$

bzw.

$$\hat{I}(K(M_1(p))) = D(M_2(I(p))). \tag{7.7}$$

Hier stellt sich im Allgemeinen aber das Problem ein, dass ein vollständig valides Modell nicht erstellt werden kann, es wird Ausnahmen geben, für die die Kommutativität nicht erfüllt ist. Holland et al. (1986) schlagen deshalb die Lösung vor, eine zweite Ebene in das Modell einzuführen. Beispielsweise könnte ein Detektor genutzt werden, um Ausnahmen zu identifizieren, für die die üblichen Abbildungsregeln außer Kraft gesetzt werden. Die Autoren nennen das modifizierte Modell einen *Quasi-Homomorphismus* oder kürzer *q-Morphismus*:

> In a q-morphism a higher level in the model, with its broader categories, provides default expectations [...] that will be used to make predictions unless some exceptional category is signaled. An exception evokes a lower level of the model, at which a different model transition function is specified to capture the exception. (S. 34)

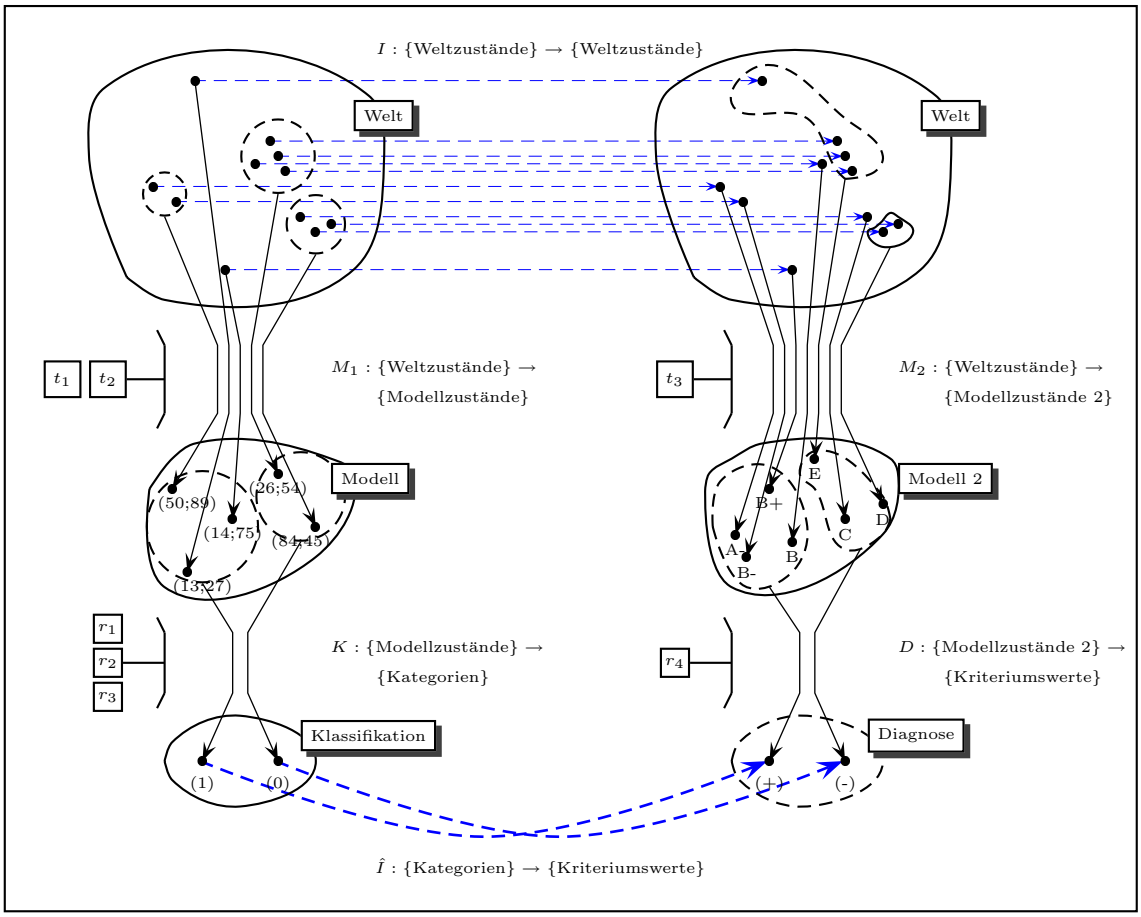


Abbildung 7.4: Das Ausgangsproblem als Homomorphismus: Die beiden Modellteile sind durch die Transitionsfunktionen I und $\hat{I}$ verbunden.

Dementsprechend wird die Modellkonstruktion als progressive Verfeinerung von q-Morphismen verstanden[14]. Versagt das Modell für spezifische Fälle, so stehen dem Modellierenden demnach drei Optionen zur Verfügung:

1. Das Modell wird als ungeeignet für den Bereich eingeschätzt, und unter Umständen sind neue Kategorien oder neue Detektoren erforderlich.

2. Die Zustände, bei denen Fehler gemacht werden, sind Ausnahmen: Ausnahmen werden durch spezifischere Klassen und eine spezifische Transitionsfunktion behandelt.

3. Die Fehler werden als unsystematisch eingeschätzt, auch wenn Vertrauen in das Modell besteht. In diesem Sinne wird die Transitionsfunktion nicht mehr deterministisch, sondern probabilistisch verstanden.

Auch diese drei Reaktionen lassen sich auf die hier vorgenommene Modellierung übertragen. Scheint es unmöglich, aufgrund der eingeholten Information die Kriteriumsklasse vorherzusagen, scheinen ungeeignete Prädiktoren vorzuliegen, und eine Revision der Annahmen scheint angezeigt. Gibt es Sonderfälle, kann das Modell entweder für Klassen von Sonderfällen sensibilisiert werden, für die Sonderbehandlungen eingeführt werden. Sind die Probleme des Modells unsystematisch, kann dies ebenfalls durch ein probabilistisches Verständnis von Transitionsfunktionen, im psychologischen Fall auch von Messfunktionen berücksichtigt werden. In der Neubildung und Ausdifferenzierung von Kategorien kann das Modell dementsprechend weiterentwickelt werden:

> The success of a q-morphic model is fundamentally limited by the categories available for encoding the environment. [...] The categories in the model are in turn dependent on the detectors that encode properties of the world. [...] As new properties are encoded and new categories become available, the predictive power of the model has the potential to increase. (Holland et al., 1986, S. 38)

Der Weg, der hier beschritten werden soll, ist die Einführung zweier Ausnahmekategorien, eine für falsch positive und eine für falsch negative Klassifikationen. Hierzu werden zwei Funktionen A_i und A_k eingeführt. Durch die erste werden Personen als Ausnahmen identifiziert und auf die zutreffende Fehlerkategorie abgebildet, durch die zweite werden Fehlerkategorien auf die korrekte Klassifikationskategorie abgebildet[15].

Abbildung 7.5 zeigt diese Lösung für das Beispiel. Im ursprünglichen Diagramm in Abbildung 7.4 ist für zwei Fälle die Kommutativität nicht erfüllt. Die neuen Abbildungen haben Priorität vor den alten Abbildungen, so dass mit ihrer Einführung das Modell einen q-Morphismus darstellt. Einige Erläuterungen sind angebracht, um einige Unterscheidungen zu klären: Die Art und Weise, wie Ausnahmen hier identifiziert werden, folgt nicht dem Prinzip der Identifikation durch Detektoren. Statt dessen werden alle Fälle als Ausnahmen betrachtet, für die die Vorhersage nicht erfolgreich ist. Es soll hier auch kein kognitives deskriptives Modell erstellt werden, das in seiner Genese einen psychologischen Prozess beschreibt, Ziel ist die Darstellung eines normativen Modells und seiner Fehler. Dementsprechend soll hier eine Kategorie akzeptiert werden, die streng genommen keine natürliche Kategorie darstellt. Die Eigenschaft, durch das Modell falsch klassifiziert zu werden, ist keine direkte Eigenschaft der Fälle selbst.

Zum anderen sollte definiert werden, welchen Status die einzelnen Abbildungen beanspruchen können. Zunächst ist $\hat{I}$ als reine Konvention zu betrachten, die in ihrer Form unverrückbar ist. Die Abbildung I hingegen ist - wie bereits angesprochen - entweder die Identität oder die Transitionsfunktion bis zum Zeitpunkt der Kriterienerhebung. Ob diese Funktion als deterministisch aufgefasst wird, ist eine Frage des

[14] Die Autoren fassen auch Analogien unter die Klasse der q-Morphismen, allerdings als q-Morphismen zweiter Ordnung: Bei Analogien werden in diesem Sinne mentale Modelle eines bestimmten Bereichs genutzt, um ein mentales Modell für einen neuen Bereich zu konstruieren:

> In many cases [...] the source is *simpler* than the target, in the sense of having fewer elements. Most generally a useful source will be one that (a) can itself be successfuly modeled and (b) preserves the critical causal properties relevant to the problem. A simpler problem is quite likely to satisfy condition (a); it may well fail condition (b), however. (Holland et al., 1986, S. 308)

[15] Sollen $w > 2$ Kriteriumskategorien berücksichtigt werden, so müssen $w^2 - w$ Fehlerkategorien eingeführt werden.

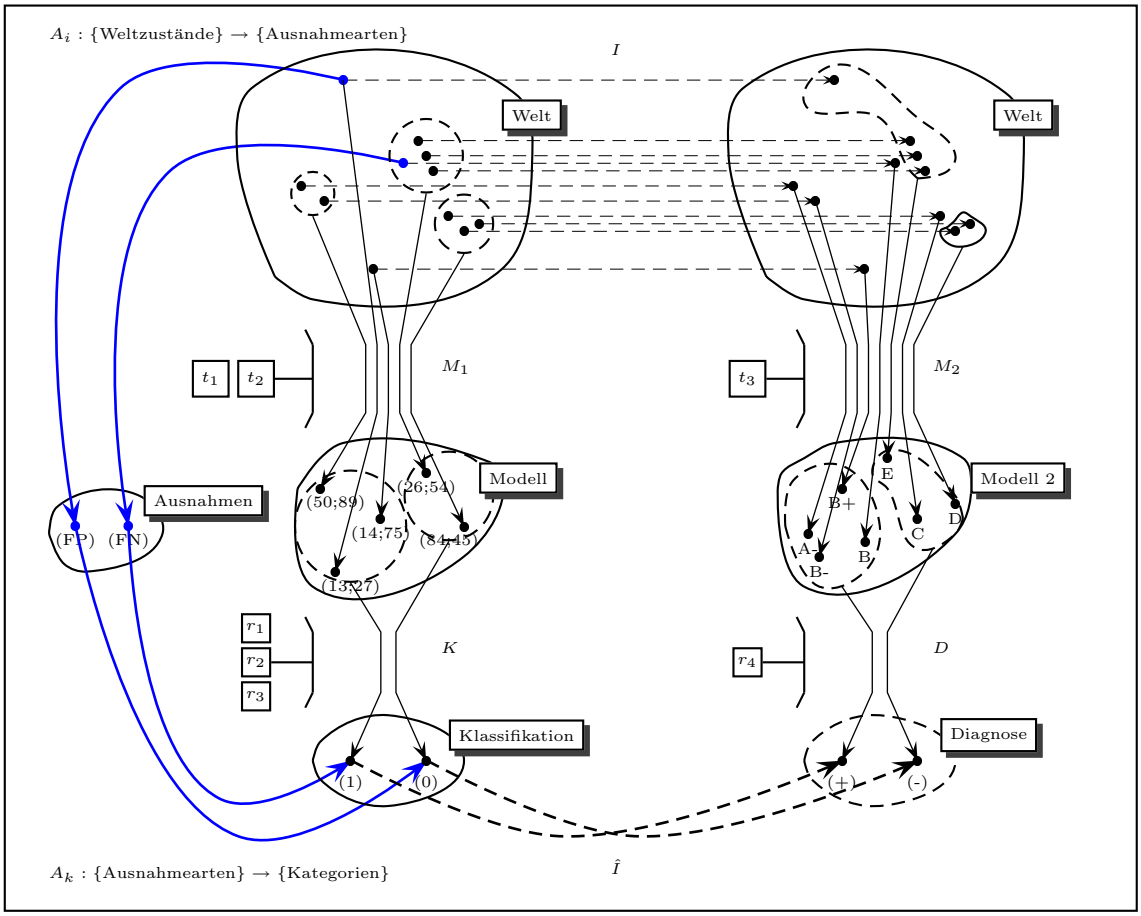


Abbildung 7.5: Das Ausgangsproblem als q-Morphismus: Das Modell wird um eine Fehlerbehandlung ergänzt. Falsch positive Klassifikationen und falsch negative Klassifikationen werden durch Sonderregeln auf die korrekten Klassifikationen abgebildet, so dass das Modell einen validen q-Morphismus darstellt.

Weltbilds und bzw. oder der Differenziertheit, mit der ein Personenzustand individuiert werden soll. Da M_1 und M_2 Messungen darstellen, und wenige psychologische Messverfahren ohne Messfehler existieren[16], können auch diese Abbildungen probabilistisch aufgefasst werden. Die Abbildung D ist wiederum eine konventionelle Festlegung, die allerdings nicht rein formal zu verstehen ist, sondern informiert sein muss durch die Details der Klassifikationsaufgabe. Unter Umständen ist hier Expertenwissen notwendig, um die Grenzen zwischen den beiden Kategorien zu definieren. Zuletzt ist K abhängig von der Liste der Entscheidungsregeln. D und K sollten als deterministisch betrachtet werden, d.h. Anwendungsfehler bei der Regelanwendung sollen in dieser Idealisierung ausgeschlossen sein.

7.2.1.2 Modell und Regelinduktionsproblem

An dieser Stelle lässt sich die Aufgabenstellung der Arbeit in Bezug auf das Modell präzisieren. Als Ausgangspunkt wird ein Datensatz genommen, in dem für jede zu klassifizierende Untersuchungseinheit Prädiktorwerte und ein binärer Kriteriumswert vorliegen. Aus diesem Grund fallen I, M_1, M_2 und D aus dem Spektrum der Untersuchung heraus. $\hat{I}$ ist konventionell festgelegt, A_i und A_k folgen ebenfalls den genannten Konventionen. Zielstellung ist die Festlegung der Regelmenge $\{r_1, r_2, \ldots, r_k\}$, so dass die Anzahl der Ausnahmen minimiert wird. Fokus ist also die Modifikation der Abbildung K, wobei die Existenz der übrigen Abbildungen vorausgesetzt wird. Damit ist die Sinnhaftigkeit des Vorgehens gebunden an eine Reihen von Prämissen, die ihrerseits im Rahmen dieser Arbeit nicht geprüft werden können. Zu diesen Voraussetzungen gehören die folgenden:

- Eine sinnvolle Menge an Prädiktoren wird identifiziert.
- Die verwendeten Messverfahren sind geeignet, um die identifizierten Prädiktoren zu messen.
- Die Messung der Prädiktoren wird reliabel genug durchgeführt, um eine sinnvolle Kategorisierung zu ermöglichen.
- Die dem Kriterium zugrunde liegenden Eigenschaften werden sinnvoll identifiziert.
- Die Messung dieser Eigenschaften wird durch ein geeignetes Messverfahren vollzogen.
- Die Messung dieser Eigenschaften wird reliabel durchgeführt.
- Die Zuteilung von Eigenschaftskategorien zu Kriteriumsklassen ist sinnvoll konzipiert und wird fehlerfrei durchgeführt.
- Das Transition zwischen Prädiktorerhebung und Kriterienerhebung ist sauber modelliert worden.

Diese Eigenschaften bleiben dabei innerhalb des Modells, sieht man weitergehend auf den Kontext der Klassifikation, so kann man die folgenden Punkte ergänzen:

- Das Intervall zwischen Prädiktorerhebung und Kriterienerhebung im Datensatz entspricht zeitlich und auch hinsichtlich der relevanten Veränderungseinflüsse dem Intervall im Anwendungsfall.
- Die Stichprobe der Fälle ist in den wesentlichen Eigenschaften repräsentativ für die Menge der Fälle im Anwendungsfall.
- Die Entscheidungssituation ist sinnvoll auf die Erfordernisse des Kontexts abgestimmt: die Entscheidung zum gewählten Zeitpunkt ist sinnvoll.
- Klassifikation und resultierende Entscheidungen und Handlungen sind den Interessen der Stakeholder angemessener als alternative Lösungen.
- Die Stichprobe ist ausreichend groß, um die Erstellung von Entscheidungsregeln zu gewährleisten, die für die Tragweite der Entscheidung angemessen sind.

[16] Der Fehleranteil bei Überprüfung von Geschlecht, Körpergröße oder Anzahl der Finger ist allerdings sicherlich kleiner als bei der Erfassung „sozialer Kompetenzen“.

- Kriterienkategorien und daraus resultierende Entscheidungen und praktische Handlungen sind angemessen aufeinander abgestimmt.
- Die Anzahl der Kriterienkategorien ist auf die Erfordernisse der Situation abgestimmt.
- Die Befugnisse und Interessen der beteiligten Parteien werden in der Formulierung der Entscheidungsprozedur angemessen berücksichtigt, so dass ihre Anwendung möglich gemacht wird.
- Die für die Entscheidung notwendigen Erhebungen und der Verlauf der Entscheidungsprozedur sind für die Beteiligten akzeptabel und führen nicht zu Reaktionen mit negativen Konsequenzen.
- Entscheidungsprozedur und ihre Anwendung sind verträglich mit ethischen Normen und Konventionen des lokalen Kontexts, mit den rechtlichen Rahmenbedingungen und menschlichen Grundrechten.

Die angesprochenen Punkte werden im Rahmen dieser Arbeit in keinerlei Hinsicht als trivial oder irrelevant eingestuft. Allerdings werden diese Punkte als inhaltlich unabhängig von der Lösung des Teilschritts behandelt, die als Aufgabenstellung der Arbeit angesetzt wird. Diese Unabhängigkeit muss qualifiziert werden: Auch die Bereitstellung von formalen Lösungen von Teilschritten unterstützt die Praxis, die auf diese Lösungen angewiesen ist. Die Legitimität dieser Praxis wird auch durch nicht-affirmative Lösungsangebote konsolidiert und die Beschränkung auf formale Texte kann mitunter sogar als Flucht vor einer (gegebenen) Verantwortung wahrgenommen werden[17]. In diesem Text soll aber keine Blankovollmacht für Klassifikationspraktiken ausgestellt werden, die kritische Reflexion der Grenzen der Modellierungen von Rationalität bilden einen Schwerpunkt im abschließenden Teil dieser Arbeit. Intendiert ist eine kritische Haltung gegenüber der Praxis, nicht aber eine negative Voreingenommenheit. Bowker und Star (1999) formulieren hierzu:

> Each standard and each category valorizes some of view and silences another. This is not inherently a bad thing - indeed it is inescapable. But it *is* an ethical choice, and as such it is dangerous - not bad, but dangerous. (S. 5 f.)

Einige der nahe liegenden Probleme, die für das Feld der Eignungsdiagnostik charakteristisch sind und dementsprechend eine besondere Relevanz für die in der Darstellung des Grundproblems dominante Einkleidung haben, konnten im Verlauf der Arbeit nicht näher ausgeführt werden. Ihre kritische Erörterung ist von anderer Seite ausführlich vorgenommen worden, einige grundsätzliche Verweise sollen im folgenden Abschnitt die Grenzen der Aufgabenstellung noch einmal in knapper Form von der abgewandten Seite beleuchten. Die Ausläufer des *Frame*-Problems reichen bis zum Ausgangspunkt der Arbeit zurück.

7.2.2 Ausgeklammerte Aspekte der Aufgabenstellung

In diesem Abschnitt sollen mit der Aufgabenstellung verbundene Problemkomplexe umrissen werden, die im Rahmen der Arbeit bisher nicht thematisiert werden konnten und somit valide Angriffspunkte darstellen: Ziel dieser Betrachtung ist dementsprechend eine Abtastung der Aufgabenstellung von außen. Im Einzelnen wird über die Rolle von Detektoren und Messinstrumenten gesprochen, über das Problem der Bestimmung und Messung von Kriterien, über Präditktorenauswahl und Verrechnung, über institutionelle Rahmenbedingungen für Personalauswahl und einige gesellschaftlich relevante Gesichtspunkte des Themas.

7.2.2.1 Detektoren und Messinstrumente

Grundsätzlich ist zu überprüfen, ob es sich bei den für die Klassifikation verwendeten Prädiktoren überhaupt um messbare Eigenschaften handelt (vgl. Suppes & Zinnes, 1963; Pfanzagl, 1968). Häufig geht die Messung der Überprüfung ihrer Legitimität aber voraus:

> Instead of candidly admitting their ignorance, quantitative psychologists promote their attributes as quantitiative and their procedures as instruments of measurement. [...] in the absence of experimental tests known to be specifically sensitive to the hypothesised additive

[17] In Präsentationen der Thematik dieser Arbeit konnte der Verfasser wiederholt die Prävalenz dieses Denkmusters beobachten.

> structure of the attributes studied, it is not known whether or not these attributes are quantitative and thus it is not known whether or not existing procedures measure them. (Suppes & Zinnes, 1963, S. 216)

Das Versäumnis ist in jenen Fällen besonders problematisch, wenn sogar gute Gründe vorliegen, die *gegen* eine Messbarkeit sprechen. Die Verwechslung von Messung und Wissenschaftlichkeit geht dabei oft auf eine falsche Vorstellung über Parallelen von Physik und Psychologie zurück. Luce (1964) mahnt hierzu an, dass sich die Psychologie wenn überhaupt auf dem Niveau der Physik der siebzehnten Jahrhunderts befindet und immer noch langsam und umständlich aus den zahlreichen psychologischen Ideen und Konzepten die fundamentalen Variablen herausschält. Dabei sieht er die Vorstellung von Messbarkeit durch Dekret ebenfalls als schlechte Praxis an:

> There are no reasons that I know of to suppose that either of these notions, hostility or aggression, and many others like them, are scalar quantities. To begin a problem in this way, is tantamount to throwing most of it away. At our present state of knowledge, to „suppose“ a measurement problem out of existence is little more than a feeble joke. (S. 366 f.)

Dabei sieht Luce manche Probleme als lösbar an durch mehr Originalität in der Modellierung, in anderen Fällen sieht er zukünftige Fortschritte in der Mathematik als notwendige Voraussetzung. Diese Methode der Rettung hat auch in der Physik nur bescheidene Erfolge verbuchen können. Ein Teil des Problems ist dabei die messtheoretische Frage; eine wesentliche, darüber hinaus gehende Frage ist aber die begriffliche Klärung der postulierten Eigenschaft. Um das Untersuchungsobjekt als Träger dieser Eigenschaft zu identifizieren, müssen zum einen die Grenzen zwischen Person und Kontext definierbar sein:

> That which is truly qualitative does not exist on a continuum so that it can be scaled. Love is not an entity that we possess more or less than others. Each person's understanding of and ability to love is different. Qualities take on their meaning from the context, cultural, historical, and individual, of derivation and application. (Stivers, 2004, S. 103)

Eine gewisse begriffliche Verwirrung zeigt sich im Übrigen bei Eigenschaften, die explizit eine soziale Situation brauchen, um sichtbar zu werden und dabei in ihrer Auswirkung von den Kontingenzen der sozialen Situation abhängig sind. So ist eine teamunabhängige „Teamkompetenz“[18] nicht zu attestieren, wenn nicht die Beschaffenheit des Teams in engen Grenzen definiert wird. Eigenschaften, die eine hervorragende und produktive Eingliederung in ein Akademikerteam im Bereich der Makroökonomie gewährleisten, können sogar negative Auswirkungen auf den Versuch haben, die eigene soziale Stellung in der lokalen Fußballmannschaft zu behaupten und umgekehrt. Berücksichtigt man zudem die Störbarkeit der Eingliederung durch zahlreiche ungewünschte Eigenschaften (Jähzorn, Aggressivität, Übellaunigkeit etc.) oder das Fehlen notwendiger Eigenschaften (mindestens basale Beherrschung von Sprache, Bewegungskoordination etc.), erscheint „Teamkompetenz“ als ein Makrobegriff ohne sinnvolle Zuordenbarkeit auf Entitäten, Eigenschaftenbündel oder Dimensionen. In diesem Fall muss der Frage der Messbarkeit also die Frage nach der begrifflichen Kohärenz des Konzeptes vorausgehen. Wittgensteins Betrachtung der Psychologie ist geprägt durch die Diagnose einer massiven Begriffsverwirrung:

> Die Verwirrung und Öde der Psychologie ist nicht damit zu erklären, daß sie eine „junge Wissenschaft“ sei; ihr Zustand ist mit dem der Physik z.B. in ihrer Frühzeit nicht zu vergleichen. [...] es bestehen nämlich, in der Psychologie, experimentelle Methoden *und Begriffsverwirrung*. [...] Das Bestehen der experimentellen Methode läßt uns glauben, wir hätten das Mittel, die Probleme, die uns beunruhigen, loszuwerden; obgleich Problem und Methode windschief aneinander vorbeilaufen. (Wittgenstein, 1993, S. 580)

Eine für Wittgenstein mögliche Ursache dieser Verhältnisse liegt in der besonderen Art und Weise, wie die Psychologie im Gegensatz zu Naturwissenschaften ihre Begriffe aus einer vorwissenschaftlichen Zeit erbt:

[18] Ähnliche Beispiele ließen sich für die meisten Modekonzepte geben. Bspw. schwankt „emotionale Intelligenz“ (vgl. Murphy, 2006) in manchen Darstellungen zwischen einem Sammelbegriff für wundersame Fähigkeiten zur Konflikttranszendenz bis hin zu einem Allheilmittel für verfahrene Organisationsstrukturen.

> Die Begriffe der Psychologie sind eben Begriffe des Alltags. Nicht von der Wissenschaft zu ihren Zwecken neu gebildete Begriffe, wie die der Physik und Chemie. Die psychologischen Begriffe verhalten sich etwa zu denen der strengen Wissenschaften wie die Begriffe der wissenschaftlichen Medizin zu denen von alten Weibern, die sich mit der Krankenpflege abgeben. (Wittgenstein, 1999, §62)

Selbst wenn ein Konzept identifiziert und leidlich scharf abgegrenzt werden kann, muss es im Folgenden in ein Messinstrument überführt werden können. Da diese Übertragung oft auf dem sprachlichen Weg versucht wird, scheitert eine große Anzahl von Ansätzen an der sprachimmanenten Ambivalenz und Interpretationsvielfalt. Gergen, Hepburn und Comer Fisher (1986) demonstriert, wie nahezu jedes Verhalten konzeptuell mit beinahe jeder Eigenschaft plausibel verbunden werden kann:

> [...] predictive validity does not appear to offer an objective crucible for interpreting trait measures. Behavior patterns, such as being persuasive, active, assertive, and so on are, like assertions on personality tests, subject to a high degree of interpretive indeterminacy. Each may be the overt result of myriad source dispositions. (Gergen et al., 1986, S. 1268)

Insgesamt fällt auf, dass viele Testverfahren denkbar schlecht darauf zugeschnitten sind, ideosynkratische Anteile des Denkens und der Interpretation verbaler Aussagen zu berücksichtigen. In den meisten Messverfahren sind deshalb Situationen ununterscheidbar, in denen Versuchspersonen erratisch reagieren oder in denen sie individuell reagieren:

> [...] individuality needs to be considered more systematically than it has been in the personnel selection literature. [...] It is too often assumed that uniqueness can be pushed into the category of the prediction error. We need to find ways to identify why subgroup differences exist rather than simply brushing them aside as ubiquitous error assuming that they represent some basic social flaw. [...]. (Guion, 1999, S. 357)

Immer noch sind zudem Messverfahren im Umlauf, die ihre Überzeugungskraft eher aus suggestiven und mythischen Quellen beziehen denn aus wissenschaftlich gestützter Qualitätsbeurteilung. Schmidt und Hunter (1998) weisen darauf hin, dass z.B. die Nutzung der Graphologie einer Zufallsauswahl nicht überlegen ist. Viele Messinstrumente auf dem Markt sind zudem an studentischen Stichproben normiert, was ihre Anwendbarkeit in entfernteren Settings zu einer Hypothese werden lässt (vgl. Landy & Shankster, 1994a). Probleme dieser Art sollen an dieser Stelle nur angerissen werden, sie werden im Bereich der psychologischen Testtheorie (vgl. Wottawa, 1980) vertiefend diskutiert. Fortschritte hinsichtlich der Messinstrumente sind allerdings nur dann gewinnversprechend, wenn die Messungen so konzipiert worden sind, dass mit den Resultaten sinnvolle Fragen beantwortet werden können:

> Superior instrumentation simply produces more precise and reliable answers, it does not make better questions. (Silverstein, 1976, S. 410)

Schärfer formuliert Guion (1991):

> A rush to measure is often ill-advised. Successul prediction of a trivial aspect of performance that happens to be easily measured will effect little improvement in important aspects of performance. Trivial measurement is a likely consequence of trivial thought. (S.334)

Aus diesem Grund soll als nächstes die Auswahl des Kriteriums für den Aufbau von Auswahlverfahren geprüft werden.

7.2.2.2 Das Kriterienproblem

Die Beobachtung des Kriteriums ermöglicht letztendlich erst die Beurteilung und die Entscheidung; mit diesem Sinne kann der Begriff aus seinen Ursprung zurückverfolgt werden:

> Der Begriff Kriterium stammt vom griechischen 'krinein'[19] (trennen, unterscheiden, urteilen, entscheiden; s.a. Kritik, Krise). Das Kriterium ist das Entschiedene (und deswegen Festgesetzte, Gewisse, Sichere, Geltende, Indiskutable); es ist zugleich das Entscheidende (Maßstab/-Messlatte /Bezugsstandard) mit der Funktion, das Eine vom Anderen unterscheiden und ihm einen spezifischen Wert zuschreiben zu können. (Neuberger, 2002, S. 289)

Das Kriterium bleibt aber immer noch auf der Ebene des Anzeichens für die Entscheidung und vor der Entscheidung, es ist nicht selbst Ziel der Entscheidung oder Konsequenz einer Entscheidung, es fällt zudem in eine andere Kategorie: Als Kriterien können sich Eigenschaften qualifizieren, die in einer normativen Praxis der Beurteilung Einfluss auf die Fällung von Urteilen haben können, weil ein Urteil in Bezug auf sie begründet werden kann. Im Beispiel der psychologischen Diagnostik handelt es sich dabei im Allgemeinen um Eigenschaften der zu diagnostizierenden Personen, bei der Personalauswahl um Eigenschaften der Bewerber, mit denen die Auswahlentscheidung begründet werden kann. Prädiktoren sind wiederum Variablen, die zum Kriterium in einem Verhältnis der Vorhersage stehen, entweder in einem kausalen oder im korrelativen Sinne und damit auch vermittelt in ein Begründungsverhältnis zum Urteil gesetzt werden können. Am Beispiel der Führungskräfteauswahl lokalisiert Neuberger (2002) die Position des Kriteriums im diagnostischen Auswahlprozess in folgender Weise:

> Formuliert man die Führungskräfteauswahl als ein rationales Problem, dann kommt es darauf an, in einer Menge von KandidatInnen diejenigen zu identifizieren, in die es sich zu investieren lohnt, weil sie *später* erfolgreich (absolutes Kriterienniveau) oder erfolgreicher als die anderen (relative Stellung in der Population) sein werden. (S. 289)

Die Beziehung zwischen Prädiktoren und Kriterium entspricht im oben ausgeführten Modell der Beziehung zwischen den Ergebnissen der Detektoren auf der linken Seite und den Kriteriumswerten auf der rechten Seite. Ghiselli und Haire (1969) mahnen hierzu an, dass die Messung des Kriteriums zu einem einzigen Zeitpunkt deshalb Probleme aufwirft, weil ein beruflicher „Erfolg“ über längere Zeiträume nicht unbedingt als stabil betrachtet werden kann, wenn Schwankungen in der durchschnittlichen Leistung berücksichtigt werden. Sicherlich lassen sich erreichte Positionen, ein erzieltes Gehalt oder die übertragene Führungsverantwortung beurteilen, aber der Begründungszusammenhang zur Entscheidung ist nicht ohne Schwierigkeiten aufzuzeigen.

Wottawa und Hossiep (1987) verweisen auf Thorndikes Konzept der Trennung von eigentlichem [*ultimate*] und aktuellem Kriterium (Thorndike, 1949). Die Vorstellung von dem, was zur Entscheidungsfindung als sinnvoller Maßstab betrachtet wird („aktuelles Kriterium“) bzw. auch das, was zu einem Zeitpunkt zur Verfügung steht, kann von der Konzeption abweichen, die für eine Fundierung der Entscheidung im Sinne des Problems sinnvoll wäre („tatsächliches Kriterium“). Berücksichtigt man zusätzlich, dass die gewählte Beobachtung oder Messung möglicherweise nicht das intendierte Konstrukt oder die gesuchte Eigenschaft erfasst, ergeben sich einige Möglichkeiten des Zusammenspiels der drei Komponenten, die mitunter absurde und auch ironische Züge annehmen können. Sind beide Kriterienarten miteinander deckungsgleich, so reduziert sich die Frage auf die Qualität und Bandbreite der Messung. Gibt es Abweichungen zwischen den Kriterienarten, kann sogar der Fall auftreten, dass ein Messverfahren etwas anderes misst als gewünscht, aus diesem Fehler aber ein Vorteil werden kann, weil in der Tat das eigentliche Kriterium gemessen wird. Umgekehrt kann eine präzise Messung des aktuellen Kriteriums die Messung des tatsächlichen Kriteriums behindern:

> Als *Beispiel* hierfür könnte die Menge an persönlichen Kontakten zwischen Vorgesetztem und Untergebenem dienen, die als subjektiver Faktor die Evaluation der Güte der Arbeit des Untergebenen durch den Vorgesetzten positiv beeinflussen kann. (Weinert, 1998, S. 265)

Auch Mischformen sind denkbar, so dass aus dem Erfolg einer Entscheidung nicht notwendigerweise auf die Korrektheit der eingesetzten Konzeptionen geschlossen werden kann (vgl. Binning & Barrett, 1989). Zudem stellt das „eigentliche Kriterium“ notwendigerweise eine Idealisierung dar, die der Komplexität menschlicher Interaktion und der Volatilität institutioneller, gesellschaftlicher und wirtschaftlicher Entwicklungen nur in einem sehr dünnen Sinne gerecht werden kann: nämlich als retrospektiv konstruierte

[19] (*κρίειν*), aber auch *kriterion* bzw. (*κριτήριον*): Mittel zur Treffung des Urteils, entscheidendes Anzeichen, auch Zeichen oder Signal des Gerichts, A.d.V.

Ideallösung unter zahlreichen *ceteris paribus*-Annahmen oder sogar nur als Ergebnis eines kontrafaktischen Experiments, in dem eine Bandbreite von Kriterien in Parallelwelten im gleichen Zeitintervall eingesetzt werden, um im Vergleich der entstandenen Zielzustände und bei jeweils restrospektiv konstruierter Ideallösung eine Entscheidung treffen zu können. Ob dieses Kriterium „existiert", ist eine Frage, die nur im Hinblick auf metaphysische Vorlieben bezüglich Lesbarkeit und Struktur der Welt beantwortet werden kann. Als „Fiktion" mit ungeklärtem Realitätsbezug kann dieses Kriterium trotzdem als Veranschaulichung dienen und soll in diesem Sinne hier betrachtet werden.

Eine diagnostische Entscheidung ist selten eine logisch abgeleitete Entscheidung, der Schluss wird in den meisten Fällen induktiver Art sein, also vom Besonderen der Beobachtung und gesammelten Erfahrung zum Allgemeinen, den gültigen generellen Zusammenhängen geführt (Wottawa & Hossiep, 1987):

> Induction, primitively, was a mere matter of expecting that events that are similar by our lights will have sequels that are similar to one another. The larger the class of mutually similar antecedent events may be, all of which had mutually similar sequels, the stronger is the assumption of a similar sequel the next time around. (Quine, 1981, S. 153)

Der Prozess des Schließens von Prädiktoren auf ein Kriterium und von einem Kriterium auf eine valide Entscheidungsgrundlage ist dementsprechend zweifach störbar. Um aber überhaupt zu einer empirischen Fundierung von Entscheidungsverfahren zu kommen, sind Informationen über Kriterienwerte erforderlich. Dies erzeugt ein Bündel von Problemen, vor allem, wenn Kriterienwerte und Entscheidungen nicht unabhängig voneinander stehen. Soll etwa die Leistung nach drei Jahren im Unternehmen als Kriterium verstanden werden, stehen mir zur Überprüfung nur eingestellte Mitarbeiter zur Verfügung, die Leistung der abgelehnten Mitarbeiter kann nicht erhoben werden. Man braucht zur Fundierung der Entscheidung in diesem Fall Daten, die nur nach getroffener Entscheidung erhoben werden können. Wenige Unternehmen gehen zur Praxis über, solange alle Bewerber einzustellen, bis ein Verfahren fundiert ist. Die zuweilen eingesetzte Methode der konkurrenten Validierung, also der gleichzeitigen Erhebung von Prädiktor- und Kriteriumswerten, ist nur in manchen Situationen sinnvoll. Zur Evaluation eines Screening-Verfahrens, das angewandt wird als einfachere, angenehmere oder kostengünstigere Alternative zu anderen, valideren Verfahren(vgl. Woike, 2003a), kann diese Variante bspw. vorteilhaft sein, in vielen anderen Fällen führt sie zu einer Problemvereinfachung, die die gefundene Lösung bis zur Nutzlosigkeit verzerren kann (Guion & Gottier, 1965). Die Stichprobe, für die Kriteriumswerte zur Verfügung stehen, ist zudem nur in seltenen Fällen repräsentativ für die Bewerber. Die Festsetzung und Erhebung der Kriterien stellt den äußeren Rand der Evaluierbarkeit und Sichtbarkeit dar. Wenn keine Sonderfallregelung besteht, die die Bedeutung der Kriterien in Einzelfällen außer Kraft setzt, kann ein Entscheidungsverfahren nicht objektiver sein als das definierte *und erhobene* Kriterium zusammen mit seiner Integration in den Entscheidungsprozess. Weinert (1998) bündelt Kriterienprobleme je nach Art und Ursache in insgesamt sechs Gruppen:

1. die *Zeitkomponente* (Abhängigkeit der Messung von der Zeit)
2. das *Konstrukt* (ultimatives, d.h. „absolutes" oder wirklich gemessenes Kriterium)
3. die *Eigenschaften des Kriteriums* (Reliabilität; Relevanz; Effizienz; Kontamination, Praktikabilität)
4. die *Anzahl* der zu wählenden Kriterien
5. *Kompositive* versus *multiple* Kriterien (deren Gewichtung und Kombination)
6. die *Kriterienentwicklung* (Kriterien für Kriterien) (S. 262)

Die beiden erstgenannten Probleme sind bereits diskutiert worden. Die dritte Thematik fokussiert zum Teil auf Eigenschaften des Kriteriums, zum Teil aber auch auf Eigenschaften der Messung des Kriteriums. Sicherlich ist es möglich, Kriterien zu konzipieren, die für die geplante Entscheidung hervorragende Eigenschaften haben, aber die sich entweder der Messung vollständig entziehen, die nur sehr unpräzise oder fehlerbehaftet messbar sind oder deren Messung nicht unter akzeptablen Bedingungen vorgenommen werden kann. Die Lösung stellt zumeist einen Kompromiss dar, der reliabel, relevant und praktikabel bleibt.

Weiterhin stellt sich die Frage, ob ein einzelnes Kriterium ausreicht oder ob ihm weitere Kriterien zur Seite gestellt werden sollten. Berufliche Leistung kann unter verschiedenen Gesichtspunkten aufgefasst

werden: Spitzenleistungen und Kontinuität bilden eine mögliche Gegenüberstellung, aber auch Motivation und Fähigkeit, Grundvoraussetzungen und Zusatzfähigkeiten etc. Bei einem zusammengesetzten Kriterium sieht Weinert (1998) zudem verschiedene Formen der Gewichtung, je nachdem, ob die verschiedenen Bestandteile als Unterfacetten eines gemeinsamen Konstrukts oder als eigenständige Dimensionen aufgefasst werden. Die Gewichtung von Facetten kann durch Experten vorgenommen werden, von den Reliabilitäten oder Korrelationen zu weiteren Variablen abhängig gemacht werden, hinsichtlich der Vorhersagbarkeit oder aufgrund einer Faktorenanalyse vorgenommen werden oder die Maximierung individueller Unterschiede sicherstellen. Weiterhin bieten sich eine Gewichtung nach monetärer Entsprechung oder eine einfache Gleichgewichtung als mögliche Strategie an. Bei multiplen Dimensionen unterscheidet Weinert weiterhin zwischen einer statischen Dimensionalität, einer dynamischen, die sich über die Zeit verändert sowie einer individuellen, die unterschiedliche Arbeitsstile berücksichtigt. Als Kriterien für Kriterien diskutiert Weinert schließlich Fragen der Reliabilität, Validität, der Umsetzbarkeit im Sinne von Eignung zu Feedback oder Festsetzung der Entlohnung und Fragen der Akzeptanz, die sowohl in Bezug auf den Anwender, aber auch in Bezug auf finanzielle Kosten, notwendige Datenerhebungen usw. geprüft werden muss. Letztendlich stellen sich hier auch Fragen der Fairness, der möglichen Diskriminierung und der ethischen Kompatibilität mit dem bestehenden Wertesystem.

Neuberger (2002) betrachtet Kriterien für Führungskräfte auf mehreren Ebenen, die verschieden eng gefasste Sichtweisen auf die Leistung darstellen. Dabei sieht er fünf Stufen in aufsteigender Breite, die von generellen Erfolgsbedingungen über Persönlichkeitsfaktoren im Spiegel von AC-Verfahren hin zur Vorgesetztenleistung, zu erreichten und verfehlten Unternehmenszielen bis zu geschaffenen und vernichteten (gesellschaftlichen) Werten führen. Wottawa und Hossiep (1987) ergänzen hier noch die Perspektive des für das Entscheidungsverfahren verantwortlichen Psychologen, der Ergebnisse anders beurteilen kann als Diagnostizierte, Auftraggeber oder Gesellschaft. Je nach Interesse des Entscheiders werden aber nicht alle Stufen Berücksichtigung finden, was die Nichtuniversalität und relative Subjektivität von Kriterien unterstreicht.

Guion (1991) betont, dass Kriterien nicht einfach untereinander ausgetauscht werden können; Forschungsergebnisse in Bezug auf ein Kriterium sind deshalb nicht ohne weiteres für ein zweites gültig oder auch nur relevant. Für jede Organisation sieht er zudem eine spezifische Hierarchie von Kriterien. Selbst die Einheit der Beurteilung ist nicht eindeutig zu bestimmen. Landy und Shankster (1994a) betonen den Aspekt der Gruppenleistung, die auch auf Gruppenzusammensetzungen zurückführbar ist. In jeder Gruppe gibt es unterstützende und befähigende Rollen - genau so wie potenziell retardierende und behindernde Rollen - die in keinem oder sogar in einem negativen Verhältnis zu Produktivitätsindizes für einzelne Mitarbeiter stehen. Der Versuch der Objektivierung von Kriterien nimmt zum Teil die Form der Überführung von Kriterienwerten in eine Währung an, so dass direkt die monetäre Konsequenz von Entscheidungen erhoben werden kann. Ähnlich ist die Einführung künstlicher Nutzendimensionen zu betrachten wie zum Beispiel in Form der multiattributiven Nutzenmessung (MAUT), in der der „Gesamtnutzen“ in Teilaspekte zerlegt wird, deren Gewichtungen durch Experten bestimmt und deren Ausprägungen im Beurteilungsfall empirisch zu bestimmen sind. Die Anzahl der notwendigen Voraussetzungen und Vereinfachungen in einem derartigen Modell allein durch die erzwungene Eindimensionalität und Invarianz der Komponentengewichte ist offensichtlich. Austin und Villanova (1992) fassen einen Überblick über 85 Jahre der Forschung zu Kriterien (1917-1992)[20] mit den Worten zusammen:

> The problems of criteria have been and remain multiple. Criteria are dynamic, multidimensional, situation-specific, and serve multiple functions. Given this complexity, it would be unreasonable to assume that a hypothetico-deductive model of explanation, which provides premises from which events can be deduced, could offer a sufficient basis for understanding criteria. (S. 862)

7.2.2.3 Prädiktoren und Verrechnungsmodelle

Die für die Auswahl des Kriteriums angestellten Überlegungen können zu einem großen Teil auch für Prädiktoren angestellt werden. Auch bei Prädiktoren stellen sich die Fragen der Erhebbarkeit, der Übereinstimmung von gemessenem und gewünschtem Konstrukt, der Akzeptanz, der Dimensionalität und der

[20] Guion (1987) sieht den Beginn der systematischen Forschungsbemühungen hinsichtlich der Personalauswahl in den Zwanziger Jahren, der Überblick deckt deshalb einen wesentlichen Teil der wissenschaftlichen Arbeit ab.

Hierarchie. Die Relevanz von Prädiktoren wird aber nicht unmittelbar auf die Entscheidung bezogen, sondern im Hinblick auf das (Entscheidungs-)Kriterium beurteilt. Stellt man den möglichen Verschiebungen von aktuellem und tatsächlichem Kriterium einerseits und dem gemessenen Kriterium andererseits nun noch die mögliche Spaltung von Prädiktorenkonstrukt und Prädiktor-Messwert gegenüber, entstehen weitere Kombinationsmöglichkeiten und Verfehlungsarten (vgl. Wottawa & Hossiep, 1987). In einem Extremfall könnte das eigentliche Kriterium durch den Prädiktoren-Messwert exakt abgedeckt werden, ohne dass das aktuelle Kriterium dem eigentlichen entspricht, ohne dass das aktuelle Kriterium exakt gemessen wird, ohne dass das Prädiktorkonstrukt ansatzweise eines der anderen Konzepte abdeckt. In diesem Fall würde allerdings nichts im Prozess auf den erreichten Erfolg hindeuten, da die Korrelation zwischen Prädiktor- und Kriteriumskonstrukt schlecht sein wird. Umgekehrt muss eine hohe Korrelation nicht auf einen Erfolg hinweisen, da die übrigen drei Komponenten der Konzeption theoretisch in beliebiger Weise zu den beiden überlappenden Messwerten positioniert sein könnten. Die Konstruktvalidierung von Messwerten kann unter Umständen bestimmte Divergenzen sichtbar werden lassen. Für die Entscheidungssituation selbst stehen aber nur Prädiktor-Messwerte und Kriteriums-Messwerte zur Verfügung, ein Entscheidungsverfahren kann nicht mehr leisten, als aufgrund von Prädiktor-Messwerten die Kriteriums-Messwerte möglichst gut vorherzusagen. Die in der Aufgabenstellung gesuchten Entscheidungsregeln übernehmen genau diese Aufgabe. Selbst eine exzellente Erfüllung ist keine Garantie für ein gutes Entscheidungsverfahren, sondern impliziert lediglich ein gutes Entscheidungsverfahren hinsichtlich der Prämissen, dass beide Messungen eine gute Qualität aufweisen, und Prädiktor-Konstrukt und eigentliches Kriterium - sofern es existiert - eine hohe Überlappung aufweisen.

Die entstehenden Fragen lassen sich nicht für jeden Anwendungsfall durch Rückgriffe auf die Forschungsliteratur beantworten. Guion (1991) weist darauf hin, dass es wenig nützlich erscheint, wenn entweder Forschungsergebnisse für eine vergleichbare Situation nicht verfügbar oder nur fragwürdige Arbeiten zur Auswahl stehen. Sorgfältige eigene Beobachtung kann in diesen Fällen bessere Arbeit leisten als das Vertrauen in problematische Beobachtungen anderer. In Bezug auf die prinzipielle Erfassbarkeit von komplexem Verhalten durch Prädiktorkonstrukte weist er auf drei Prinzipien hin, nach denen komplexes Verhalten nicht voll mit einfachen Mitteln vorausgesagt werden kann, dass dieses Verhalten nur zu einem Teil eine Funktion der Charakteristika von Individuen sein wird und dass das Verhalten nicht für alle Menschen in gleicher Weise vorhersagbar sein wird. Selbst wenn diese Einsicht als trivial begutachtet werden kann, weist sie auf eine weitere Grenze hin, der sich selbst das „eigentliche Kriterium“ stellen muss[21].

Die Suche nach Prädiktoren sollte neben offensichtlichen Abhängigkeiten wie Beziehungen zwischen Prädiktor und Kriterium auch durch das Modell ihrer späteren Verrechnung beeinflusst werden. Auch bei Prädiktoren ist die sinnvolle Einsetzbarkeit für die bestehende Situation zu prüfen:

> ... a home-made personality or interest measure, carefully and competently developed for a specific situation, is a better bet for prediction than is a standard personality measure with a standard system of scoring. (Guion & Gottier, 1965, S. 159)

Auch bei den Prädiktoren sollte Wert darauf gelegt werden, dass bei einer Einzeltestung nicht Merkmale getestet werden, die sinnvoll nur bei Gruppen oder ganzen Organisationen getestet werden können oder die zu einer Internalisierung organisationaler Probleme genutzt werden. Scheitern zahlreiche Projekte an übertrieben eng gesetzten Zeitfenstern, mag das Scheitern der Projekte zwar von persönlichen Schwierigkeiten in Teams begleitet sein, die Einstellung „teamfähiger“ Mitarbeiter wird das grundsätzlichere Problem aber nicht beheben können. Ähnlich kritisch - wenn nicht zynisch - sind Konstruktionen wie „Zufriedenheitskompetenz“ oder „Erholungskompetenz“ zu bewerten:

> Die Personalauswahl kann die als gegeben betrachteten Rahmenbedingungen - insbesondere die sachbezogenen Arbeitsbedingungen - zwar berücksichtigen, damit ist jedoch der Erfolg der Stellenbesetzung nicht garantiert. Beispielsweise wird bei einer monotonen oder unbefriedigenden Tätigkeit auch diesbezüglich resistenteres Personal keine optimale Arbeitsleistung erbringen. (Schwarb, 1996, S. 28)

[21] Mit ausreichender methodologischer Fantasie könnten allerdings akzidentelle Vorhersagemöglichkeiten ohne Stabilität im nachfolgenden Zeitintervall mit einer Überwindung der Problematik verwechselt werden.

Ehrlicher wäre ein Ansatz, der die gewünschten Eigenschaften direkt als „funktionale Anpassungsfähigkeit an die nicht-plastischen Strukturen der spezifischen Organisation“ benennen würde. Die Vorstellung einer vollständigen monadischen Rekonstruierbarkeit bzw. Prognostizierbarkeit sozialer Dynamik sollte zudem die Einführungsveranstaltungen soziologischer oder sozialpsychologischer Fakultäten nicht unwiderlegt verlassen dürfen. Zudem sollte das Bedürfnis der Legitimation durch Quantifizierung nicht mit einer wissenschaftlichen Fundierung verwechselt werden:

> [...] sciences clamor to be quantitative; they clamor for something to measure. This is both good and bad. It is very good indeed if the measurable quantity can be found to play a significant role in the subject matter of the science in question. It is bad if in the quest for something to measure the scientist turns his back on the original concerns of his science and is borne away, however smoothly, on a tangent of trivialities. Ills of mathematization, as well as successs, can be laid to the quest for quantitativity. (Quine, 1981, S. 153)

Die Einbettung in freischwebende „Kompetenzmodelle“ mit unzulänglicher Rückführung auf den organisationalen Alltag stellt auch aus diesen Erwägungen nicht die Ideallösung für die Frage der Prädiktorenauswahl dar. Trotzdem wird der angerissene Problemkomplex im Rahmen der vorliegenden Arbeit als *außerhalb der Fragestellung* betrachtet, was gleichzeitig die Leistungsfähigkeit aller erstellten Aufgabenlösungen gravierend beschneiden muss. Gegen den aufkommenden Ton der Resignation ob der Nichterfüllbarkeit einer idealen Entscheidungsprozedur soll aber entschieden eingewandt werden, dass es zuweilen die Ideale sind, die einer Korrektur bedürfen:

> ... finality is the mark of runic explanations, not scientific ones; the road of inquiry is always open, and it reaches beyond the horizon. (Kaplan, 1964, S. 341)

7.2.2.4 Institutionelle Rahmenbedingungen

Entscheidungsprozeduren können vielleicht in einem luftleeren Raum der Ideen konzipiert und modifiziert werden, aber sie werden niemals in diesem Raum realisiert und implementiert. Im Aufeinandertreffen der Interessen des Diagnostikers mit den Strukturen von Organisationen machen sich Reibung und Schwerkraft mehr als deutlich bemerkbar. Von Seiten der Diagnostiker wird beharrlich auf den großen Nutzen der eigenen Praxis für die auftraggebenden Organisationen verwiesen (vgl. Schmidt & Hunter, 1983):

> In economic terms, the gains from increasing the validity of hiring methods can amount over time to literally millions of dollars. However, this can be viewed from the opposite point of view: By using selection methods with low validity, an organization can lose millions of dollar in reduced production. (Schmidt & Hunter, 1998, S. 273)

Natürlich erscheint eine solche Überlegung bei Annahme einer Vollbeschäftigung fast als Nullsummenspiel[22], aber die grundsätzliche Skepsis der Institutionen erscheinen vor diesem Hintergrund fast kontraintuitiv. Allerdings stellt Klehe (2004) diesbezüglich fest, dass Organisationen einschließlich Universitäten häufig Auswahlmethoden mit schlechter psychometrischer Qualität und schwacher juristischer Verantwortbarkeit einsetzen. Hier sollen einige der entstehenden Schwierigkeiten beleuchtet werden, die ebenfalls nicht in dem oben dargestellten Modell berücksichtigt werden konnten.

Eines der angesprochenen Probleme kann als Widerstand von Seiten der Organisation, aber auch des Diagnostikers gegen Erfolgskontrollen aufgefasst werden (Wottawa, 1989). Für das Management kann schon die Notwendigkeit einer Evaluation der eigenen Praxis auf Skepsis stoßen, erst recht, wenn diese Betrachtung das Potenzial hat, tatsächliche Verbesserungsmöglichkeiten zu identifizieren. Eine zukünftige Verbesserung macht automatisch den gegenwärtigen Zustand schlechter *als er sein könnte*. Ein „Bei mir gibt es keine Fehler“-Prinzip (Wottawa, 1989) setzt sich vor allem dann durch, wenn Lernfähigkeit in einem Unternehmen weniger geachtet wird als die ungetrübte Suggestion von Perfektion und Rationalisierungen des Suboptimalen. Weiterhin haben vergangene Fehlentscheidungen ein menschliches Gesicht, und die Identifizierung von „fälschlich Eingestellten“ ist zum einen sozialer Zündstoff, zum anderen selten konsensfähig. Mitunter wird selbst dem Diagnostiker der Einblick in die dementsprechend sensiblen

[22] Beispielsweise wird der Pool meiner Bewerber in der durchschnittlichen Leistung sinken, wenn *andere* Unternehmen leistungsstarke Selektionsmethoden einsetzen - hier setzt eine Spirale der „Aufrüstung“ an.

Erfolgsdaten verwehrt, so dass für ihn die Befürwortung einer Evaluationsmaßnahme ein viel direkterer Hinweis auf die Möglichkeit vergangener Fehler ist. Image und Selbstwertgefühl können unter der öffentlichen Zurschaustellung von Defiziten leiden, zudem ist gerade der junge „Diagnostiker im Training“ eine ähnlich unheilsschwangere Erscheinung wie der „Arzt im Praktikum“ oder ein „lernender Richter“ (Wottawa, 1989). Ein Psychologe im ökonomischen Kontext hat zudem mit der Stereotypisierung als Minderleister zu kämpfen, der außer leerer Rede und vermuteten Pathologien wenig zu bieten hat. Lediglich der Nimbus der Orakels, die projizierte Meisterschaft des Unbewussten und die beinahe paranormal anmutende Befähigung zum Durchschauen der Mitarbeiter können in extremen Fällen den persönlichen Raum des Diagnostikers schützen:

> Bei einem Offenlegen der (tatsächlichen) Entscheidungsgründe stellt sich aber das Problem des „gläsernen“, leicht ersetzbaren Spezialisten - die Stärke der Position des Spezialisten liegt ja gerade darin, daß er für das Unternehmen wegen seiner besonderen Kompetenz der Entscheidungsfindung, die Konkurrenten fehlt, nicht ohne weiteres ersetzbar ist. (Wottawa, 1989, S. 34f)

Sachdienlich ist die Verwechslung von Para-Psychologie und psychologischer Diagnostik allerdings nicht, was als Einsicht aber ein basales Wissen über professionelle Aspekte der Tätigkeit erfordert, welches keineswegs als zur Allgemeinbildung gehörend betrachtet werden muss. Klehe (2004) betrachtet eingehend die Konflikte zwischen organisationalen Stakeholdern und ihren Machtbefugnissen, die die Arbeit des Diagnostikers erschweren. Die von Schmidt und Hunter (1998) gerühmte vermögenssteigernde Leistungsfähigkeit psychologischer Verfahren kann zunächst einmal nicht ohne Angabe eines Zeitintervalls bewertet werden. Weiterhin sieht Klehe Diagnostik im Spannungsfeld zwischen rechtlicher Normiertheit und Bewerberakzeptanz:

> In the case of personnel selection, (a) long-term economic fitness benefits from the adoption of well-developed and validated selection procedures, while (b) short-term economic fitness is usually greater with the use of less refined procedures. Similarly, (c) legal regulations promote the adoption of well-developed selection procedures, while (d) applicant rejections may sometimes be more favorable towards less refined procedures [...]. (Klehe, 2004, S. 329)

Langfristige Vorteile und - unbeobachtete - Rechtmäßigkeit der Prozeduren stehen allerdings nicht unbedingt im Interesse der Aktionäre, deren Einfluss deshalb gegen eine wissenschaftliche Fundierung der Eignungsdiagnostik wirksam werden kann:

> [...] organizations that are under pressure to maximize shareholder wealth in the form of quarterly dividends and increasing stock value are not very likely to invest in the costs of developing and validating selection procedures that will only pay off in the longer term. This should make publicly traded organizations more likely to choose short-term strategies than organizations owned privately or by the public sector. (Klehe, 2004, S. 330)

Auch entstehende Kosten durch rechtliche Klagen oder Reputationsverlust mit nachfolgenden Problemen im Recruiting gehören zu den erst langfristig zu verbuchenden Kosten. Die Überzeugungskraft der Vorteile diagnostischer Technologien ist aufgrund der fehlenden Sichtbarkeit geringer als die Demonstrationen der Überlegenheit von Maschinen oder Informationstechnologien, und da eine Beziehung zu Nutzenerwägungen erst konstruiert werden muss, sind Manager weniger offen für vorgeschlagene Innovationen im diagnostischen Bereich. Sichtbar sind hingegen die Kosten hinsichtlich Entwicklung, Validierung und Verwaltung der Selektionsprozeduren sowie Verzögerung von Einstellungen. Diese persönlichen Einschätzungen sind aber für die Einführung von Verfahren häufig wichtiger als deren nachweislicher, aber nicht im Spezifischen nachgewiesener Nutzen:

> [...] it remains to be noted that it is not the actual long-term benefit obtainable but decision makers' perceptions thereof that will determine an organization's interest in using any specific selection procedure. (Klehe, 2004, S. 329)

Die intendierte Vermeidung von negativen Reaktionen bei Kandidaten erfordert ebenfalls Zugeständnisse. Dazu gehört nicht nur der offensichtliche Verzicht auf aversiv, unfair oder unzulässig erlebte Prozeduren zugunsten von positiv erlebten Verfahren, sondern auch die Berücksichtigung subjektiver Theorien über die Diagnostizität von Verfahren, die selten mit der tatsächlichen Leistungsfähigkeit übereinstimmen. Mikropolitische Prozesse können zudem zu Störungen im Prozess führen, wenn Sichtbarkeitsrelationen und Machtverhältnisse durch Verfahren und ihre Ergebnisse neu ausgehandelt werden können:

> [...] organizations might prefer unstructured selection procedures, since their inherent ambiguity may prevent scrutiny and monitoring of the selection process by outside parties, such as the human resource department or government agencies. (Klehe, 2004, S. 334)

Klehe bezeichnet den Prozess der Abkopplung durch Einführung von Sichtgrenzen auch als *Buffering*. Auch für den Diagnostiker selbst können Erwägungen über den Alltag seiner Arbeit zu einer Auswahl von Methoden führen, die primär den eigenen Bedürfnissen gerecht werden:

> [...] recruiters may regard structured interviews as a monotonous and boring exercise that reduces their ability to build a rapport with the interviewee as well as their power in deciding whom to hire and whom not. (S. 334)

Unstrukturierte Interviews werden nach Klehe zudem so eingeschätzt, dass in ihnen anders als in standardisierten Verfahren Wertevermittlung und Selbstpräsentation des Unternehmens ihren Raum finden können. In dieser Hinsicht tragen verschiedene wissenschaftlich diskreditierte Verfahren symbolische Funktionen, die unter Umständen für die Beteiligten eine größere Tragweite haben als die nachweisbare Leistungsfähigkeit. Eine Legitimation durch wissenschaftliche Ergebnisse fällt dabei auch aus dem Grund schwer, dass die für die Erbringung dieser Ergebnisse zuständige Wissenschaft von vielen Entscheidern nicht als vollwertige Autorität angesehen wird:

> [...] even executive MBA students studying human resource management were reluctant to consider scientific utility information when deciding on the adoption of a personnel selection procedure. Since practitioners are neither dependent on nor controlled by human resource scientists, the pressure to conform to scientific standards is perceived as being low. (Klehe, 2004, S. 337)

Dabei sollte der Einfluss der Personalauswahl in der Organisation nicht unterschätzt werden:

> Es zeigt sich, dass die Personalauswahl nicht nur eine (Ersatz-)Beschaffungsfunktion hat, sondern auch die Unternehmung neu definiert. [...] Eine Stellenbesetzung ist immer gleichzeitig eine Maßnahme der Arbeitsgestaltung, der Organisations- und Personalentwicklung. (Schwarb, 1996, S. 29)

Eine steigende Beachtung der Eignungsdiagnostik, die etwa in dem Moment verstärkt werden könnte, in dem konkurrierende Firmen mit diagnostischen Prozeduren Erfolge verbuchen (Klehe, 2004), wäre dabei hinsichtlich einiger der beschriebenen Effekte folgenlos. Wottawa (1989) betont die steigende Bedeutung der Diagnostik, wenn organisationskulturelle Aspekte stärker in den Vordergrund rücken sollten. Bei einer solchen Umorientierung wäre allerdings unter Umständen zum einen zu erwarten, dass die Ökonomie ihre eigene diagnostische Kompetenz entweder „entdeckt“ oder tatsächlich entwickelt, und zum anderen, dass das Ausmaß der mikropolitischen Dynamik sich vergrößern würde. Eine weitere Konsequenz, die eine stärkere Institutionalisierung von Entscheidungsverfahren nach sich ziehen könnte, wird von Stinchcombe (1990) beleuchtet. Demnach unterliegen rationale Techniken in Organisationen sehr stark dem Einfluss von Moden - wie sie auch für den Bereich der Unternehmenskultur festgestellt werden können:

> The general point is that fashions in rational technique can spread faster than their value justifies, and value choices can be concealed in the choice of technique not understood by the people or organizations whose utilities are supposedly being maximized by rational means. (Stinchcombe, 1990, S. 310)

Als Grund für diese Annahme sieht er die Erwägung, dass die Übernahme scheinbar bewährter und institutionalisierter Techniken die eigenen Berechnungskosten ohne Qualitätsverlust reduziert. Eine solche Methode kann dabei leicht den wissenschaftlichen Kontrollprozess außer Kraft setzen. Schwarb (1996) weist zuletzt auf den Umstand hin, dass die Beziehung von Ressourcenmanagement und Personalmanagement durch das unpräzise Verständnis von Personal als Ressource verdeckt wird:

> In den Produktionsprozessen wird nicht die Ressource Personal eingesetzt, sondern der Ressourcenlieferant Personal liefert die Ressource Arbeit. Die Besonderheit dabei ist, dass der Ressourcenlieferant untrennbar mit der Ressource selbst gekoppelt ist [...]. (Schwarb, 1996, S. 28)

Für die Organisation ist deshalb das Personal nur in Bezug auf die Abgabe der Ressource Arbeit bedeutsam, nicht hinsichtlich seiner Qualitäten *an sich*. Diese Abgabe wiederum ist gekoppelt an die innere Befindlichkeit des Personals, die sich der direkten Kontrolle und Selektion entzieht. In der Konsequenz wird deshalb aber auch die Gestalt der Arbeit selbst unkontrollierbar:

> Die Personalauswahl kann bestenfalls Voraussetzungen schaffen, damit die erforderliche Ressource bereitgestellt wird. Je nach Ausgestaltung der Arbeitssituation kann jedoch ganz unterschiedliches Personal die gleiche benötigte Leistung erbringen. Aber auch die Mitarbeitenden selbst werden eine noch so präzis definierte Arbeit beeinflussen und gestalten. Das heißt, die Gestalt der Arbeit wird je nach Mitarbeitenden eine andere sein. (Schwarb, 1996, S. 28)

Diese potenzielle Plastizität kann im Standardvorgehen der Bestimmung von Anforderungsprofilen aber unter Umständen gar nicht berücksichtigt werden, das entstehende neue Problemlösungspotenzial kann unter Umständen nicht registriert werden, wenn die Personalauswahl nicht als Möglichkeit der Stellengestaltung, sondern nur im Hinblick auf eine Notwendigkeit der Stellenbesetzung betrachtet wird. Gleichzeitig muss unter Berücksichtigung von Lernpotenzialen auch die Kompensierbarkeit von Defiziten berücksichtigt werden. Selbst unter ökonomischen Gesichtspunkten ist nicht zu erwarten, dass bei der Erweiterung der Dimensionalität des Suchraums von reinen Prädiktorprofilen auf Prädiktorprofile plus verfügbare und auf die Profile abgestimmte Personalentwicklungsmaßnahmen die Reihenfolge der Bewerber konstant bliebe. Eine solche Entwicklungsmöglichkeit ließe sich weiterhin nicht durch eine zusätzliche Skala „Lernpotenzial“ zufriedenstellend erfassen. Durch die Etablierung von Regeln wandelt sich zudem mit andauernder Verwendung der Regeln die Gestalt der Organisation, so dass nicht zu erwarten ist, dass ein Regelsystem in seiner Leistungsfähigkeit für Entscheidungen konstant bleiben kann.

7.2.2.5 Gesellschaftliche Perspektiven auf Eignungsdiagnostik

Losgelöst vom engeren Kontext der Organisation können Auswahlverfahren auch auf gesellschaftlicher Ebene einer Kritik unterzogen werden. Fragen der Chancengerechtigkeit, der Rechte der Arbeitnehmer sowie Fragen nach Fairness und ethischer Akzeptanz der Verfahren werden zum Gegenstand der Diskussion. Feger (1991) sieht beispielsweise eine soziale Funktion des Messens:

> Unter der sozialen Funktion des Messens verstehe ich zweierlei. Zum einen konstituiert Messen in einer spezifischen Weise eine gemeinsame Welt. Zum anderen arrangiert jede Gesellschaft Messungen als Methoden, um Entscheidungen herbeizuführen. (S. 86)

Messungen sind eines der Werkzeuge, mit denen eine gemeinsame Umgebung sortiert und mit Bedeutung ausgelegt wird. Sie schaffen Vergleichsdimensionen und Orientierungen, allerdings immer auf Kosten alternativer Interpretationen und Muster. Für Grubitzsch (1991) ist der Wunsch nach Klassifikation die Reaktion auf eine undurchschaubare und chaotische und möglicherweise sozial gefährliche Situation. Diagnosen haben eine Ordnungsfunktion, mit der Störungen abgedeckt werden, die damit auch distanziert bleiben können. Die Diagnosemöglichkeit für das Bedrohliche ermöglicht die illusorische Befreiung auch von bedrohlichen Anteilen im gewöhnlichen Kern des Alltäglichen. Auf der positiven Seite ermöglicht eine Pluralisierung der Kategorien die Erfahrung eigener positiv konnotierter Außergewöhnlichkeit auf zumindest einigen der neuen Dimensionen bis hin zu den Extremformen von Rekordsammlungen und begleitenden Fernsehshows. Problematisch sind Sortierungen zum einen dann, wenn die auferlegte Ordnung

den Blick auf lösbare Probleme verstellt oder Verantwortlichkeit schafft, wo eine Kontrolle systematisch fehlt. Zum anderen werden Sortierungen und Klassifikationen dann suspekt, wenn ihr Zustandekommen auf einem Ungleichgewicht von Machtverhältnissen beruht:

> Der in der Moderne gestartete Versuch, Entscheidungen mit Hilfe der Wissenschaften zu optimieren, Fehlentscheidungen zu minimieren und den Menschen durch wissenschaftlich fundierte Aussagen „aufzuklären“, ist im 20. Jahrhundert problematisch geworden. Die Grenze zur Manipulation und Ideologie scheint gerade in der Beziehung von Wissenschaft zur Gesellschaft offensichtlich. (Karafyllis & Schmidt, 2002, S. 14)

In diesem Sinne steht gerade die Personalabteilung in einem Unternehmen unter Verdacht, Menschlichkeit und Manipulation gegeneinander auszuspielen, und nicht in jedem Unternehmen kann davon ausgegangen werden, dass dieser Verdacht zurückgewiesen werden muss. Stivers (2004) urteilt über die eingesetzten Techniken:

> Human relations techniques (like therapy) attempt to produce adjusted employees who, no matter how bleak the circumstances of their employment, retain a cheerful disposition and a positive outlook. Rational management techniques, however, serve the paramount purpose of management - to predict and control the future. In creating a closed system of statistical information, procedures, and decision rules, rational management magically assures us that we can control economic markets and employees despite the fact that a technological civilization is the most unpredictable of civilizations. [...] The obsession with rational control and prediction leads to an everincreasing irrationality. (S. 106)

Dabei fällt auf, dass in vielen Fällen nicht die diagnostischen Techniken selbst, sondern der Umgang mit den erhobenen Informationen zu einer gesellschaftlich relevanten Problematik führen, wie Wottawa und Thierau (2003) an einem drastischen Beispiel demonstriert:

> Wie umstritten die moralische Bewertung von Fortschritten in der wissenschaftlichen Erfassung von „evaluationsrelevanten“ Aspekten sein kann, zeigt sich zum Beispiel in den einfachen, billig gewordenen Möglichkeiten der Ultraschalldiagnostik, die es auch in einem Land wie Indien ermöglicht, das Geschlecht des Kindes schon in einer frühen Phase der Schwangerschaft zuverlässig zu bestimmen und dort zum gezielten Abbruch von Schwangerschaften mit weiblichen Föten führt. (S. 17)

„Selektion“ steht gerade in Deutschland unter dem Schatten der Praktiken und Euphemismen zu Zeiten des Nationalsozialismus, der auch die Diagnose von Behinderung heute noch berührt. Es ist leicht, darauf hinzuweisen, dass Daten kategoriell keine natürlichen Konsequenzen zeitigen können, auch wenn naive empiristische Wissenschaftsauffassungen in die Nähe dieses Irrglaubens führen können:

> Nach klassischer empiristischer Ansicht genügt es, in der Wissenschaft auf die Daten zu „hören“. Sie allein sagen uns, welche Theorien wir akzeptieren sollten. Man könnte das als ein weiteres Dogma des Empirismus bezeichnen, das ich das „Dogma der einfachen Rechtfertigung“ (DER) nennen werde. Es besagt, dass schon die Daten allein für oder gegen eine bestimmte Theorie sprechen bzw. weitergehend, dass es hier eine lineare Rechtfertigungsbeziehung von Daten zu Theorien gibt, so dass sich die Rechtfertigung von Theorien letztlich ganz auf bestimmte Daten zurückführen lassen muss. (Bartelborth, 2004, S. 13)

Geht der Zweck allerdings der Erhebung voraus, stellt die Bereitstellung von Methoden ein stillschweigendes Einverständnis dar, das vom einzelnen Diagnostiker zu verantworten ist. Gleichwohl bietet Diagnostik auch die Möglichkeiten zur subversiven Tätigkeit, wenn Darstellung von Methoden und die eigene Auffassung von Methoden auseinander klaffen. In diesem Sinne könnte eine negativ gewichtete Körpergröße, die als Minimierung der gefährdeten Hautoberfläche bei Arbeiten in Unfallzonen sicherlich den Frauenanteil unter den Eingestellten fördern. Ob sich dieses Vorgehen aber ethisch legitimieren lässt, ist äußerst fraglich. Westmeyer (2004) geht vom Ende der so genannten Krise der Diagnostik in den 70er Jahren des letzten Jahrhunderts aus. Er sieht aber auch die Grundlage der damals geäußerten Kritik als gegenstandslos, allerdings mit einem nicht unproblematischen Argument:

> Die von der Krisenrhetorik weniger beeindruckten Kolleginnen und Kollegen wiesen darauf hin, dass in jeder Gesellschaft, gleich welcher Form, in den verschiedensten Zusammenhängen Selektionsentscheidungen zu treffen wären. Selbst wenn auf den Einsatz psychologisch-diagnostischer Verfahren gänzlich verzichtet würde, änderte sich daran nichts. Es stellte sich allerdings die Frage, ob die dann auf andere Weise zu Stande gekommenen Entscheidungen besser fundiert und vertretbarer wären - gerade auch unter Berücksichtigung der Interessen der betroffenen Personen - als Entscheidungen, die sich auf psychologisch-diagnostische Verfahren stützten. Im Übrigen bestätigte sich auch in diesem Fall, was nach Auffassung der Wissenschaftstheorie für alle Wissenschaften gelten soll: Bestimmte Ideen, Theorien oder Methoden werden nicht schon deshalb aufgegeben, weil sie mit Fehlern behaftet sind, sondern erst dann, wenn es überzeugendere Alternativen gibt. (Westmeyer, 2004, S. 14 f.)

Diese Betrachtung verkennt die Dynamik, in der aus einer wissenschaftlichen Sorglosigkeit die Unbedenklichkeit des Einsatzes des zeitgenössischen Entwicklungsstandes gefolgert wird. Die Wahrscheinlichkeit der Entwicklung „überzeugenderer“ Theorien ist nicht unabhängig von der Akzeptanz der bestehenden. Eine schädliche Theorie ist auch dann aufzugeben, wenn es keine Alternative zu ihr gibt, gegebenenfalls ist die Praxis zu ändern, die den Rückgriff auf diese Theorie erforderlich macht. Status-Quo-Argumente sind in doppelter Weise gefährlich, zum einen entwerten sie auf rhetorischem Weg valide Kritikpunkte, zum anderen verlassen sie sich oft auf die *ungeprüfte* Hypothese, dass die Aufgabe einer Theorie zu einer Verschlechterung führen *muss* - zudem angemahnt werden kann, dass die mögliche Verschlechterung aus Sicht des Wissenschaftlers eine andere sein kann als jene aus der Sicht des Betroffenen[23].

Was diese Erwägungen aber deutlich machen, sind die Vorteilhaftigkeit einer breiten Basis und Sichtbarkeit für die Entwicklung von Zielen für Entscheidungen und Entscheidungskriterien, den Nutzen einer Explizitheit von Entscheidungsregeln und einer öffentlichen Rechtfertigung für diese und die positiven Konsequenzen einer Transparenz des eignungsdiagnostischen Prozesses. Diese Erwägungen sollen auf die Komplexität verweisen, die sich hinter der Etablierung einfacher Entscheidungsregeln verbergen kann. Die allgemeinere Beziehung zwischen Einfachheit und Komplexität soll im folgenden Abschnitt unersucht werden.

7.3 Über Einfachheit, Komplexität und die Welt hinter den Heuristiken

In diesem Abschnitt sollen die epistemologischen Ränder für den kognitionspsychologischen Ansatz der ABC-Gruppe überprüft werden. Es wird die Frage nach der epistemischen Verantwortbarkeit des Einsatzes einfacher Heuristiken in sozial bedeutungsvollen Entscheidungssituationen aufgeworfen, um dann das Verhältnis zwischen Einfachheit und Komplexität etwas allgemeiner zu diskutieren. Zuletzt soll untersucht werden, welchen Arten von Unsicherheit sich der Ansatz stellen kann, und wo sich Grenzen abzeichnen.

7.3.1 Einfache Heuristiken und epistemisch verantwortbare Entscheidungen

Es soll untersucht werden, inwiefern die Nutzung einfacher Heuristiken zur Bearbeitung des Ausgangsproblems legitimiert werden kann. Bishop (2000) nutzt als Beispiel für die eigenen Überlegungen eine diagnostische Fragestellung, anhand der eine solche Verbindung nachvollzogen werden kann:

> Imagine you're serving on an undergraduate admissions committee. There are various strategies you might adopt for predicting the future academic success of applicants. You might review all the available relevant evidence, GPA, test scores, letters of recommendation, quality of high school, etc., try to make the judgments that best fit your evidence and then spend hours wrangling with colleagues. Less onerous strategies are possible. For example, you might predict randomly, or on the basis of a single piece of evidence, like applicants' GPA's or test scores or the weight of their dossiers. Among the feasible reasoning strategies, which is the most epistemically responsible? (Bishop, 2000, S. 179)

Dabei soll davon ausgegangen werden, dass die beschriebene epistemische Verantwortbarkeit sich mit der Zielsetzung eines Diagnostikers decken sollte, aufgrund vorhandener Information zu bestmöglichen

[23] Zumal die Aufgabe der Theorie zunächst *allein* den Wissenschaftler schädigt, der sie vertreten hat.

Entscheidungen zu gelangen. Bishop argumentiert gegen die arbeitsintensive Strategie mit epistemologischen Argumenten, die hier analysiert werden sollen. Zunächst führt Bishop aus, was er für einen Entscheider bzw. Denker bedeutet, epistemisch verantwortlich zu handeln:

> (V) To be epistemically responsible is to display in one's reasoning the virtue (or virtues) epistemic internalists take to be central to warrant or justification, e.g., coherence, having good reasons, fitting the evidence. (Bishop, 2000, S. 180)

Beim Problem der Zulassung besteht eine Übernahme des Prinzips nach Bishop in der Maximierung der Chance für valide Vorhersagen („reliable judgments“[24]). In einer zusätzlichen Annahme - der *consilience assumption* - wird für epistemologische Verantwortlichkeit zudem eine spezielle Verbindung zur Wahrheit gefordert:

> The consilience assumption survives because except for the village skeptic, no one really believes that we live in a world in which consilience systematically fails. We believe that systematic consilience failures would require powerful evil demons, perfectly reliable clairvoyants, epistemically serendipitous brain lesions, and other creatures quirky and contrived. In our everyday lives, when pondering vacation plans, investment strategies or what to have for dinner most of us manage to reckon reliably and responsibly while putting aside worries about dreamers, demons and madmen. (Bishop, 2000, S. 181 f.)

Damit lässt sich die Annahme in folgender Weise formulieren:

> (C) In normal (non-skeptical) circumstances and in the long run being more epistemically responsible tends to lead to more reliable judgments. The more responsible a reasoning strategy, the more reliable it will tend to be, and the most responsible reasoning strategy will typically be the most reliable. (Bishop, 2000, S. 182)

Dabei geht Bishop davon aus, dass pragmatische Gründe eine maximale epistemische Verantwortlichkeit zu jedem Zeitpunkt verhindern müssen, zudem führen unlösbare oder unwichtige Probleme zu Einschränkungen. Zuweilen können Kosten hinsichtlich der Qualität der Wahl auch bewusst in Kauf genommen werden. Bishop verweist zunächst auf Studien zum *Bootstrapping* (Goldberg, 1970; Dawes, 1971) und zufälligen linearen Modellen (Dawes & Corrigan, 1974) und damit auf bereits geschilderte Phänomene wie die Leistungsfähigkeit der so genannten „improper linear models“:

> (FM) If a prediction problem is such that no proper model will be especially reliable and if the cues are reasonably predictive and somewhat redundant, then the reliability of a linear model's predictions will not be particularly sensitive to what weights are assigned to the cues (except for their sign). (Bishop, 2000, S. 187)

Gemäß der Ergebnisse dieser Studien scheint auf lange Sicht im Beispiel der Hochschulauswahl also eine Strategie der Summierung von Hochschulrang und Testergebnisrang eine arbeitsintensive Strategie in der Qualität der Ergebnisse zu übertreffen. Dieses kontraintuitive Fazit ist ein Beispiel für den Fall, dass eine scheinbar unverantwortliche Strategie der Informationsintegration die Ergebnisse der scheinbar verantwortlichen Strategie übertrifft. Daraus kann aber nicht gelernt werden, dass eine mechanische Prozedur Menschen in der Leistung übertrifft, da die „unangemessenen“ Modelle von geschickten Menschen ersonnen wurden. Diese Modelle bezeichnet Bishop auch als Mitglieder der Klasse der unvernünftigen, aber sehr reliablen Strategien (IRS):

> [...] for problems of social prediction, *we can increase our reliability by rejecting labor intensive strategies that intuitively strike us as epistemically responsible and using instead simple but intuitively improper methods of integrating the information we have.* The unit weight model is an irresponsible but very reliable strategy (or IRS) for making social predictions. (Bishop, 2000, S. 188)

[24] Bishop spricht von „reliable judgments“, was sich aber im Kontext als „verlässlich in Bezug auf die Zielsetzung“ und damit als „valide“ interpretieren lässt.

Für die Klasse der so definierten *IRS* sind für Bishop folgende vier Eigenschaften besonders bemerkenswert:

- Eine IRS hat relativ transparente Bedingungen für eine angemessene Anwendung, die relativ einfach entdeckt werden können.
- Eine IRS ist für einen durchschnittlich intelligenten Menschen relativ einfach anzuwenden ohne technische Hilfsmittel oder eine statistische Ausbildung.
- Bei angemessener Anwendung erbringt die Strategie sehr gute Ergebnisse, die hinter den Ergebnissen unzugänglicher[25] idealer Strategien nicht weit zurückstehen und besser sind als die Ergebnisse von um Verantwortbarkeit bemühten Menschen.
- Die Anwendung einer IRS verletzt die Intuition und das Gespür für das, was es heißt, epistemisch verantwortlich zu sein, so dass eine Tendenz zur Ablehnung der IRS besteht.

Für Bishop (2000) klassifizieren sich die Rekognitionsheuristik und TTB als Beispiele für *IRS*, bspw. aufgrund des kontraintuitiven *Less is More*-Effekts. Bishop beschreibt eine Argumentation, in der ein Konflikt mit einer internalistischen Konzeption epistemischer Rationalität vermieden werden könnte: wenn der Entscheider nichts von den IRS-Ergebnissen weiß und dafür nicht verantwortlich gemacht werden kann, dann handelt er verantwortlich. Wenn er die Ergebnisse hingegen kennt, dann ist er genau dann epistemisch irrational, wenn er die IRS nicht anwendet. Das Problem an dieser Argumentation ist die Notwendigkeit, eine IRS als kompatibel mit den Tugenden epistemischer Verantwortlichkeit einzustufen. Insbesondere in Fällen, in denen *kein* qualitativ äquivalentes klassisches Modell zur Verfügung steht, entsteht ein Dilemma. Bishop vergleicht die Situation mit dem Lernvorgang, der durch die Entdeckung von Placebo-Effekten einsetzen musste und in dem Doppelblind-Versuche zur bevorzugten Methode avancierten:

> After the IRS results, in making social predictions, responsible cognizers prefer conclusions derived from well-constructed unit weight models to those of human experts. But as I have tried to suggest, this analogy radically underestimates the havoc the IRS results will bring to our epistemic practices. (Bishop, 2000, S. 203)

Wenn nicht (C) als Annahme aufgegeben wird, kann die Anwendung einer IRS nicht als epistemisch verantwortlicher eingestuft werden als die Anwendung einer arbeitsintensiven Strategie. In der Praxis führt diese Bedingung in vielen Fällen zu einer Abwehrhaltung gegen die vorgeschlagenen Strategien, wie Bishop am Beispiel des Festhaltens an unstrukturierten Interviews argumentiert: Trotz bekannter wissenschaftlicher Befunde gegen den prädiktiven Nutzen dieser Verfahren werden sie aufgrund des Anscheins der Erbringung entscheidungsrelevanter Information beibehalten:

> For the sake of consilience, we must smother our inclination to attend to good reasons, to attend to fit with evidence, to attend to coherence. This is a nontrivial task. [...] If it's fiercely difficult for us to relinquish just one kind of evidence (unstructured interviews), imagine how difficult it will be for us to relinquish (for specific kinds of reasoning problems) deeply held epistemic values, e.g., good reasons, fit with evidence, coherence, etc. (Bishop, 2000, S. 203)

Bishop betrachtet hingegen das Ignorieren des größten Teils verfügbarer Informationen bei begrenzten Ressourcen als epistemisch optimales Verhalten, ebenso der Verzicht auf individuelle Gewichtungen bei sozialen Vorhersagen [26]: Als Konsequenz aus den Ergebnissen sieht Bishop deshalb die Notwendigkeit, (V) aufzugeben, da (V) und (C) nicht kompatibel erscheinen, (C) aber nicht verworfen werden soll und argumentiert damit für eine reliabilistische Betrachtung von epistemischer Verantwortlichkeit, ähnlich wie Nozick eine reliabilistische Perspektive auf Rationalität entwirft (Nozick, 1993):

[25] Strategien sind dann unzugänglich, wenn sie von einem zeitlich und kognitiv begrenzten Denker nicht ausgeführt werden können: Ein Beispiel wäre die vollständige Analyse von kombinatorisch explodierten Suchräumen.

[26] Bishop qualifiziert diese Aussagen allerdings nicht hinsichtlich der untersuchten Stichprobengrößen und der Stabilität der betrachteten sozialen Strukturen.

> My own suspicion is that the IRS results will drive us toward a generally reliabilist view of epistemic responsibility, especially if more IRS's are discovered; that is, the more we are forced to quash (V) in our general epistemic prescriptions and in our judgments of particular episodes, the more likely we are to give up (V) as part of our conception of what it is to be epistemically responsible. If this happens, epistemic responsibility would be exhausted by consilience. (Bishop, 2000, S. 205)

Bishop weist zudem darauf hin, dass die Anwendung von IRS-Methoden keine hypothetische Praxis darstellt:

> When you apply for credit and when newborns are identified as at risk for Sudden Infant Death syndrome, it's almost surely an IRS, and probably a unit weight model, that's doing the predictive work (Lovie and Lovie 1986). It's not just that IRS's are coming. They're here, and they're here to stay. (Bishop, 2000, S. 205)

Diese Analyse kann zunächst einmal in Bezug auf das Konzept ökologischer Rationalität interpretiert werden: durch eine exklusive Gewichtung des kriterienorientierten Erfolges von Strategien ist die reliabilistische Strategie in diesem Ansatz die Ausgangssituation. Das Beispiel könnte aus dieser Betrachtungsrichtung lediglich als Hinweis darauf gewertet werden, dass auch klassische Strategien hinsichtlich ihrer ökologischen Rationalität untersucht werden können und dass normative Empfehlungen aus Erwägungen der Optimalität hinsichtlich einer vorgegebenen Menge von Strategien für einen Anwendungsbereich erwachsen. Die vorgeschlagenen in (V) enthaltenen Prinzipien implizieren dabei größtenteils Normen, die mit dem ABC-Ansatz nicht kompatibel sind oder abgelehnt werden. Die von Bishop als *IRS*-Ergebnisse genannten Ergebnisse sind in der ABC-Perspektive nicht als kontraintuitiv zu bezeichnen, weil die Präferenz für *IRS* auch über entsprechende Strukturen von Umgebungen gerechtfertigt werden können, wobei die Form der Rechtfertigung auch hier als reliabilistisch bezeichnet werden kann. In diesem Fall wird die vorgeschlagene Konsequenz trivial, wie Bishop und Trout (2002) herausstellen:

> Given a choice between two reasoning strategies where one of them is known to be less reliable and harder to use, to insist that the responsible reasoner adopt that strategy is to insist upon epistemic masochism, not epistemic responsibility. (Bishop & Trout, 2002, S. S206)

Was allerdings in dieser Perspektive vollständig ausgeblendet wird, sind Sekundärfolgen des Wechsels eines Auswahlverfahrens wie die Einführung spezieller politisch oder gesellschaftlich relevanter Verzerrungen, Veränderungen der Umgebungen durch Veränderung von Zulassungsbedingungen, etc. In jedem Fall scheint es zur Verantwortlichkeit in einem volleren Sinne zu gehören, alternativ mögliche Strategien untereinander zu vergleichen, mögliche Konsequenzen zu antizipieren und im Verlauf der Zeit die Angemessenheit der Entscheidungsregeln kontinuierlich zu überprüfen - mit einem möglichen Dilemma bei interpersonell variierenden Intuitionen und Bewertungen. Dabei kann ebenfalls nichts gegen die Vorstellung sprechen, dass eine IRS bis zu einer bestimmten Beobachtungszahl von Fällen die Methode der Wahl darstellt, um ab einem bestimmten Punkt durch ein Modell abgelöst zu werden, dass in Bishops Klassifikation keine IRS darstellt, deren Vorteilhaftigkeit aber unter Berücksichtigung des möglichen *Overfitting* aber anhand der gewachsenen Stichprobe demonstriert werden kann.

Bishop und Trout (2002) laufen insbesondere Gefahr, die für die statistische Bearbeitung notwendigen Kodierungen zu ignorieren und die Ergebnisse methodologisch naiv direkt in die untersuchten sozialen Bereiche zu projizieren, was als arrogant und unsachgemäß zurückgewiesen werden könnte:

> This focus on outcomes means that, without relying on outcome information in such domains as psychotherapy, oncology, the psychology of criminal behavior, etc., „expert“ claims originating in subjective evaluation can be safely ignored for what they are: sentimental autobiography. We cannot begin to repair the damage done by our indulgence of these internalist conceits, conceits that have persisted beyond the decades that exposed them. Incorrectly diagnosed cancers, dangerous criminals released, innocent people put to death, needless neglect of progressive brain disease, the misidentification of psychotics [...] these failures demand frank admission. (Bishop & Trout, 2002, S 206 f.)

Weder Kriterien noch Informationen werden auf ihre Genese und Vertrauenswürdigkeit hinterfragt, noch wird die Verbindung zwischen statistischen Empfehlungen und praktischen Konsequenzen in Frage gestellt, so dass eine medizinische Krebsdiagnose auf die gleiche Stufe gestellt wird wie das Ergebnis eines juristischen Prozesses oder einer psychotherapeutischen Dynamik. In dem, was die Autoren als sentimentale Autobiographien karrikieren, mögen sich in der Tat übertriebene Selbstbilder und Kompetenzüberzeugungen verbergen, aber mit Sicherheit ruht hier auch ein großer Teil dessen, was sich einer outputbasierten Analyse entzieht, aber einen breiten Bereich dessen berührt, was Menschen und Menschlichkeit ausmacht.

Die Autoren machen hier den Fehler, nicht zu berücksichtigen, dass Methoden in praktischen Anwendungen einen Einfluss auf Datenerhebung und Datenaufbereitung haben, unter Umständen sogar auf die Auswahl von Handlungsmöglichkeiten, die aufgrund der Daten zur Verfügung stehen. Die vorgestellten Argumente können genutzt werden, um zwischen Methoden zu unterscheiden, die eine identische Datenerhebung und -aufbereitung verlangen, aber nicht zu einer Rechtfertigung dieser Praxis ohne eine Berücksichtigung von Alternativen, die mit einer anderen Datenbasis operieren oder eine andere Form der Behandlung an eine Diagnose knüpfen. Ansonsten entstünde der Eindruck einer Eindeutigkeit von Informationsbasis und Handlungsraum, die *gerade* bei sozialen Vorhersagen nur aus einer positivistischen Perspektive leidlich akzeptabel erscheinen mag.

Für die computerunterstützte Diagnostik fallen allerdings Einschränkungen hinsichtlich der Komplexität von Auswertungen fort, solange die Korrektheit der Berechnungssoftware sichergestellt werden kann. Solange die Berechnung von Computern in für den Entscheidungsprozess akzeptabler Zeit bewerkstelligt werden kann, spielen kleinere Laufzeitdifferenzen selten eine Rolle. Relevant hingegen ist weiterhin die Frage der Interpretierbarkeit der Ergebnisse, aber auch die Frage der zu erbringenden Kosten für die Informationsbasis. Die Aufgabe des Diagnostikers bei der Suche nach Entscheidungsregeln kann in bestimmter Hinsicht nicht als eine Optimierung betrachtet werden: Nicht alle denkbaren Entscheidungsregeln, nicht alle denkbaren Information und nicht alle denkbaren Entscheidungsempfehlungen können berücksichtigt werden, die Einflussmöglichkeiten auf diesen Prozess sind begrenzt, Kriterien möglicherweise diffus, unbekannt oder umstritten. Einfache Heuristiken bieten sich als Alternativen zu klassischen Verfahren an, müssen aber auf ihre Vorteilhaftigkeit im Anwendungsfall hin überprüft werden, wenn nicht Entscheidungen getroffen werden sollen, die mit einem realistischen Aufwand *merklich* verbessert werden können.

Zudem gehört es zu den zentralen menschlichen Kompetenzen, Ausnahmebedingungen für Regelmäßigkeiten zu identifizieren. Auch wenn die Studien zur differenziellen Kompetenz von statistischen Modellen und Experten zu einer Vorsicht vor Selbstüberschätzung anraten, würde kaum einem Menschen zu raten sein, Geld auf einen Sportler zu setzen, der sich beim Warmlaufen das Bein bricht, egal ob dieses Merkmal als Prädiktor in ein Modell aufgenommen wurde oder nicht. Over (2000) weist auf den Umstand hin, dass Borges et al. (1999) in der Entscheidung, Geld auf die Vorhersagen einer gruppenbasierten Rekognitionsheuristik zu setzen, *gegen* die Vorhersagen der eigenen Rekognitionsleistungen handeln mussten:

> [...] they were sophisticated enough to discount their own recognition abilities and to invest in the portfolio for the category least familiar with stock markets: Munich pedestrians. In fact, they must have used higher-level cognition to override the output of their recognition heuristics, and nothing in the book even tries to describe such a process. (Over, 2000, S. 189)

Epistemische Verantwortlichkeit kann in vergleichbarer Weise nicht ohne die Nutzung ähnlicher höherstufiger Prozesse zugeschrieben werden, auch diese Frage entkommt damit nicht den Ausläufern des *Frame*-Problems.

7.3.2 Einfachheit und Komplexität

Für Harman (2004) gehört Einfachheit zu den Komponenten, die in das menschliche System des Denkens eingebaut sind. Das Streben nach Einfachheit wird damit zu einer Basiseigenschaft des Menschen und kann nicht abgeleitet werden aus der Annahme, dass simplere Hypothesen wahrscheinlicher sind. Einfachheit stellt damit für Harman einen Bias menschlicher Lernverfahren dar, er stimmt aber mit Forster darin überein, dass kein Lernverfahren ohne Bias auskommt:

> Our initial ordering of hypotheses might be based on how easy it is to use a hypothesis to answer questions in which we are interested. This would be one sort of simplicity ordering, where „simple“ means simple to use. (Harman, 2004, S. 52)

Für Harman besteht ein Trade-Off zwischen Einfachheit und Genauigkeit, eine Argumentation für eine höhere objektive Wahrscheinlichkeit einfacherer Hypothesen findet er nicht plausibel:

> Perhaps we can say that simpler theories are more likely to be true than more complicated theories that account for the same data. [...] On the other hand, there seems to be no direct noncircular argument that simpler theories are more likely to be true in the sense of objective likelihood [...]. (Harman, 2004, S. 53)

Für Harman muss damit die Suche nach Einfachheit nicht aus praktischen Gründen gerechtfertigt werden, statt dessen gehört Einfachheit für ihn zu den Prinzipien, die den induktiven Bias menschlichen Lernens formen, die aber zum Erreichen bestimmter Ergebnisse fundamental notwendig sind:

> [...] a system without inductive bias in favour of simpler hypotheses [...] would not be able to reach conclusions that ought to be reached. The point is that it would be deeply irrational to reason in a way that leads to scepticism. Any sceptical system simply gives the wrong results about what one should believe. (Harman, 2004, S. 55)

Gegen den Vorwurf der Zirkularität argumentiert er, dass Denken nicht mit der Produktion logischer Argumente gleichgesetzt werden kann:

> Reasoning is reasoned change in view that improves coherence in a way that helps to answer questions in which you are interested. (Harman, 2004, S. 55)

In diesem Sinne kann Einfachheit nützlich sein, damit interessante Fragen beantwortet werden können, damit kann aber auch Simons Diagnose der „Komplexität der Welt“ relativiert werden. Wilson (1973) weist auf den Umstand hin, dass Simon im Unterschied zu den meisten Menschen nicht mit Einfachheit, sondern mit Komplexität beginnt:

> Anyone can point out how „complex“ the real world is by post hoc reference to factors that might have been considered prior to decision and action. From the standpoint of the scientific observer, whose ideals real people never quite measure up to, it may even make sense to characterize administrative man's view of this „buzzing, blooming confusion“ as „a drastically simplified model“ of what we must take on faith from Simon constitutes the „real“ world. But I would have put most of this the other way round. (Wilson, 1973, S. 279)

Nach Tversky und Simonson (1993) können Fehler nicht durch eine unzulässige Simplifizierung, sondern durchaus auch gerade durch die Suche nach Komplexität entstehen, indem unnötige Details gesammelt und eine Entscheidung verschleppt wird.

Eine Variante des Komplexitätserhaltungssatzes von Forster für Theorien des Sozialverhaltens beschreibt Weick (1979) in Bezugnahme auf Thorngates Postulat der kommensurablen Komplexität:

> [...] this postulate states that it is impossible for a theory of social behavior to be simultaneously general, accurate, and simple. The more general a simple theory is, for example, the less accurate it will be in predicting specifics. (Weick, 1979, S. 35)

Damit könnten die Abhängigkeiten auch als Uhr dargestellt werden, die den notwendigen Trade-Off verdeutlicht (vgl. Abbildung 7.6): Auf dem Ziffernblatt sind die Attribute der Einfachheit, der Allgemeinheit und der Genauigkeit angeordnet. Jede Theorie kann als auf der Kreislinie positioniert konzipiert werden, d.h. keine Theorie erreicht alle drei Eigenschaften gleichzeitig, ein Versuch erscheint zudem wenig wünschenswert:

> Failure to accept the inevitable tradeoffs implied in the GAS formulation seems to be at the heart of many current research problems. Investigators act as if they can simultaneously accomplish all three aims in their explanations, and that delusion is at the heart of much trivial, inconclusive research. The solution would seem to be either robust compromises or alternation rather than attempts to accomplish all three. (Weick, 1979, S. 36)

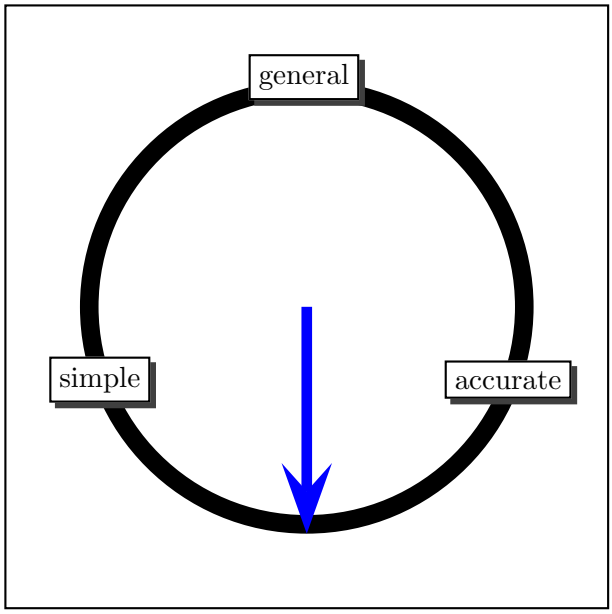


Abbildung 7.6: Die Veranschaulichung von Thorngates Postulat der kommensurablen Komplexität nach Weick (1979). Jede Art der Theorie liegt auf der Kreislinie positioniert.

Der Ansatz der Evolutionspsychologie und auch der ABC-Ansatz versuchen, über das Zugeständnis von Bereichsspezifität einfach und genau zu bleiben. Der Versuch, nun wieder eine allgemeine Theorie erhalten zu können, indem eine Systematik der Bereiche oder eine Verwaltung der Module konzeptualisiert werden, erweist sich dementsprechend als Stolperstein, der einen Schatten auf die Leistungen des Ansatzes wirft. In diesem Sinne lässt sich auch der Hinweis von Tversky und Simonson (1993) verstehen, dass simple Strategien manchmal komplexere deskriptive Modelle benötigen, da der Kontext mit modelliert werden muss. Annahmen des bereichsübergreifenden Problemlösens werden andererseits ungenau, wenn sie zu einfach konzipiert werden und unhandlich und überkomplex, wenn sie genau werden sollen. Es spricht bei näherer Betrachtung der Verhältnisse in sehr unterschiedlichen Forschungsfeldern einiges für das Postulat, auch wenn es sicherlich keine präzisen Voraussagen ableiten lässt. Freitas (2002) betont die ursprüngliche Formulierung der vielzitierten Metapher von *Occam's razor*, die ein Entscheidungskriterium bieten soll, um den Entscheider gegen unnötige Komplexität voreinzunehmen:

> Occam's razor is a philosophical principle proposed by William of Occam in the 14th century. Translated into English, it essentially says that: „*Entities should not be multiplied without necessity*“ or „Plurality should not be assumed without necessity“[...]. (Freitas, 2002, S. 38)

In Bezugnahme auf Domingos (1998) diskutiert Freitas zwei Varianten der „Klinge“, eine „scharfe“ und eine „stumpfe“ Variante:

- Die „stumpfe Klinge“: Gegeben zwei Modelle mit gleichem Fehler im Trainingsset, sollte das einfachere präferiert werden, da es mit höherer Wahrscheinlichkeit einen kleineren Fehler in einem Testdatensatz aufweisen wird.
- Die „scharfe Klinge“: Gegeben zwei Modelle mit dem gleichen Fehler in der Generalisierung auf neue Datensätze, sollte das einfachere präferiert werden, weil Einfachheit an sich wünschenswert ist.

Dabei entspricht ein Argument in Form der stumpfen Klinge einer der Begründungen des ABC-Programms für die Robustheit von einfachen Heuristiken. Allerdings ist dieses Argument nach Freitas nicht valide:

> Among all rule sets equally consistent with training data, all are equally likely to maximize predictive accuracy on an unseen test set. Whether the simplest or most complex rule set will really maximize predictive accuracy on an unseen test set essentially depends on whether the sought relationships in the data being mined are simple or complex. In other words, choosing a rule set because it is simpler than other rule sets that are equally consistent with the data is a form of inductive bias, i.e., a bias favoring simpler rules. (Freitas, 2002, S. 38 f.)

Die scharfe Version ist hingegen für Freitas akzeptabel, da die Einfachheit einer Regelmenge zu einfacherer Interpretierbarkeit und fehlerärmerer Anwendung führen kann. Ein Prinzip, das einen impliziten Trade-Off zwischen Einfachheit und Genauigkeit explizieren soll, ist das Prinzip der *Minimum Description Length* (MDL) (vgl. Quinlan & Rivest, 1989). Freitas (2002) stellt das Prinzip in folgender Form vor:

> Given a set of competing hypotheses (e.g. rule sets) and a data set, this principle recommends that one choose, as the „best“ hypothesis, the one that minimizes the sum of two terms, namely:
>
> 1. the length of the hypothesis; and
> 2. the length of the data given the hypothesis, i.e., the length of the data when encoded using the hypothesis as a predictor for the data.
>
> The second term represents the length of the encoding of the data instances that are „exceptions“ to the hypothesis. Both terms are measured in bits of information, which, one could argue, is a nice chararacteristic of the MDL principle. (Freitas, 2002, S. 39)

Eine komplexere Hypothese braucht damit eine längere Beschreibung als eine einfache und ist der einfacheren damit bei gleicher Genauigkeit unterlegen. Gleichzeitig haben eine komplexe Hypothese, die jeden Datenpunkt einzeln vorhersagt und eine Variante ohne Hypothese, die die Abweichung jedes einzelnen Datenpunkts bestimmen muss, die gleiche Länge, was intuitiv sinnvoll erscheint. Auf diese Weise gehen Freiheitsgrade von Modellen in scheinbar „natürlicher“ Weise in die Beschreibung ein:

> Apparently, it has the advantage that two incommensurable criteria for evaluating the quality of a hypothesis, its simplicity and its classification accuracy, are made „commensurable“ by measuring them in terms of the number of bits needed to encode the hypothesis and the exceptions to the hypothesis. (Freitas, 2002, S. 39)

Ein Problem ergibt sich in der Wahl des Enkodierschemas. Wenn komplexere Hypothesen eine höhere Wahrscheinlichkeit haben, können sie kürzer kodiert werden als einfache, sind deshalb aber noch lange nicht prädiktiver. Wird getreu dem ursprünglichen Hypothesenraum kodiert, gibt es hingegen einen Bias für Einfachheit. Um einen Entscheidungsbaum zu kodieren, könnte zum einen jeder Knoten einzeln kodiert werden, zum anderen könnte als Ausgangsbasis ein voller Baum genommen werden, um jeden Schnitt einzeln zu kodieren. Die Beziehung zwischen Komplexität und Kodierungslänge wäre für die beiden Enkodierungsschemata entgegengesetzt. Eine „natürliche Kodierung“ wird in den seltensten Fällen zu finden sein.

Eine ähnliche Frage wirft die Überlegung der Architekturabhängigkeit von Einfachheitskriterien auf. Kognitiven Architekturen wird zuweilen der Status von Forschungsprogrammen zugesprochen (Cooper, 2006), die Annahme einer Architektur hat weitreichende Konsequenzen für den Vergleich zwischen kognitiven Leistungen. Dementsprechend sieht Oaksford (2000) bspw. die Notwendigkeit, die Erfolgsmaße von Take-The-Best zu relativieren:

> Take-The-Best only has a clear advantage over other algorithms if cognitive processes are assumed to be serial. Given the extensive research programs aimed at establishing the viability of instance-based and conectionist architectures as general accounts of cognitive architecture [...], it seems that considerable caution must be attached to a measure of speed that presupposes a serial architecture. (Oaksford, 2000, S. 760)

Neben der generellen Frage der Gewichtung von Geschwindigkeit hält er damit die Abhängigkeit dieses Maßes von der Architektur fest und sieht schnelle, parallele und integrative Ansätze als Konkurrenten. Gigerenzer (2004a) räumt ein, dass die Einfachheit einer Heuristik von den evolvierten oder gelernten Kapazitäten des Organismus abhängig ist:

> A heuristic is *simple* relative to the evolved or learned capacities of an organism. (Gigerenzer, 2004a, S. 63)

Dementsprechend könnte auch diese Beziehung als Beispiel für den Komplexitätserhaltungssatz Forsters herangezogen werden. Damit eine Heuristik einfach ausgeführt werden kann, ist unter Umständen eine komplexe zugrunde liegende Architektur notwendig. Gleichzeitig werden Maße der Einfachheit notwendigerweise speziesspezifisch:

> For humans acting consciously, a weighted additive calculation is much harder than tallying (weights all unity), but for an insect specialised on a specific task, evolving can have hardwired the weights [...]; then the insect needs simply tally these ready-weighted cues, yet it achieves what to us looks like a harder weighted-additive calculation [...]. (Hutchinson & Gigerenzer, 2005b, S. 107 f.)

Hutchinson und Gigerenzer (2005b) sehen ein weiteres Problem in der Notwendigkeit der Trennung zwischen einfacher Konstruierbarkeit und einfacher Beschreibbarkeit, da nur über Plausibilitätsargumente eingeschätzt werden kann, welche Algorithmen für einen Organismus schwierig zu entwickeln sind, was beinahe unabhängig von der Leichtigkeit der Beschreibung der Ergebnisse sein kann: Als Beispiel kann eine perfekt lineare Beziehung leicht beschrieben aber äußerst schwierig natürlich erzeugt werden. Es sollte nicht unterschätzt werden, dass bei der Einschätzung von Einfachheit im Sinne einer Beschreibung gerade jene kognitiven Mechanismen beteiligt sein können, deren Einschätzung unternommen wird. Es spricht deshalb einiges dafür, dass Menschen gerade jene Aufgaben oder Leistungen als einfach klassifizieren, in Bezug auf welche sie eine höhere strukturelle Komplexität aufweisen, die den Umgang mit diesen Aufgaben einfach *macht*.

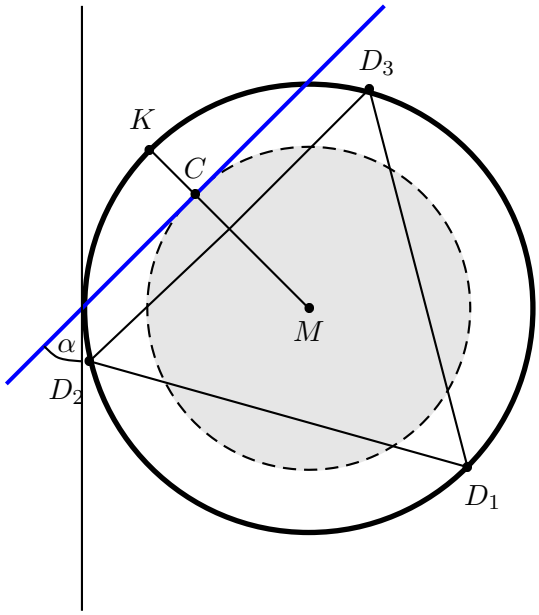


Abbildung 7.7: Das Problem von Bertrand (1907) nach Forster (1999): Der zufällige Wurf eines Strohhalms in einen Kreis kann modelliert werden als (a) zufällige Wahl der Länge der Strecke $\overline{MC}$ mit $0 \leq l(\overline{MC}) \leq l(\overline{MK})$, (b) zufällige Wahl des Winkels α zwischen Strohhalm und Tangente an einem Schnittpunkt des Strohhalms mit dem Kreis, (c) als zufällige Wahl der Fläche des Kreises mit dem Radius $\overline{MC}$, also zufälliger Wahl von $\pi \cdot l(\overline{MC})^2$; die Wahrscheinlichkeit p, dass die Länge des Strohhalms innerhalb des Kreises die Länge einer Seite des gleichseitigen Dreiecks $\Delta D_1 D_2 D_3$ mit Eckpunkten auf der Kreislinie übersteigt, ergibt sich als (a) $p = \frac{1}{2}$, (b) $p = \frac{1}{3}$ und (c) $p = \frac{1}{4}$.

Zuletzt soll an einem Beispiel demonstriert werden, wie sich die Komplexität von Voraussetzungen verstecken kann hinter einer scheinbaren Voraussetzungsfreiheit. Die Problematik eines Vorgehens, das die inhaltliche Form und Substanz eines Problems als sekundäre Frage auffasst und sie der formalen Lösung des Problems unterordnet, kann mit einer Aufgabe illustriert werden, die am Anfang des letzten Jahrhunderts von Bertrand betrachtet wurde (Bertrand, 1907) und hier in der Version von Forster (1999) betrachtet werden soll. Die Aufgabe lässt sich in folgender Weise formulieren:

Ein Strohhalm wird zufällig in einen Kreis geworfen. Wie groß ist die Wahrscheinlichkeit unter der Annahme, dass der Strohhalm den Kreis an zwei Punkten schneidet, dass die Strecke zwischen diesen bei-

den Punkten länger ist als die Seitenlänge des gleichseitigen Dreiecks, dessen Ecken auf dem Kreisumfang liegen?

Das Problem ist in Abbildung 7.7 illustriert. Das zunächst kontraintuitive Ergebnis des Versuches, die Frage zu beantworten, ist die Feststellung, dass die Frage nicht eindeutig beantwortbar ist, wenn nicht zusätzliche Annahmen über den Werfenden getroffen werden. Insbesondere bieten sich drei Lesarten des „zufälligen Werfens“ an, die alle drei formalisierbar sind und mit einem Begriff von Gleichwahrscheinlichkeit operieren:

> The purpose of Bertrand's example is to show that Laplace's principle of indifference does not (by itself) solve this problem because there are many ways of defining equally possible situations. Three of these are assinging a uniform probability distribution to: (a) the distance between the center of the chord and the center of the circle; (b) the angle of intersection of the chord on the circumference [...]; or (c) the area of the interior circle touching the center of the chord [...]. These assignments respectively lead to answers of 1/2, 1/3, and 1/4 to Bertrand's question. (Forster, 1999, S. 557)

Die drei Lesarten sind in der Abbildung illustriert. Nicht alle Antworten werden als in gleicher Weise ökologisch valide bei Annahme eines menschlichen Werfers betrachtet, aber letztendlich müsste eine genaue Analyse dessen, was es heißt einen Strohhalm zufällig zu werfen, der Beantwortung der Frage vorausgehen. Das formale Instrumentarium kann nur Alternativen anbieten, nicht aber bei der Auswahl zwischen den Alternativen unterstützen, letztendlich sind alle drei Lösungen korrekt, auch wenn unterschiedliche implizite Annahmen gemacht werden (Michalewicz & Fogel, 2000).

Wie soll jetzt aber die Antwort einer Versuchsperson bewertet werden, die nach einer Stunde Bearbeitungszeit und dem Ausfüllen von fünf Notizblockseiten die Antwort „zwei Fünftel“ gibt? Da aus der Aufgabenstellung kein absoluter Maßstab für richtig oder falsch gewonnen werden kann und die drei betrachteten Lösungswege nicht die einzig möglichen sein müssen, ist die Bewertung der Antwort ohne ein Studium der Notizen und die Diskussion des gewählten Lösungsansatzes kaum möglich[27]. Die Norm, mit deren Hilfe sich die Antwort bewerten ließe, wäre in diesem Fall nicht einmal unabhängig von der Versuchsperson, denn mit der Setzung der unterspezifizierten Prämissen durch diese verbinden sich unterschiedliche Bewertungsmaßstäbe[28]. Weiterhin besteht bei dieser Aufgabe die Möglichkeit, dass die Antwort auf die Frage zum Zeitpunkt ihrer Formulierung noch nicht feststeht. Wird zum Beispiel eine konkrete Person als Werfer designiert, um die Frage durch eine ausreichende Anzahl wiederholter Würfe empirisch zu beantworten, so müsste das Ergebnis dieses Experiments abgewartet werden, bevor eine auf formalem Weg gewonnene Lösung bewertet werden kann. In diesem Fall wäre immerhin die Anwendung eines frequentistischen Begriffes der Wahrscheinlichkeit möglich, während die Sinnhaftigkeit des Experiments fraglich bleiben würde.

Hinter einer scheinbar einfachen Aufgabe liegt ein Interpretationsprozess versteckt, der im Moment des Verständnisses der Aufgabenstellung bereits abgeschlossen ist, damit aber bereits in eine von mehreren konkurrierenden Hypothesen investiert hat. Die formale Struktur dieses Problems ermöglicht die Aufdeckung dieser Ambiguität, in anderen Problemen kann die Sachlage durchaus verwickelter sein.

7.3.3 Hinter den Heuristiken

Gigerenzer (2004a) vergleicht den Forschungsansatz der ABC-Gruppe mit der Exploration einer neuen, unvertrauten Landmasse:

> The new land of rationality we set out to explore is shrouded in a dim mist of uncertainty, populated with people who have limited time and knowledge, but with smart heuristics at their disposal. Welcome, and I hope you feel at home in this world. (Gigerenzer, 2004a, S. 83)

Hinter den trügerischen Fassaden einer lediglich lokalen Unsicherheit ruht weiterhin die Möglichkeit der niemals zu erlangenden Gewissheit, die sich als ontologische Unsicherheit von einer rein epistemischen

[27] Allerdings könnten die Antworten „Null.“ oder „Eins.“ zumindest als implausibel ausgewiesen werden, da beide Antworten ein Werferverhalten implizieren, das mit den Angaben der Aufgabenstellung nur mit großem Aufwand in Deckung zu bringen ist.

[28] In jedem Fall wäre die Frustration einer Versuchsperson gerechtfertigt, die sie bei der Entgegennahme der schlichten Antwort „Falsch, richtig ist ein Viertel“ empfinden würde.

trennen lässt. Für Russell (1912/1967) ist Unsicherheit und menschliche Fallibilität eine unumstößliche Grenze für philosophische Bestrebungen:

> Philosophy may claim justly that it diminishes the risk of error, and that in some cases it renders the risk of error so small as to be practically negligible. To do more than this is not possible in a world where mistakes must occur; and more than this no prudent advocate of philosophy would claim to have performed. (Russell, 1912/1967, S. 88)

Bereits bei Cusanus findet sich die Akzeptanz unverrückbarer Grenzen der Erkenntnis. Er vergleicht das Bemühen des (endlichen) Verstandes um die (unendliche und bei Cusanus auch göttliche) Wahrheit mit dem ebenso vergeblichen Versuch, ein Polygon durch Hinzufügen von Ecken zu einem Kreis zu gestalten. Erst durch ein Eingeständnis der fundamentalen Unsicherheit, die Akzeptanz eines Nichtwissens des Verstandes, das damit zu einem gelehrten Nichtwissen wird[29], kann eine Annäherung an die Wahrheit erfolgen. Für Dawes (1988) ist Unsicherheit eine Facette der menschlichen Welt, die für sich genommen weder bedrohlich noch gefährlich wirken muss. Ein permanenter Versuch, Sinn zu entnehmen aus einer nichtreduzierbaren Unsicherheit kann hingegen pathologische Konsequenzen zeitigen. Ähnlich können subkomplexe Erklärungen zur Unterdrückung von Unsicherheit zutiefst schädlich wirken, wenn falsche Verantwortliche identifiziert oder illusorische Erfolge reklamiert werden. Einfache Erklärungen sind in vielen Fällen attraktive Erklärungen und auch einfache Heuristiken könne zu einer Bevorzugung einfacher Erklärungen führen. Einfache Erklärungen sind aber in vielen Fällen nicht kostenlos:

> The price of our retreat from uncertainty is often paid by others. Some individuals believe that if people are poor, or are on the street or addicted or ill, they must somehow have done something to deserve that fate. In the face of such deservingness, help is futile. Moreover, given the biasing effect of retrospective memory [...] these people themselves may accept that judgment; for if they, too, believe in „just desserts," that could play a crucial role in determining their recall of what it is they *did*. (Dawes, 1988, S. 264)

Unsicherheit kann als ein Abgrund erscheinen, der keinen Halt bietet und den Menschen fundamental bedroht, da gemäß der Definition Sicherheit negiert wird, die als Ziel und fundamentales Gut geschätzt wird. Eine vollständig sichere und vorhersehbare Welt würde diese Bewertung aber *mit Sicherheit* umkehren [30]:

> Without uncertainty, there would be no hope, no ethics, and no freedom of choice. It is only because we do not know what the future holds for us (e.g., the exact time and manner of our own deaths) that we can have hope. It is only because we do not know the consequences of our actions (e.g., the exact effect of our treatment of others) that we face ethical choice. (Dawes, 1988, S. 272)

Selbst der Wunsch nach Sicherheit müsste nach ihrer Erlangung verlöschen, ein wesentlicher Teil der menschlichen Lebenswelt wäre verloren. Vollständige Unsicherheit ist wiederum ebenso wenig wünschenswert wie vollständige Sicherheit, aber in den Hohlstellen der Vorhersagbarkeit liegen auch positive Überraschungen, Spielräume verborgen sowie die Chance, Routinen zu durchbrechen und fundamental Neues zu entdecken. Ambiguitätstoleranz kann als eine Facette des positiven Umgangs mit Unsicherheit aufgefasst werden, deren Verbindung zu Autoritarismus auch auf politische Konsequenzen des Umgangs mit Unsicherheit verweist. Eine niedrigere Ambiguitätstoleranz muss zum Rückzug aus zahlreichen Bereichen des täglichen Lebens oder ihrer sehr selektiven Wahrnehmung führen:

> Those who choose to deny the reality of uncertainty invent a stable world of their own. Such people's natural desire to reduce uncertainty, which may be basic to the whole cognitive

[29]

> Ex quibus concludimus praecisionem veritatis in tenebris nostrae ignorantiae incomprehensibiliter lucere. Et haec est illa docta ignorantia, quam inquisivimus. (Nikolaus von Kues, 1440/1994, S. 112)

[30] Kurt Vonnegut führt im Roman *Timequake* die zutiefst aversiven Konsequenzen einer exakten Wiederholung von zehn Jahren Zeit aus.

> enterprise of understanding the world, runs amok to the point that they come to believe uncertainty does not exist. The „statistician's definition of an optimist“ as „someone who believes that the future is uncertain“ is not as cynical as it may first appear to be. (Dawes, 1988, S. 273)

Vergleichbar argumentiert Foley (2000) für die Notwendigkeit, Zweifel und Skeptizismus auszuhalten, da beide nicht nur destruktiv zu werten sind, sondern eine Vorbedingung für jede Epistemologie darstellen:

> The way to respond to skeptical doubts is not to legislate against them metaphysically, and it is not to dismiss them as meaningless, self-defeating, or even odd. It is rather to live with them. It is to recognize that what makes epistemology possible also makes skeptical worries inevitable - namely, our ability to make our methods of inquiry themselves into an object of inquiry. Within the context of such an inquiry, the worry that we might have widely mistaken beliefs is as natural as it is ineradicable. If this illustrates our whimsical condition, then so be it, but it is, after all, not so surprising. (Foley, 2000, S. 187)

Rationalität stellt sich diesen Zweifeln, wenn sie eine lebensfähige Rationalität darstellen will, ohne jede Unsicherheit wäre Rationalität ohnehin eine eher alberne Angelegenheit, genauso ohne jegliche Sicherheit. Rationalität kann dementsprechend auch nicht die Aufgabe haben, eine perfekte Sicherheit zu garantieren, diese Einsicht ist für ihren zuversichtlichen Einsatz unabdingbar:

> The reality of our intellectual lives is that we are working without nets. No procedure, no amount of reflection no amount of evidence gathering can guarantee that we won't fall into error, perhaps even massive error. We are thus forced to choose between proceeding in a way that we on reflection would take to be effective and proceeding in a way that we would not take to be effective. If we are rational, we opt for the former. (Foley, 2000, S. 190)

In diesem Sinne verlangt eine fundamental unsichere Welt von Menschen mehr ab als die Anwendung einfacher Heuristiken oder reliabler Prozeduren. Eine Welt, die in ihren Fundamenten unsicher steht, vor allem in jenen Strukturen, die auf sozialen Pfeilern ruhen, verlangt die Fähigkeit zur Improvisation in einer Grundhaltung, die nicht nur im Gesicht der Welt, sondern auch in der eigenen Person bereit ist, fundamental Neues zu entdecken und dazu den Mut hat, dem Eindruck zu folgen. Vielleicht muss man einräumen, dass Unsicherheit auf Sicherheit gegründet werden kann aber nicht umgekehrt:

> Man kann erst zweifeln, wenn man Gewisses gelernt hat; wie man sich erst verrechnen kann wenn man rechnen gelernt hat. Dann ist es allerdings unwillkürlich. (Wittgenstein, 1984a, §410)

Der Zweifel und die Unsicherheit können erst dann als solche abgegrenzt werden, wenn die Grenze von beiden Seiten erfahren worden ist, wobei die Zweifellosigkeit und Sicherheit den individuellen Ausgangspunkt bilden, der im Zuge der menschlichen Erfahrung aber im Allgemeinen verlassen werden muss, um nicht Brüche und massive Verwerfungen zu riskieren. Heuristiken und Routinen helfen dem Menschen, sich dort zurechtzufinden, wo ihn der Weg hinführt, aber ähnlich wie die Vorwärtsbewegung nicht ohne eine körperliche Instabilität zwischen den Schritten vollzogen werden kann, ist eine gedankliche Navigation durch die Welt nicht ohne kognitive Instabilität zwischen Einsichten und Konsolidierungen erfolgreich. Die Vergrößerung des eigenen Repertoires an Reaktionsweisen und Verarbeitungstechniken ist auf diesem Weg niemals abgeschlossen. Wie ein Lyriker die eigene Sprache vorfindet, um sie in ungewohnte Formen zu gießen, begegnet ein kreativer Denker den eigenen kognitiven Strukturen, um sie zu hinterfragen und mit ihnen zu experimentieren. Im Kern mag es zutreffen, dass ein großer Teil der Verarbeitung durch einfach erscheinende Mechanismen und Heuristiken beinahe automatisch geleistet wird, aber Einfachheit erkennt nur jener, der von Komplexität umgeben ist, und es wäre eine willkommene Ironie der Welt, wenn das Komplexeste des Menschen aus den einfachsten Bausteinen erwächst. Das Wiedererlangen von Einfachheit im Bewusstsein dieser Komplexität gehört zu den schwierigsten Aufgaben des Menschen, die mit jenem schwach Umrissenen untrennbar verbunden stehen, das begleitet von vagen Handbewegungen lediglich nahezu hilflos als „Menschsein“ betitelt werden kann.

Damit ist in dieser Arbeit ein Bogen geschlagen worden von einer Fragestellung in der Klassifikation von Menschen zu Klassifikationsmethoden und hin zum klassifizierenden Menschen selbst. Einfache

Heuristiken erweisen sich als begrüßenswerte Bestandteile der menschlichen Rationalität, auch wenn die Kategorisierung in „einfach“ und „komplex“ im Zuge der Betrachtung an Trennschärfe einbüßen musste. Die Rationalität, die in den Analysen aufgedeckt wurde, ist eine begrenzte Rationalität, deren Grenzen in den menschlichen Beschränkungen und den menschlichen Umgebungen zu suchen sind, aber auch in der endlichen konzeptuellen Reichweite eines individuell-kognitiven Zuschnitts. Vermieden wurde in dieser Arbeit die Ausweitung in die Sphäre des Sozialen, der Kommunikation, Semiotik, Mikropolitik, Sinngestaltung und Interpretation, der politischen Kritik und der menschlichen Lebenswelten. Territoriale Entdeckungen sind begleitet von der Unsicherheit, wie weit sich die Landmasse erstreckt, deren Strand die Boote angesteuert haben. Nicht jede Insel sollte deshalb frühzeitig zum Kontinent ausgerufen oder mit weitreichenden Namen belegt werden. Die benannten Ziele müssen deshalb auf die Schiffe kommender Expeditionen verlagert werden, und die aufgeschlagenen Zelte können sich vielleicht nicht als Heimat, aber zumindest als Zwischenlager erweisen, von dem aus die verbleibenden weißen Flecken der Karte vorsichtig (unter Verwendung löschbarer Tinte) mit Konturen bedeckt werden sollen.

Literaturverzeichnis

Adams, N. M. & Hand, D. J. (2000). Improving the Practice of Classifier Performance Assessment. *Neural Computation*, *12*, 305–311.

Aeberhard, S., Coomans, D. & de Vel, O. (1992). *The classification performance of RDA* (Tech. Rep. no. 92-01). Dept. of Computer Science and Dept. of Mathematics and Statistics, James Cook University of North Queensland.

Ahmad, R. & Bath, P. A. (2004). The use of Cox regression and genetic algorithm (CoRGA) for identifying risk factors for mortality in older people. *Health Informatics Journal*, *10*(3), 221–236.

Aickelin, U. & Dowsland, K. A. (1998). Exploiting problem structure in a genetic algorithm approach to a nurse rostering problem. *Journal of Scheduling*, *3*(3), 139–153.

Aickelin, U. & Dowsland, K. A. (2002). Enhanced Direct and Indirect Genetic Algorithm Approaches for a Mall Layout and Tenant Selection Problem. *Journal of Heuristics*, *8*, 503–514.

Alba, E. & Troya, J. M. (2002). Improving flexibility and efficiency by adding parallelism to genetic algorithms. *Statistics and Computing*, *12*, 91–114.

Aldershof, B. & Carducci, O. M. (1999). Stable Marriage and Genetic Algorithms: A Fertile Union. *Journal of Heuristics*, *5*, 29–46.

Amber, S. W. (2005). *The Elements of UML 2.0 Style.* Cambridge: Cambridge University Press.

Amelang, M. & Zielinski, W. (2002). *Psychologische Diagnostik und Intervention* (3., korr., akt. u. überarb. Aufl.). Berlin: Springer.

Anderson, R. W. (2000). The Baldwin effect. In T. Bäck, D. B. Fogel & Z. Michalewicz (Hrsg.), *Evolutionary Computation 1: Basic Algorithms and Operators* (S. 308–317). Bristol: Institute of Physics Publishing.

Araki, H., Kaji, T., Yamamoto, M., Suzuki, K. & Ohuch, A. (2001). A Genetic Algorithm for the Optimal Sequential Partitioning Problem. *Electrical Engineering in Japan*, *135*(4), 215–221.

Ariew, A. & Lewontin, R. (2004). The Confusions of Fitness. *Brit. J. Phil. Sci.*, *55*, 347–363.

Asuncion, A. & Newman, D. J. (2007). *UCI Machine Learning Repository* (http://www.ics.uci.edu/∼mlearn/MLRepository.html). University of California, Irvine, School of Information and Computer Sciences.

Audi, R. (2004). Theoretical Rationality: Its Sources, Structure, and Scope. In A. R. Mele & P. Rawling (Hrsg.), *The Oxford handbook of rationality* (S. 17–44). New York: Oxford University Press.

Austin, J. T. & Villanova, P. (1992). The Criterion Problem: 1917-1992. *Journal of Applied Psychology*, *77*(6), 836–874.

Axelrod, R. (1986). An Evolutionary Approach to Norms. *American Political Science Review*, *80*(4), 1095–1111.

Axelrod, R. (1997). *The complexity of cooperation: agent-based models of competition and collaboration.* Princeton, NJ: Princeton University Press.

Bäck, T. (2000). Binary strings. In T. Bäck, D. B. Fogel & Z. Michalewicz (Hrsg.), *Evolutionary Computation 1: Basic Algorithms and Operators* (S. 132–135). Bristol: Institute of Physics Publishing.

Balzert, H. (2001). *Lehrbuch der Software-Technik: Software-Entwicklung* (2. Aufl.). Heidelberg: Spektrum.

Banzhaf, W. (2000). Interactive evolution. In T. Bäck, D. B. Fogel & Z. Michalewicz (Hrsg.), *Evolutionary Computation 1: Basic Algorithms and Operators* (S. 228–234). Bristol: Institute of Physics Publishing.

Barriga, K. J., Hamman, R. F., Hoag, S., Marshall, J. A. & Shetterly, S. M. (1996). Population screening for glucose intolerant subjects using decision tree analyses. *Diabetes Research and Clinical Practice*, *34*, S17–S29.

Bartelborth, T. (2004). Wofür sprechen die Daten? *Journal for General Philosophy of Science*, *35*, 13–40.

Bartram, D. (2000). Internet Recruitment and Selection: Kissing Frogs to find Princes. *International Journal of Selection and Assessment*, *8*(4), 261–274.

Baudrillard, J. (1994). *Simulacra and Simulation.* Ann Arbor: University of Michigan Press.

Beckers, R., Deneubourg, J.-L. & Goss, S. (1992). Trails and U-turns in the selection of the shortest path by the ant Lasius Niger. *Journal of Theoretical Biology*, *159*, 397–415.

Bertrand, J. (1907). *Calcul des probabilites.* Paris: Hermann et Fils.

Binning, J. F. & Barrett, G. V. (1989). Validity of Personnel Decisions: A Conceptual Analysis of the Inferential and Evidential Bases. *Journal of Applied Psychology*, *74*(3), 478–494.

Birchenhall, C., Kastrinos, N. & Metcalfe, S. (1997). Genetic algorithms in evolutionary modelling. *Journal of Evolutionary Economics*, *7*, 375–393.

Birnbaum, M. H. (1999). How to show that $9 > 221$: Collect judgments in a between-subjects design. *Psychological Methods*, *4*, 243–249.

Bishop, M. A. (2000). In Praise of Epistemic Irresponsibility: How Lazy and Ignorant Can You Be? *Synthese*, *122*, 179–208.

Bishop, M. A. & Trout, J. D. (2002). 50 Years of Successful Predictive Modeling Should Be Enough: Lessons for Philosophy of Science. *Philosophy of Science*, *69*, S197–S208.

Blickle, T. (2000). Tournament selection. In T. Bäck, D. B. Fogel & Z. Michalewicz (Hrsg.), *Evolutionary Computation 1: Basic Algorithms and Operators* (S. 181–186). Bristol: Institute of Physics Publishing.

Boll, H. (2003). *Evolutionäre Verfahren zur Optimierung von Produktionsplänen mittels impliziter Kooperation.* Dissertation, Rheinisch-Westfälische Technische Hochschule Aachen.

Booker, L. B. (2000). Recombination. Binary strings. In T. Bäck, D. B. Fogel & Z. Michalewicz (Hrsg.), *Evolutionary Computation 1: Basic Algorithms and Operators* (S. 256–270). Bristol: Institute of Physics Publishing.

Borges, B., Goldstein, D. G., Ortmann, A. & Gigerenzer, G. (1999). Can Ignorance Beat the Stock Market? In G. Gigerenzer, P. M. Todd & the ABC Research Group (Hrsg.), *Simple heuristics that make us smart* (S. 59–72). New York: Oxford University Press.

Borland GmbH. (2001). *Object Pascal: Sprachreferenz.* Langen: Borland GmbH.

Bortz, J. (1993). *Statistik für Sozialwissenschaftler* (4., vollst. überarb. Aufl.). Springer.

Bowker, G. C. & Star, S. L. (1999). *Sorting Things Out: Classification and Its Consequences.* Cambrdige, MA: MIT Press.

Boyd, M. (2001). On Ignorance, Intuition, and Investing: A Bear Market Test of the Recognition Heuristic. *The Journal of Psychology and Financial Markets*, *2*(3), 150–156.

Brandstätter, E., Gigerenzer, G. & Hertwig, R. (2006). The Priority Heuristic: Making Choices Without Trade-Offs. *Psychological Review*, *113*(2), 409–432.

Breiman, L., H.Friedman, J., Olshen, R. & Stone, C. J. (1984). *Classification and Regression Trees.* Monterey, CA: Wadsworth.

Bröder, A. (2000). Assessing the Empirical Validity of the „Take-The-Best“ Heuristic as a Model of Human Probabilistic Inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 1332–1346.

Bröder, A. (2001). Die eingeschränkte Fruchtbarkeit eines unvollständigen Forschungsprogramms: Kommentar zu Hertwig und Hoffrage (2001). *Psychologische Rundschau*, *52*(3), 159–163.

Bröder, A. (2002). Take The Best, Dawes' Rule, and Compensatory Decision Strategies: A Regression-based Classification Method. *Quality & Quantity*, *36*, 219–238.

Bröder, A. (2003). Decision Making With the „Adaptive Toolbox“: Influence of Environmental Structure, Intelligence, and Working Memory Load. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(4), 611–625.

Bröder, A. (2005). *Entscheiden mit der „adaptiven Werkzeugkiste“: Ein empirisches Forschungsprogramm.* Lengerich: Pabst.

Broggi, A., Cellario, M., Lombardi, P. & Porta, M. (2003). An Evolutionary Approach to Visual Sensing for Vehicle Navigation. *IEEE Transactions on Industrial Electronics*, *50*(1).

Broome, J. (1990). Should a Rational Agent Maximize Expected Utility? In K. Schweers Cook & M. Levi (Hrsg.), *The Limits of rationality* (S. 132–145). Chicago: University of Chicago Press.

Brunetti, A. & Golosio, B. (2001). Fit of EDXRF spectra with a genetic algorithm. *X-Ray Spectrometry*, *30*, 32–36.

Büttner, J. (1997). Diagnostic validity as a theoretical concept and as a measurable quantity. *Clinica Chimica Acta*, *260*, 131–143.

Cantù, M. (2001). *Mastering Delphi 6.* San Francisco: Sybex.

Caverni, J.-P. & Péris, J.-L. (1990). The anchoring-adjustment heuristic in an 'information rich, real world setting': knowledge assessment by experts. In J.-P.Caverni, J.-M. Fabre & M. Gonzalez (Hrsg.), *Cognitive biases* (S. 35–45). Amsterdam: North-Holland.

Cestnik, G., Konenenko, I. & Bratko, I. (1987). Assistant-86: A Knowledge-Elicitation Tool for Sophisticated Users. In I. Bratko & N. Lavrac (Hrsg.), *Progress in Machine Learning* (S. 31–45). Wilmslow: Sigma.

Chang, B. C. H., Ratnaweera, A., Halgamuge, S. K. & Watson, H. C. (2004). Particle Swarm Optimisation for Protein Motif Discovery. *Genetic Programming and Evolvable Machines*, *5*, 203–214.

Chang, C. K., Christensen, M. J. & Zhang, T. (2001). Genetic Algorithms for Project Management. *Annals of Software Engineering*, *11*, 107–139.

Chase, V. M., Hertwig, R. & Gigerenzer, G. (1998). Visions of Rationality. *Trends in Cognitive Sciences*, *2*(6), 206–214.

Chater, N., Oaksford, M., Nakisa, R. & Redington, M. (2003). Fast, frugal and rational: How rational norms explain behavior. *Organizational Behavior and Human Decision Processes*, *90*, 63–86.

Chatterjee, A. & Siarry, P. (2004). Nonlinear inertia weight variation for dynamic adaptation in particle swarm optimization. *Computers & Operations Research*, *33*, 859–871.

Chatterjee, S., Laudato, M. & Lynch, L. A. (1996). Genetic algorithms and their statistical applications: an introduction. *Computational Statistics & Data Analysis*, *22*, 633–651.

Chaudhury, P. & Bhattacharyya, S. P. (2000). Stochastic Construction of Reaction Paths: A Genetic Algorithm-Based Approach. *International Journal of Quantum Chemistry*, *76*, 161–168.

Chen, M. & Zalzala, A. M. S. (1997). A Genetic Approach to Motion Planning of Redundant Mobile Manipulator Systems Considering Safety and Configuration. *Journal of Robotic Systems*, *14*(7), 529–544.

Chen, S.-H. & Ni, C.-C. (2000). Simulating the Ecology of Oligopolistic Competition with Genetic Algorithms. *Knowledge and Information Systems*, *2*, 285–309.

Cheng, M.-Y. & Lin, C.-S. (1997). Genetic Algorithm for Control Design of Biped Locomotion. *Journal of Robotic Systems*, *14*(5), 365–373.

Chiu, C.-C. & Chen, W.-T. (2001). Comparison of Image Reconstruction by Using Near-Field and Far-Field Data for an Imperfect Conductor. *Int J RF and Microwave CAE*, *11*, 79–83.

Chtioui, Y., Bertrand, D. & Barba, D. (1998). Feature Selection by a Genetic Algorithm. Application to Seed Discrimination by Artificial Vision. *J Sci Food Agric*, *76*, 77–86.

Chuanwen, J. & Bompard, E. (2005). A self-adaptive chaotic particle swarm algorithm for short term hydroelectric system scheduling in deregulated environment. *Energy Conversion and Management*, *46*, 2689–2696.

Coello Coello, C. A. & Aguirre, A. H. (2002). Design of combinational logic circuits through an evolutionary multiobjective optimization approach. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, *16*, 39–53.

Conradie, A., Miikkulainen, R. & Aldrich, C. (2002). Adaptive control utilising neural swarming. In W. B. Langdon, E. Cantú-Paz, K. Mathias, R. Roy, D. Davis, R. Poli, K. Balakrishnan, V. Honavar, G. Rudolph, J. Wegener, L. Bull, M. A. Potter, A. C. Schultz, J. F. Miller, E. Burke & N. Jonoska (Hrsg.), *Gecco 2002: Proceedings of the genetic and evolutionary computation conference* (S. 60–67). New York: Morgan Kaufmann Publishers.

Cooper, R. (2000). Simple heuristics could make us smart; but which heuristics do we apply when? *Behavioral and Brain Sciences*, *23*, 746.

Cooper, R. P. (2006). Cognitive Architectures as Lakatosian Research Programs: Two Case Studies. *Philosophical Psychology*, *19*(2), 199–220.

Cordón, O., Herrera, F. & Stützle, T. (2002). A Review on the Ant Colony Optimization Metaheuristic: Basis, Models and New Trends. *Mathware & Soft Computing*, *9*, 1–35.

Cormen, T. H., Leiserson, C. E., Rivest, R. L. & Stein, C. (2001). *Introduction to Algorithms (2nd ed.).* Cambridge, MA: MIT.

Cortés, P., Larrañeta, J., Onieva, L., García, J. M. & Caraballo, M. S. (2001). Genetic algorithm for planning cable telecommunication networks. *Applied Soft Computing*, *1*, 21–33.

Costall, A. & Still, A. (1991). Introduction: Cognitivism as an approach to cognition. In A. Still & A. Costall (Hrsg.), *Against Cognitivism: Alternative Foundations for Cognitive Psychology* (S. 1–7). Hemel Hempstead, Hertfordshire: Harvester.

Cox, M. T. (2005). Metacognition in computation: A selected research review. *Artificial Intelligence*, *169*, 104–141.

Czerlinski, J., Gigerenzer, G. & Goldstein, D. G. (1999). How Good Are Simple Heuristics? In G. Gigerenzer, P. M. Todd & the ABC Research Group (Hrsg.), *Simple heuristics that make us smart* (S. 97–118). New York: Oxford University Press.

Daston, L. (2001). *Wunder, Beweise und Tatsachen: Zur Geschichte der Rationalität.* Frankfurt a. M.: S. Fischer.

Davidson, D. (2004). Paradoxes of Irrationality. In D. Davidson (Hrsg.), *Problems of Rationality* (S. 169–187). Oxford: Clarendon.

Dawes, R. (1971). A Case Study of Graduate Admissions: Application of Three Principles of Human Decision Making. *American Psychologist*, *26*, 180–188.

Dawes, R. & Corrigan, B. (1974). Linear Models in Decision Making. *Psychological Bulletin*, *81*, 95–106.

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, *34*, 571–582.

Dawes, R. M. (1982). The robust beauty of improper linear models in decision making. In D. Kahneman, P. Slovic & A. Tversky (Hrsg.), *Judgment under uncertainty: Heuristics and biases* (S. 391–407). New York: Cambridge University Press.

Dawes, R. M. (1988). *Rational Choice in an Uncertain World.* Fort Worth: Harcourt Brace.

Dawid, H. & Kope, M. (1998). On economic applications of the genetic algorithm: a model of the cobweb type. *Journal of Evolutionary Economics*, *8*, 297–315.

Dawkins, R. (1987). *The blind watchmaker: why the evidence of evolution reveals a universe without design.* New York: Norton.

De Jong, K. A. (1975). *An analysis of the behavior of a class of genetic adaptive systems.* Dissertation, University of Michigan.

De la Fontaine, A., Glas, P., Schafsteller, C. & Wottawa, H. (2001). E-Recruiting - Hintergründe und Trends. Beispiele für erste Erfahrungen bei der Siemens AG. In H. C. Riekhoff (Hrsg.), *Strategien der Personalentwicklung* (5. Aufl., S. 359–380). Wiesbaden: Gabler.

Deb, K. (2000a). Encoding and decoding functions. In T. Bäck, D. B. Fogel & Z. Michalewicz (Hrsg.), *Evolutionary Computation 2: Advanced Algorithms and Operators* (S. 4–11). Bristol: Institute of Physics Publishing.

Deb, K. (2000b). Introduction to selection. In T. Bäck, D. B. Fogel & Z. Michalewicz (Hrsg.), *Evolutionary Computation 1: Basic Algorithms and Operators* (S. 166–171). Bristol: Institute of Physics Publishing.

Dennett, D. C. (1998). *Brainchildren: Essays on Designing Minds.* Cambridge, MA: MIT.

Detrano, R., Janosi, A., Steinbrunn, W., Pfisterer, M., Schmid, J., Sandhu, S., Guppy, K., Lee, S. & Froelicher, V. (1989). International application of a new probability algorithm for the diagnosis of coronary artery disease. *American Journal of Cardiology*, 304–310.

Dhami, M. K. & Ayton, P. (2001). Bailing and Jailing the Fast and Frugal Way. *Journal of Behavioral Decision Making*, *14*(2), 141–168.

Dhar, V., Chou, D. & Provost, F. (2000). Discovering Interesting Patterns for Investment Decision Making with GLOWER - A Genetic Learner Overlaid with Entropy Reduction. *Data Mining and Knowledge Discovery*, *4*, 251–280.

DIN 33430. (Ausgabe: 2002-06). *Anforderungen an Verfahren und deren Einsatz bei berufsbezogenen Eignungsbeurteilungen.* Berlin: DIN Deutsches Institut für Normierung.

Domingos, P. (1998). Occam's two razors: the sharp and the blunt. In R. Agrawal & P. Stolorz (Hrsg.), *Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining (KDD '98)* (S. 37–43). AAAI Press.

Domínguez, O. & Arcos, M. J. (2000). Speciation of Chromium by Adsorptive Stripping Voltammetry Using Pyrocatechol Violet. *Electroanalysis*, *12*(6), 449–458.

Dorigo, M. & Di Caro, G. (1999). Ant colony optimization: A new meta-heuristic. In *1999 Congress on Evolutionary Computation* (S. 1470–1477). Piscataway, NJ: IEEE Service Center.

Dorigo, M. & Gambardella, L. M. (1997). Ant colonies for the travelling salesman problem. *BioSystems*, *43*, 73–81.

Dreyfus, H. L. (2007). Why Heideggerian AI Failed and How Fixing it Would Require Making it More Heideggerian. *Philosophical Psychology*, *20*(2), 247–268.

Durant, E. A. (2002). *Hearing Aid Fitting with Genetic Algorithms.* Dissertation, University of Michigan.

Eby, D., Averill, R. C., Gelfand, B., Punch, W. F., Mathews, O. & Goodman, E. D. (1997). An Injection Island GA for Flywheel Design Optimization. In H.-J. Zimmermann (Hrsg.), *Eufit '97 – 5th European Congress on Intelligent Techniques and Soft Computing* (S. 687–691). Aachen: Verlag Mainz, Wissenschaftsverlag.

Eiben, A. E. (2000). Recombination. Multiparent recombination. In T. Bäck, D. B. Fogel & Z. Michalewicz (Hrsg.), *Evolutionary Computation 1: Basic Algorithms and Operators* (S. 289–307). Bristol: Institute of Physics Publishing.

Eiben, A. E. & Ruttkay, Z. (2000). Constraint-satisfaction problems. In T. Bäck, D. B. Fogel & Z. Michalewicz (Hrsg.), *Evolutionary Computation 2: Advanced Algorithms and Operators* (S. 75–86). Bristol: Institute of Physics Publishing.

Einhorn, H. J. (1972). Alchemy in the Behavioral Sciences. *The Public Opinion Quarterly*, *36*(3), 367–378.

Einhorn, H. J. & Hogarth, R. M. (1981). Behavioral Decision Theory: Processes of Judgment and Choice. *Annual Review of Psychology*, *32*, 53–88.

Eisenführ, F. & Weber, M. (2003). *Rationales Entscheiden* (4., neu bearb. Aufl.). Berlin: Springer.

Elegbede, C. (2005). Structural reliability assessment based on particles swarm optimization. *Structural Safety*, *27*, 171–186.

Eliseev, A. V. & Nelen, M. I. (1998). Use of Molecular Recognition to Drive Chemical Evolution: Mechanisms of an Automated Genetic Algorithm Implementation. *Chem. Eur. J.*, *4*(5), 825–834.

Elster, J. (1989). *Solomonic Judgements: Studies in the Limitations of Rationality.* Paris: Cambridge University Press.

Emmeche, C. (1994). *The Garden in the Machine: The Emerging Science of Artificial Life.* Princeton, NJ: Princeton University Press.

Erlich, Z., Gelbard, R. & Spiegler, I. (2002). Data Mining by Means of Binary Representation: A Model for Similarity and Clustering. *Information Systems Frontiers*, *2*(2), 187–197.

Eshelman, L. J. (2000). Genetic algorithms. In T. Bäck, D. B. Fogel & Z. Michalewicz (Hrsg.), *Evolutionary Computation 1: Basic Algorithms and Operators* (S. 64–80). Bristol: Institute of Physics Publishing.

Eskridge, W. H. (1991). Overriding Statutory Interpretation Decisions. *Yale Law Journal*, *101*, 331–417.

Etzel, S. & Küppers, A. (2002). Moderne Potenzialanalyse mit dem Testsystem „pro facts“. *Wirtschaftspsychologie*, *9*(1), 21–27.

Everitt, B. (2001). *Cluster Analysis* (4. Aufl.). London: Arnold.

Ezziane, Z. (2002). Solving the 0\1 knapsack problem using an adaptive genetic algorithm. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, *16*, 23–30.

Falk, R. & Greenbaum, C. W. (1995). Significance Tests Die Hard: The Amazing Persistence of a Probabilistic Misconception. *Theory & Psychology*, *5*(1), 75–98.

Feeney, A. (2000). Simple heuristics: From one infinite regress to another? *Behavioral and Brain Sciences*, *23*, 749-750.

Feger, H. (1991). Über das Messen in der Psychologie. In K. Grawe, R. Hänni, N. Semmer & F. Tschan (Hrsg.), *Über die richtige Art, Psychologie zu betreiben* (S. 79–88). Göttingen: Hogrefe.

Feldman, D. C. & Klaas, B. S. (2002). Internet job hunting: a field study of applicant experiences with on-line recruiting. *Human Resource Management*, *41*(2), 175–192.

Fischer, J. E., Steiner, F., Zucol, F., Berger, C. & al., L. M. nad W. Bossart et. (2002). Use of simple heuristics to target macrolide prescription in children with community-acquired pneumonia. *Archives of Pediatrics and Adolescent Medicine*(156), 1005–1008.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annual Eugenics*, *7*(II), 179–188.

Fodor, J. (2000). *The Mind Doesn't Work That Way.* Cambridge, MA: MIT Press.

Fodor, J. (2005). Reply to Steven Pinker 'So How Does The Mind Work?'. *Mind & Language*, *20*(1), 25–32.

Fogel, D. B. (2000a). Other selection methods. In T. Bäck, D. B. Fogel & Z. Michalewicz (Hrsg.), *Evolutionary Computation 1: Basic Algorithms and Operators* (S. 201–204). Bristol: Institute of Physics Publishing.

Fogel, D. B. (2000b). Real-valued vectors. In T. Bäck, D. B. Fogel & Z. Michalewicz (Hrsg.), *Evolutionary Computation 1: Basic Algorithms and Operators* (S. 136–138). Bristol: Institute of Physics Publishing.

Fogel, D. B. (2000c). Recombination. Real-valued vectors. In T. Bäck, D. B. Fogel & Z. Michalewicz (Hrsg.), *Evolutionary Computation 1: Basic Algorithms and Operators* (S. 270–274). Bristol: Institute of Physics Publishing.

Fogel, D. B. & Angeline, P. J. (2000). Guidelines for a suitable encoding. In T. Bäck, D. B. Fogel & Z. Michalewicz (Hrsg.), *Evolutionary Computation 1: Basic Algorithms and Operators* (S. 160–162). Bristol: Institute of Physics Publishing.

Foley, R. (2000). Skepticism and Rationality. In E. Sosa & J. Kim (Hrsg.), *Epistemology: An Anthology* (S. 182–191). Malden, MA: Blackwell.

Forster, M. R. (1999). How Do Simple Rules 'Fit to Reality' in a Complex World? *Minds and Machines*, *9*, 543–564.

Freedman, D. A. (1991). Statistical Models and Shoe Leather. *Sociological Methodology*, *21*, 291–313.

Freedman, D. A. (1999). From Association to Causation: Some Remarks on the History of Statistics. *Statistical Science*, *14*(3), 243–258.

Freitas, A. A. (2002). *Data Mining and Knowledge Discovery with Evolutionary Algorithms.* Berlin: Springer.

Fränti, P. (2000). Genetic algorithm with deterministic crossover for vector quantization. *Pattern Recognition Letters*, *21*, 61–68.

Fukuyama, Y., Takayama, S., Nakanishi, Y. & Yoshida, H. (1999). A Particle Swarm Optimization for Reactive Power and Voltage Control in Electric Power Systems. In W. Banzhaf, J. Daida, A. E. Eiben, M. H. Garzon, V. Honavar, M. Jakiela & R. E. Smith (Hrsg.), *Proceedings of the genetic and evolutionary computation conference* (Bd. 2, S. 1523–1528). Orlando, Florida, USA: Morgan Kaufmann.

Fuller, S. (2000). Against an uncritical sense of adaptiveness. *Behavioral and Brain Sciences*, *23*, 750–751.

Funke, J. (2003). *Problemlösendes Denken.* Stuttgart: Kohlhammer.

Funke, U. (1993). Computergestützte Eignungsdiagnostik mit komplexen dynamischen Szenarios. *Zeitschrift für Arbeits- und Organisationspsychologie*, *37 (N.F. 11)*(3), 109–118.

Gallant, M. & Boudreau, R. (2002). The Synthesis of Planar Parallel Manipulators with Prismatic Joints for an Optimal, Singularity-Free Workspace. *Journal of Robotic Systems*, *19*(1), 13–24.

Gardiner, E. J., Willett, P. & Artymiuk, P. J. (2001). Protein Docking Using a Genetic Algorithm. *PROTEINS: Structure, Function, and Genetics*, *44*, 44–56.

Gardner, M. (1970). The Fantastic Combinations of John Conway's New Solitaire Game 'Life'. *Scientific American*, *223*(4), 120–123.

Garfinkel, A. & Mitarbeiter. (1999). Prognostic Value of Dobutamine Stress Echocardiography in Predicting Cardiac Events in Patients With Known or Suspected Coronary Artery Disease. *Journal of the American College of Cardiology*, *3*, 708–716.

Gatarski, R. (2002). Breed better banners: design automation through on-line interaction. *Journal of Interactive Marketing*, *16*(1), 2–13.

Gelbard, R. & Spiegler, I. (2000). Hempel's Raven Paradox: a positive approach to cluster analysis. *Computers & Operations Research*, *27*, 305–320.

Gergen, K. J., Hepburn, A. & Comer Fisher, D. (1986). Hermeneutics of Personality Description. *Journal of Personality and Social Psychology*, *50*(6), 1261–1270.

Germanson, T. P., Lanzino, G., Kongable, G. L., Torner, J. C. & Kassell, N. F. (1998). Risk Classification after Aneurysmal Subarachnoid Hemorrhage. *Surg Neurol*, *49*, 155–163.

Ghiselli, E. E. & Haire, M. (1969). The validation of selection tests in the light of the dynamic character of criteria. *Personnel Psychology*, *13*, 225–231.

Gies, D. & Rahmat-Samii, Y. (2003). Particle swarm optimization for reconfigurable phasedifferentiated array design. *Microwave and optical technology letters*, *38*(3), 168–175.

Gigerenzer, G. (1981). *Messung und Modellbildung in der Psychologie.* München: E. Reinhard.

Gigerenzer, G. (1991). From Tools to Theories: A Heuristic of Discovery in Cognitive Psychology. *Psychological Review*, *98*(2), 254–267.

Gigerenzer, G. (1993). The bounded rationality of probabilistic mental models. In K. Manktelow & D. Over (Hrsg.), *Rationality: Psychological and philosophical perspectives* (S. 286–313). London: Routledge.

Gigerenzer, G. (1996). On Narrow Norms and Vague Heuristics: A Reply to Kahnemann and Tversky (1996). *Psychological Review*, *103*(3), 592–596.

Gigerenzer, G. (1997). Bounded Rationality: Models of Fast and Frugal Inference. *Swiss Journal of Economics and Statistics*, *133*(2/2).

Gigerenzer, G. (1998). Surrogates for Theories. *Theory & Psychology*, *8*(2), 195–204.

Gigerenzer, G. (2001a). Content-blind norms, no norms, or good norms? A reply to Vranas. *Cognition*, *81*, 93–103.

Gigerenzer, G. (2001b). Decision Making: Nonrational Theories. In N. Smelser & P. Baltes (Hrsg.), *International Encyclopedia of the Social and Behavioral Sciences* (Bd. 5, S. 3304–3309). Oxford: Elsevier.

Gigerenzer, G. (2001c). The adaptive toolbox: Towards a Darwinian rationality. In J. A. French, A. Kamil & D. W. Leger (Hrsg.), *Nebraska Symposium on Motivation: Evolutionary, Psychology and Motivation* (S. 113–143). Lincoln: University of Nebraska Press.

Gigerenzer, G. (2001d). The Adaptive Toolbox. In G. Gigerenzer & R. Selten (Hrsg.), *Bounded Rationality: The Adaptive Toolbox* (S. 37–50). Cambridge: MIT Press.

Gigerenzer, G. (2002). *Calculated Risks: How to Know When Numbers Deceive You.* New York: Simon & Schuster.

Gigerenzer, G. (2004a). Fast and Frugal Heuristics: The Tools of Bounded Rationality. In D. Koehler & N. Harvey (Hrsg.), *Blackwell handbook of judgment and decision making* (S. 62–88). Oxford: Blackwell.

Gigerenzer, G. (2004b). Striking a Blow for Sanity in Theories of Rationality. In M. Augier & J. G. March (Hrsg.), *Models of man: Essays in memory of Herbert A. Simon* (S. 389–409). Cambridge, MA: MIT Press.

Gigerenzer, G. (2005a). I Think, Therefore I Err. *social research*, *72*(1), 195–218.

Gigerenzer, G. (2005b). Is the mind irrational or ecologically rational? In F. Parisi & V. Smith (Hrsg.), *The law and economics of irrational behavior* (S. 37–67). Stanford, CA: Stanford University Press.

Gigerenzer, G. (2006). Bounded and Rational. In R. Stanton (Hrsg.), *Contemporary debates in cognitive science* (S. 115–133). Oxford: Blackwell.

Gigerenzer, G. & Dieckmann, A. (2006). Oblivious Confusion. In R. Rapp, P. Sedlmeier & G. Zunker-Rapp (Hrsg.), *Perspectives on Cognition: A Festschrift for Manfred Wettler* (S. 19–29). Lengerich: Pabst.

Gigerenzer, G. & Goldstein, D. G. (1996a). Mind as Computer: Birth of a Metaphor. *Creativity Research Journal*, *9*(2 & 3), 131-144.

Gigerenzer, G. & Goldstein, D. G. (1996b). Reasoning the Fast and Frugal Way: Models of Bounded Rationality. *Psychological Review*, *103*(4), 650–669.

Gigerenzer, G. & Goldstein, D. G. (1999). Betting on One Good Reason: The Take The Best Heuristic. In G. Gigerenzer, P. M. Todd & the ABC Research Group (Hrsg.), *Simple heuristics that make us smart* (S. 75–95). New York: Oxford University Press.

Gigerenzer, G. & Kurzenhäuser, S. (2005). Fast and frugal heuristics in medical decision making. In R. Bibace, J. D. Laird, K. L. Noller & J. Valsiner (Hrsg.), *Science and medicine in dialogue* (S. 3–15). Praeger: Westport, CT.

Gigerenzer, G. & Selten, R. (2001). Rethinking Rationality. In G. Gigerenzer & R. Selten (Hrsg.), *Bounded Rationality: The Adaptive Toolbox* (S. 1–12). Cambridge: MIT Press.

Gigerenzer, G., Swijtink, Z., Porter, T., Daston, L., Beatty, J. & Krüger, L. (1989). *Th Empire of Chance: How probability changed science and everyday life.* Cambridge: Cambridge University Press.

Gigerenzer, G. & Todd, P. M. (1999). Fast and Frugal Heuristics: The Adaptive Toolbox. In G. Gigerenzer, P. M. Todd & the ABC Research Group (Hrsg.), *Simple heuristics that make us smart* (S. 3–34). New York: Oxford University Press.

Glymour, C., Madigan, D., Pregibon, D. & Smyth, P. (1997). Statistical Themes and Lessons for Data Mining. *Data Mining and Knowledge Discovery*, *1*, 11–28.

Gómez, M. & Bielza, C. (2004). Node deletion sequences in influence diagrams using genetic algorithms. *Statistics and Computing*, *14*, 181–198.

Goh, K. S., Lim, A. & Rodrigues, B. (2003). Sexual Selection for Genetic Algorithms. *Artificial Intelligence Review*, *19*, 123–152.

Goldberg, D. E. (1989). *Genetic algorithms in search, optimization, and machine learning.* Reading, MA: Addison-Wesley.

Goldberg, L. (1970). Man vs. Model of Man: A Rationale, Plus Some Evidence, for a Method of Improving on Clinical Inferences. *Psychological Bulletin*, *73*, 422–432.

Goldstein, D. G. & Gigerenzer, G. (1999). The Recognition Heuristic: How Ignorance Makes Us Smart. In G. Gigerenzer, P. M. Todd & the ABC Research Group (Hrsg.), *Simple heuristics that make us smart* (S. 37–58). New York: Oxford University Press.

Goldstein, D. G. & Gigerenzer, G. (2002). Models of Ecological Rationality: The Recognition Heuristic. *Psychological Review*, *109*(1), 75–90.

Goodie, A., Ortmann, A., Nerissa Davis, J., Bullock, S. & Werner, G. M. (1999). Demons Versus Heuristics in Artificial Intelligence, Behavioral Ecology, and Economics. In G. Gigerenzer, P. M. Todd & the ABC Research Group (Hrsg.), *Simple heuristics that make us smart* (S. 327–355). New York: Oxford University Press.

Goodman, N. (1965). *Fact, fiction and forecast* (2. Aufl.). Cambridge, MA: Harvard University Press.

Gosepath, S. (2002). Eine einheitliche Konzeption von Rationalität: Ein Zugang aus Sicht der Analytischen Philosophie. In N. C. Karafyllis & J. C. Schmidt (Hrsg.), *Zugänge zur Rationalität der Zukunft* (S. 29–52). Stuttgart: Metzler.

Goss, S., Aron, S., Deneubourg, J.-L. & J.Pasteels. (1989). Self-organized shortcuts in the argentine ant. *Naturwissenschaften*, *76*, 579–581.

Gür Ali, F. . & Wallace, W. A. (1997). Bridging the gap between business objectives and parameters of data mining algorithms. *Decision Support Systems*, *21*, 3–15.

Grabmeier, J. & Rudolph, A. (2002). Techniqus of Ckuster Algorithms in Data Mining. *Data Mining and Knowledge Discovery*, *6*, 303–360.

Greaney, V. & Kelleghan, T. (1984). *Equality of Opportunity in Irish Schools.* Dublin: Educational Company.

Green, L. & Mehr, D. (1997). What alters physicians' decisions to admit to the coronary care unit? *The Journal of Family Practice*, *45*, 219–226.

Grefenstette, J. (2000a). Proportional selection and sampling algorithms. In T. Bäck, D. B. Fogel & Z. Michalewicz (Hrsg.), *Evolutionary Computation 1: Basic Algorithms and Operators* (S. 172–180). Bristol: Institute of Physics Publishing.

Grefenstette, J. (2000b). Rank-based selection. In T. Bäck, D. B. Fogel & Z. Michalewicz (Hrsg.), *Evolutionary Computation 1: Basic Algorithms and Operators* (S. 187–194). Bristol: Institute of Physics Publishing.

Groner, R. & Groner, M. T. (1991). Heuristische versus algorithmische Orientierung als Dimension des individuellen kognitiven Stils. In K. Grawe, R. Hänni, N. Semmer & F. Tschan (Hrsg.), *Über die richtige Art, Psychologie zu betreiben* (S. 315–330). Göttingen: Hogrefe.

Grubitzsch, S. (1991). Tests im diagnostischen Prozeß: Grenzen und Probleme. In S. Grubitzsch (Hrsg.), *Testtheorie - Testpraxis: Psychologische Tests undPrüfverfahren im kritischen Überblick* (S. 184–243). Hamburg: Rowohlt.

Guion, R. M. (1987). Changing views for personnel selection research. *Personnel Psychology*, *40*, 199–213.

Guion, R. M. (1991). Personnel Assessment, Selection, and Placement. In M. D. Dunnette & L. M. Hough (Hrsg.), *Handbook of Industrial and Organizational Psychology (2. Aufl., Vol. 2)* (S. 327–397). Palo Alto: Consulting Psychologists Press.

Guion, R. M. (1999). Some virtues of dissatisfaction in the science and practice of personnel selection. *Human Resource Managemen Review*, *8*(4), 351–365.

Guion, R. M. & Gottier, R. F. (1965). Validity of personality measures in personnel selection. *Personnel Psychology*, *18*, 135–164.

Guntsch, M. (2004). *Ant Algorithms in Stochastic and Multi-Criteria Environments.* Dissertation, Universität Fridericana, Karlsruhe.

Haller, H. & Krauss, S. (2002). Misinterpretations of Significance: A Problem Students Share with Theit Teachers? *Methods of Psychological Research Online*, *7*(1), 1–20 [Internet: http://www.mpr-online.de].

Hancock, P. J. B. (2000). A comparison of selection mechanisms. In T. Bäck, D. B. Fogel & Z. Michalewicz (Hrsg.), *Evolutionary Computation 1: Basic Algorithms and Operators* (S. 212–227). Bristol: Institute of Physics Publishing.

Hand, D. J. (2000). Data Mining: New challenges for statisticians. *Social Science Computer Review,*, *18*(4), 442–449.

Harman, G. (2004). Practical Aspects of Theoretical Reasoning. In A. R. Mele & P. Rawling (Hrsg.), *The Oxford handbook of rationality* (S. 45–56). New York: Oxford University Press.

He, S., Wu, Q., Wen, J., Saunders, J. & Paton, R. (2004). A particle swarm optimizer with passive congregation. *BioSystems*, *78*, 135–147.

Heider-Friedel, C., Strobel, A. & Westhoff, K. (2006). Anforderungsprofile zukunftsorientiert und systematisch entwickeln - Ein Bericht aus der Unternehmenspraxis zur Kombination des Bottom-up- und Top-Down-Vorgehens bei der Anforderungsanalyse. *Wirtschaftspsychologie*, *13*(1), 23–31.

Herrera, F., Lozano, M. & Moraga, C. (1999). Hierarchical Distributed Genetic Algorithms. *International Journal of Intelligent Systems*, *14*, 1099–1121.

Hertwig, R. & Hoffrage, U. (2001a). Eingeschränkte und ökologische Rationalität: Ein Forschungsprogramm. *Psychologische Rundschau*, *1*(52), 11–19.

Hertwig, R. & Hoffrage, U. (2001b). Empirische Evidenz für einfache Heuristiken: Eine Antwort auf Bröder. *Psychologische Rundschau*, *52*(3), 162–165.

Hertwig, R. & Hoffrage, U. (2002). Technology needs psychology: How natural frequencies foster insight in medical and legal experts. In P. Sedlmeier & T. Betsch (Hrsg.), *etc: frequency processing and cognition* (S. 285–302). New York: Oxford University Press.

Hertwig, R., Hoffrage, U. & Martignon, L. (1999). Quick Estimation. letting the Environment Do the Work. In G. Gigerenzer, P. M. Todd & the ABC Research Group (Hrsg.), *Simple heuristics that make us smart* (S. 209–234). New York: Oxford University Press.

Hesse, H. (1971). *Lektüre für Minuten.* Frankfurt a. M.: Suhrkamp.

Ho, S.-Y. & Lee, K.-Z. (2003). Design and Analysis of an Efficient Evolutionary Image Segmentation Algorithm. *Journal of VLSI Signal Processing*, *35*, 29–42.

Hoffrage, U. & Gigerenzer, G. (2004). How to Improve the Diagnostic Inferences of Medical Eperts. In E. Kurz-Milcke & G. Gigerenzer (Hrsg.), *Experts in science and society* (S. 249–268). New York: Kluwer.

Hoffrage, U., Hertwig, R. & Gigerenzer, G. (2005). Die ökologische Rationalität einfacher Entscheidungs- und Urteilsheuristiken. In H. Siegenthaler (Hrsg.), *Rationalität im Prozess kultureller Evolution: Rationalitätsunterstellungen als eine Bedingung der Möglichkeit substantieller Rationalität des Handelns* (S. 65–89). Tübingen: Mohr Siebeck.

Hoffrage, U. & Reimer, T. (2004). Models of Bounded Rationality: The Approach of Fast and Frugal Heuristics. *management revue*, *15*(4), 437–459.

Holden, N. & Freitas, A. A. (2004). Web page classification with an ant colony algorithm. In X. Yao, E. Burke, J. A. Lozano, J. Smith, J. J. Merelo-Guervós, J. A. Bullinaria, J. Rowe, P. T. A. Kabán & H.-P. Schwefel (Hrsg.), *Parallel problem solving from nature - ppsn viii* (Bd. 3242, S. 1088–1097). Birmingham, UK: Springer-Verlag.

Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems.* Ann Arbor, MI: Univerity of Michigan Press.

Holland, J. H., Holyoak, K. J., Nisbett, R. E. & Thagard, P. (1986). *Processes of Inference, Learning, and Discovery.* Cambridge: MIT Press.

Hollmann, H. (1991). *Validität in der Eignungsdiagnostik: Neue Ansätze einer sachgerechten Bewertung und effizienten Verbesserung.* Göttingen: Hogrefe.

Hong, T.-P., Huang, K.-Y. & Lin, W.-Y. (2001). Applying genetic algorithms to game search trees. *Soft Computing*, *6*, 277–283.

Horton, P. & Nakai, K. (1996). A Probablistic Classification System for Predicting the Cellular Localization Sites of Proteins. *Intelligent Systems in Molecular Biology*, 149–169.

Hosmer, D. W. & Lemeshow, S. (1989). *Applied logistic regression.* New York: Wiley.

Hossiep, R. & Paschen, M. (1998). *Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung (BIP).* Göttingen: Hogrefe.

Hsu, W. H., Welge, M., Redman, T. & Clutter, D. (2002). High-Performance Commercial Data Mining: A Multistrategy Machine Learning Application. *Data Mining and Knowledge Discovery*, *6*, 361–391.

Hubbard, R., Parsa, R. A. & Luthy, M. R. (1997). The Spread of Statistical Significance Testing in Psychology: The Case of the Journal of Applied Psychology, 1917-1994. *Theory & Psychology*, *7*(4), 545–554.

Hunt, A. & Thomas, D. (2000). *The Pragmatic Programmer: From Journeyman to Master.* Reading, MA: Addison Wesley.

Hutchinson, J. M. C. & Gigerenzer, G. (2005a). Connecting behavioural biologists and psychologists: Clarifying distinctions and suggestions for further work. *Behavioural Processes*, *69*, 159–163.

Hutchinson, J. M. C. & Gigerenzer, G. (2005b). Simple heuristics and rules of thumb: Where psychologists and behavioural biologists might meet. *Behavioural Processes*, *69*, 97–124.

Jackson, J. E. (1991). *A User's Guide to Principal Components.* New York: Wiley.

Jahn, A. (2001). Stochastic optimization of LMS resource assignments. *International Journal of Satellite Communications*, *19*, 141–157.

Jin, A. Y., Leung, F. Y. & Weaver, D. F. (1999). Three Variations of Genetic Algorithm for Searching Biomolecular Conformation Space: Comparison of GAP 1.0, 2.0, and 3.0. *Journal of Computational Chemistry*, *20*(13), 1329–1342.

Jones, T. (1995). *Evolutionary algorithms, fitness landscapes and search.* Dissertation, The University of New Mexico, Albuquerque, NM.

Juslin, P. & Persson, M. (2002). PROBabilities from EXemplars (PROBEX): a „lazy“ algorithm for probabilistic inference from generic knowledge. *Cognitive Science*, *26*, 563–607.

Kahneman, D., Slovic, P. & Tversky, A. (Hrsg.). (1982). *Judgment under uncertainty: Heuristics and biases.* New York: Cambridge University Press.

Kahneman, D. & Tversky, A. (1996). On the Reality of Cognitive Illusions. *Psychological Review*, *103*(3), 582–591.

Kahneman, D. & Tversky, A. (Hrsg.). (2000). *Choices, values and frames.* New York: Cambridge University Press.

Kamphausen, S., Höltge, N., Wirsching, F., Morys-Wortmann1, C., Riester, D., Goetz, R., Thürk, M. & Schwienhorst, A. (2002). Genetic algorithm for the design of molecules with desired properties. *Journal of Computer-Aided Molecular Design*, *16*, 551–567.

Kannan, S., Slochanal, S. M. R., Subbaraj & Padhy, N. P. (2004). Application of particle swarm optimization technique and its variants to generation expansion planning problem. *Electric Power Systems Research*, *70*, 203–210.

Kannetzky, F. (2000). *paradoxes denken: Theoretische und praktische Irritationen des Denkens.* Paderborn: mentis.

Kaplan, A. (1964). *The conduct of inquiry.* San Francisco: Chandler.

Karafyllis, N. C. & Schmidt, J. C. (2002). Rationalität und Pluralität: Versuch einer einführenden Bestimmung im Horizont von Wissenschaft, Kultur und Gesellschaft. In N. C. Karafyllis & J. C. Schmidt (Hrsg.), *Zugänge zur Rationalität der Zukunft* (S. 9–28). Stuttgart: Metzler.

Katsikopolous, K. V. & Martignon, L. (2006). Naïve heuristics for paired comparisons: Some results on their relative accuracy. *Journal of Mathematical Psychology*, *50*, 488–494.

Katsikopoulos, K. V. & Fasolo, B. (2006). New Tools for Decision Analysis. *IEEE Transactions on Ssystems, Man, and Cybernetics - Part A: Systems and Humans*, *36*(5), 960–967.

Katsikopoulos, K. V., Martignon, L. & Woike, J. K. (i. Vorb.). *Categorization without Probabilities and Similarities: A Family of Simple Heuristics* (Zur Publikation eingereichtes Manuskript). Max-Planck-Institut für Bildungsforschung, Berlin.

Katsikopoulos, K. V., Woike, J. K. & Brighton, H. (in Vorb.). Fast and frugal trees: A theory of robust classification. In D. Fum, A. Stocco & F. Del Missier (Hrsg.), *Proceedings of the 7th international conference on cognitive modeling.*

Keane, M. T. (1988). *Analogical Problem Solving.* Chichester: Horwood.

Kennedy, J. & Eberhart, R. C. (1995). Particle swarm optimization. In *Proc. of the IEEE Int. Conf. on Neural Networks* (S. 1942–1948). Piscataway, NJ: IEEE Service Center.

Kennedy, J. & Eberhart, R. C. (2001). *Swarm Intelligence.* San Diego: Academic Press.

Kersting, M. & Hornke, L. F. (2003). Qualitätssicherung und -optimierung in der Diagnostik: die DIN 33430 und notwendige Begleit- und Folgeinitiativen. *Psychologische Rundschau*, *54*(3), 175–184.

Ketelaar, T. & Todd, P. M. (2001). Framing our thoughts: Ecological rationality as evolutionary psychology's answer to the frame problem. In H. R. Holcomb (Hrsg.), *Conceptual challenges in evolutionary psychology. innovative research strategies* (S. 179-211). Dordrecht: Kluwer.

Kirbach, C. & Montel, C. (2002). PERLS: Ein neues System für das Internet-Recruiting. *Wirtschaftspsychologie*, *9*(1), 39–43.

Kivijärvi, J., Fränti, P. & Nevalainen, O. (2003). Self-Adaptive Genetic Algorithm for Clustering. *Journal of Heuristics*, *9*, 113–129.

Klehe, U.-C. (2004). Choosing How to Choose: Institutional Pressures Affecting the Adoption of Personnel Selection Procedures. *International Journal of Selection and Assessment*, *12*(4), 327–342.

Kleinmuntz, B. (1990). Why We Still Use Our Heads Instead of Formulas: Toward an Integrative Approach. *Psychological Bulletin*, *107*(3), 296–310.

Künzel, R. & Wottawa, H. (1985). Hinreichend, notwendig oder korrelativ? bedingungen für das zustandekommen von leidensdruck und therapiemotivation. *Zeitschrift für Diferentielle und Diagnostische Psychologie*, *6*(3), 175–184.

Kohonen, T. (1982a). Clustering, Taxonomy, and Topological Maps of Patterns. In *Proceedings of the IEEE International Conference on Pattern Recognition* (S. 114–128).

Kohonen, T. (1982b). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, *43*, 59–69.

Korczak, J. & Roger, P. (2002). Stock timing using genetic algorithms. *Applied Stochastic Models in Business and Industry*, *18*, 121–134.

Koyamada, K., Yamada, Y., Nishio, T. & Kotera, H. (2001). Compact Modeling Approach Using Genetic Algorithms for Accurate Thermal Simulation. *Heat Transfer - Asian Research*, *30*(1), 28–39.

Kragelund, L. V. (1997). Solving a Timetabling Problem using Hybrid Genetic Algorithms. *Software - Practice and Experience*, *27*(10), 1121–1134.

Krauth, J. (1980). Ein Vergleich der Konfigurationsfrequenzanalyse mit der Methode der log-linearen Modelle. *Zeitschrift für Sozialpsychologie*, *11*, 233–247.

Krivokapich, J., Child, J. S., Walter, D. O. & Garfinkel, A. (1999). Prognostic Value of Dobutamine Stress Echocardiography in Predicting Cardiac Events in Patients With Known or Suspected Coronary Artery Disease. *Journal of the American College of Cardiology*, *33*(3), 708–716.

Kubinger, K. (1999). Forschung in der Psychologischen Diagnostik: Programmatische Betrachtungen. *Psychologische Rundschau*, *50*(3).

Kuo, T. & Hwang, S.-Y. (1993). A genetic algorithm with disruptive selection. In S. Forrest (Hrsg.), *Proc. of the Fifth Int. Conf. on Genetic Algorithms* (S. 65–69). San Mateo, CA: Morgan Kaufmann.

Kurgan, L. A., Cios, K. J., Tadeusiewicz, R., Ogiela, M. & Goodenday, L. S. (2001). Knowledge Discovery Approach to Automated Cardiac SPECT Diagnosis. *Artificial Intelligence in Medicine*, *23*(2), 149–169.

Landy, F. J. & Shankster, L. J. (1994a). Personnel selection and placement. *Annual Review of Psychology*, *45*, 261–296.

Landy, F. J. & Shankster, L. J. (1994b). Personnel Selection and Placement. *Annual Review of Psychology*, *45*, 261–296.

Langton, C. G. (1989a). Artificial Life. In C. G. Langton (Hrsg.), *Artificial Life* (Bd. VI, S. 1–48). Reading, MA: Addison-Wesley.

Langton, C. G. (1989b). Preface. In C. G. Langton (Hrsg.), *Artificial Life* (Bd. VI, S. xv–xxvi). Reading, MA: Addison-Wesley.

Larrañaga, P., Kuijpers, C., Murga, R., Inza, I. & Dizdarevic, S. (1999). Genetic Algorithms for the Travelling Salesman Problem: A Review of Representations and Operators. *Artificial Intelligence Review*, *13*, 129–170.

Lauret, P., Boyer, H., Riviere, C. & Bastide, A. (2005). A genetic algorithm applied to the validation of building thermal models. *Energy and Buildings*, *37*, 858–866.

Lawrenz, C. & Westerhoff, F. (2003). Modeling Exchange Rate Behavior with a Genetic Algorithm. *Computational Economics*, *21*, 209–229.

Leardi, R. (2000). Application of genetic algorithm-PLS for feature selection in spectral data sets. *Journal of Chemometrics*, *14*, 643–655.

Leardi, R. (2001). Genetic algorithms in chemometrics and chemistry: a review. *Journal of Chemometrics*, *15*, 559–569.

Lee, D. (1997). Multiobjective design of a marine vehicle with aid of design knowledge. *International Journal for Numerical Methods in Engineering*, *40*, 2665–2677.

Lee, J., Lee, K. & Rho, S. (2002). An evolutionary perspective on strategic group emergence: a genetic algorithm-based model. *Strategic Management Journal*, *23*, 727–746.

Lee, M. D. & Cummins, T. D. (2004). Evidence accumulation in decision making: Unifying the „take the best“ and the „rational“ models. *Psychonomic Bulletin & Review*, *11*(2), 343–352.

Li, J.-P., Balazs, M. E., Parks, G. T. & Clarkson, P. J. (2002). A Species Conserving Genetic Algorithm for Multimodal Function Optimization. *Evolutionary Computation*, *10*(3), 207–234.

Li, Q. S., Liu, D. K., Zhang, N., Tam, C. M. & Yang, L. F. (2001). Multi-level design model and genetic algorithm for structural control system optimization. *Earthquake Engineering and Structural Dynamics*, *30*, 927–942.

Lipnitskii, A. A. (2002). Use of genetic algorithms for solution of the rectangle packing problem. *Cybernetics and Systems Analysis*, *38*(6), 943–946.

Liu, B., Wang, L., Jin, Y.-H., Tang, F. & Huang, D.-X. (2005). Improved particle swarm optimization combined with chaos. *Chaos, Solitons and Fractals*, *25*, 1261–1271.

Liu, J., Iba, H. & Ishizuka, M. (2001). Selecting Informative Genes with Parallel Genetic Algorithms in Tissue Classification. *Genome Informatics*, *12*, 14–23.

Locke, J. (1690/1998). *An Essay Concerning Human Understanding.* Ware: Wordsworth.

Lovbjerg, M., Rasmussen, T. K. & Krink, T. (2001). Hybrid particle swarm optimiser with breeding and subpopulations. In L. Spector, E. D. Goodman, A. Wu, W. Langdon, H.-M. Voigt, M. Gen, S. Sen, M. Dorigo, S. Pezeshk, M. H. Garzon & E. Burke (Hrsg.), *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2001)* (S. 469–476). San Francisco, California, USA: Morgan Kaufmann.

Lowe, E. J. . (1993). Rationality, deduction and mental models. In K. Manktelow & D. Over (Hrsg.), *Rationality: Psychological and philosophical perspectives* (S. 211–230). London: Routledge.

Luce, R. D. (1964). The Mathematics Used in Mathematical Psychology. *The American Mathematical Monthly*, *71*(4), 364–378.

Luce, R. D. (2000). Fast, frugal, and surprisingly accurate heuristics. *Behavioral and Brain Sciences*, *23*, 757–758.

Mackinnon, M. J. & Glick, N. (1999). Data mining and knowledge discovery in data bases - an overview. *Austral. & New Zealand J. Statist.*, *41*(3), 255–275.

Mahfoud, S. W. (2000). Boltzmann selection. In T. Bäck, D. B. Fogel & Z. Michalewicz (Hrsg.), *Evolutionary Computation 1: Basic Algorithms and Operators* (S. 195–200). Bristol: Institute of Physics Publishing.

Mangasarian, O. L. & Wolberg, W. H. (1990). Cancer diagnosis via linear programming. *SIAM News*, *23*(5), 1–18.

Mark Bower, N. M. amd & Rowe, J. E. (1999). An evolutionary approach to constructing prognostic models. *Artificial Intelligence in Medicine*, *15*, 155–165.

Markham, I. S., Mathieu, R. G. & Wray, B. A. (1998). A rule induction approach for determining the number of Kanbans in a just-in-time production system. *Computers in industrial engineering*, *34*(4), 717–727.

Martignon, L. (2001). Comparing Fast and Frugal Heuristics and Optimal Models. In G. Gigerenzer & R. Selten (Hrsg.), *Bounded Rationality. the Adaptive Toolbox* (S. 147–172). Cambridge: MIT Press.

Martignon, L. & Hoffrage, U. (2002). Fast, frugal and fit: simple heuristics for paired comparison. *Theory and Decision*, *52*, 29–71.

Martignon, L., Katsikopoulos, K. & Woike, J. K. (i. Vorb.). The ecological rationality of fast and frugal trees: A case study in lexicographic models for classification. In P. M. Todd, G. Gigerenzer & the ABC Research Group (Hrsg.), *Ecological rationality: Intelligence in the world.* New York: Oxford University Press.

Martignon, L. & Schmitt, M. (1999). Simplicity and Robustness of Fast and Frugal Heuristics. *Minds and Machines*, *9*, 565–593.

Martignon, L., Vitouch, O., Takezawa, M. & Forster, M. R. (2003). Naive and yet Enlightened: From Natural Frequencies to Fast and Frugal Decision Trees. In D. Hardman & L. Macchi (Hrsg.), *Thinking: Psychological Perspectives on Reasoning, Judgment and Decision Making* (S. 189–211). New York: Wiley and Sons.

Marx, G., Vangerow, B., Hecker, B., Leuwer, M., Jankowski, M., Piepenbrock, S. & Rueckoldt, H. (1998). Predictors of respiratory function deterioration after trransfer of critically ill patients. *Intensive Care Medicine*, *24*, 1157–1162.

Mayley, G. (1996). Landscapes, learning costs and genetic assimilation. *Evolutionary Computation*, *4*(3). (Special edition: Evolution, Learning, and Instinct: 100 Years of the Baldwin Effect)

Mazzeo, S. & Loiseau, I. (2004). An Ant Colony Algorithm for the Capacitated Vehicle Routing. *Electronic Notes in Discrete Mathematics*, *18*.

McKenzie, C. R. M. (2003). Rational models as theories - not standards - of behavior. *TRENDS in Cognitive Sciences*, *7*(9), 403–406.

McLaughlin, A. (1970). Rationality and Total Evidence. *Philosophy of Science*, *37*(2), 271–278.

Medin, D. L. & Schaffer, M. (1978). Context model of classification learning. *Psychological Review*, *85*, 207–238.

Mehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence.* Minneapolis: University of Minnesota Press.

Mehl, P. E. (1957). When shall we use our heads instead of the formula? *Journal of Counseling Psychology*, *4*, 268–273.

Merkle, D. (2002). *Ameisenalgorithmen - Optimierung und Modellierung.* Dissertation, Universität Fridericana, Karlsruhe.

Michalewicz, Z. (1996). *Genetic Algorithms + Data Structures = Evolution Programs (3rd revised and extended edition).* Berlin: Springer.

Michalewicz, Z. (2000a). Introduction to constraint-handling techniques. In T. Bäck, D. B. Fogel & Z. Michalewicz (Hrsg.), *Evolutionary Computation 2: Advanced Algorithms and Operators* (S. 38–40). Bristol: Institute of Physics Publishing.

Michalewicz, Z. (2000b). Other constraint-handling methods. In T. Bäck, D. B. Fogel & Z. Michalewicz (Hrsg.), *Evolutionary Computation 2: Advanced Algorithms and Operators* (S. 69–74). Bristol: Institute of Physics Publishing.

Michalewicz, Z. & Fogel, D. B. (2000). *How to solve it: Modern heuristics.* Berlin: Springer.

Michell, J. (1999). *Measurement in Psychology. Critical History of a Methodological Concept.* Cambridge: Cambridge Universtiy Press.

Miesen, J., Schuhfried, G. & Wottawa, H. (1999). ELIGO: Eine vorläufige Antwort auf Grundproleme der testgestützten Eignungsdiagnostik. *Wirtschaftspsychologie*, *6*(1), 16–24.

Mitchell, M. (1996). *An Introduction to Genetic Algorithms.* Cambridge, MA: MIT Press.

Monett, D., Méndez, J. A., Abraham, G. A., Gallardo, A. & Román, J. S. (2002). An Evolutionary Approach to the Estimation of Reactivity Ratios. *Macromolecular Theory and Simulations*, *11*, 525–532.

Montel, C. (2005). *Empirische Fundierung eignungsdiagnostischer Entscheidungen: Multiple Cutoffs als Basis einer lernfähigen Diagnostik.* Dissertation, Ruhr-Universität Bochum.

Montel, C. & Wottawa, H. (2001). CALL ME - Ein neues Instrument unterstützt die Personalarbeit in Call-Centern. *Wirtschaftspsychologie*, *8*(1), 74–79.

Morton, A. (2000). Heuristics all the way up? *Behavioral and Brain Sciences*, *23*, 758–759.

Mostaghim, S. & Teich, J. (2003). The role of e-dominance in multi objective particle swarm optimization methods. In R. Sarker, R. Reynolds, H. Abbass, K. C. Tan, B. McKay, D. Essam & T. Gedeon (Hrsg.), *Proceedings of the 2003 congress on evolutionary computation cec2003* (S. 1764–1771). Canberra: IEEE Press.

Murphy, K. R. (Hrsg.). (2006). *A Critique of Emotional Intelligence: What are the Problems and How Can They Be Fixed?* Mahwah, NJ: Lawrence Erlbaum.

Nakashima, M., Maruyama, Y., Umeda, N., Kyura, N. & Hasegawa, T. (1998). Basic Experiments on Robot-Base Vibration Control of the Hot-Line Work Robot System Using Genetic Algorithm. *Electrical Engineering in Japan*, *123*(2), 798–807.

Nandy, S., Chaudhury, P. & Bhattacharyya, S. P. (2002). Stochastic Diagonalization of Hamiltonian: A Genetic Algorithm-Based Approach. *International Journal of Quantum Chemistry*, *90*, 188–194.

Neuberger, O. (2002). *Führen und führen lassen* (6., völl. neu bearb. u. erw. Aufl.). Stuttgart: Lucius & Lucius.

Newell, B. R. (2005). Re-visions of rationality? *TRENDS in Cognitive Science*, *9*(1), 11–15.

Newell, B. R., Rakow, T., Weston, N. J. & Shanks, D. R. (2004). Search Strategies in Decision Making: The Success of „Success“. *Journal of Behavioral Decision Making*, *17*, 117–137.

Newell, B. R. & Shanks, D. R. (2003). Take the Best or Look at the Rest? Factors Influencing „One-Reason“ Decision Making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(1), 53–65.

Newell, B. R. & Shanks, D. R. (2004). On the Role of Recognition in Decision Making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(4), 923–935.

Newell, B. R., Weston, N. J. & Shanks, D. R. (2003). Empirical tests of a fast and frugal heuristic: Not everyone „takes-the-best“. *Organizational Behavior and Human Decision Processes*, *91*, 82–96.

Nikolaus von Kues. (1440/1994). De docta ignorantia. In P. Wilpert (Hrsg.), *Die belehrte Unwissenheit: Buch 1* (4., erw. Aufl., S. 3–113). Cambridge, MA: MIT Press.

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*, 104–114.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115*, 39–57.

Novkovic, S. (1998). A Genetic Algorithm Simulation of a Transition Economy: An Application to Insider-Privatization in Croatia. *Knowledge and Information Systems*, *11*, 221–243.

Nozick, R. (1993). *The nature of rationality.* Princeton: Princeton University Press.

Oaksford, M. (2000). Speed, frugality, and the empirical basis of take-the-best. *Behavioral and Brain Sciences*, *23*, 760–761.

Oaksford, M. & Chater, N. (1993). Reasoning theories and bounded rationality. In K. Manktelow & D. Over (Hrsg.), *Rationality: Psychological and philosophical perspectives* (S. 31–60). London: Routledge.

Oaksford, M. & Chater, N. (1998). *Rationality in an Uncertain World: Essays on the Cognitive Science of Human Reasoning.* Hove: Psychology Press.

Oda, A., Nagao, H., Kitagawa, Y., Shigeta, Y. & Yamaguchi, K. (2000). Theoretical Studies on Magnetic Behavior in Clusters by the Genetic Algorithms. *International Journal of Quantum Chemistry*, *80*, 646–656.

Oppenheimer, D. M. (2003). Not so fast! (and not so frugal!): rethinking the recognition heuristic. *Cognition*, *90*, B1-B9.

Ourique, C. O., Biscaia, E. C. & Pinto, J. C. (2002). The use of particle swarm optimization for dynamical analysis in chemical processes. *Computers and Chemical Engineering*, *26*, 1783–1793.

Over, D. E. (2000). Ecological rationality and its heuristics. *Thinking and Reasoning*, *6*(2), 182–192.

Oyama, A., Obayashi, S. & Nakamura, T. (2001). Real-coded adaptive range genetic algorithm applied to transonic wing optimization. *Applied Soft Computing*, *1*, 179–187.

Ozcan, E. & Mohan, C. K. (1999). Particle swarm optimization: Surfing the waves. In P. J. Angeline, Z. Michalewicz, M. Schoenauer, X. Yao & A. Zalzala (Hrsg.), *Proceedings of the congress on evolutionary computation* (Bd. 3, S. 1939–1944). Mayflower Hotel, Washington D.C., USA: IEEE Press.

Özyildirim, S. & Criss, N. B. (2001). Survival of Rationalism Between Hostility and Economic Growth. *Journal of Peace Research*, *38*(4), 515–535.

Pachur, T. & Hertwig, R. (2006). On the Psychology of the Recognition Heuristic: Retrieval Primacy as a Key Determinant of Its Use. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(5), 983–1002.

Pantoja, M. F., García, S. G., Hernández-López, M. A., Bretones, A. R. & Martín, R. G. (2002). Design of an ultra-broadband v antenna for microwave detection of breast tumors. *Microwave and Optical Technology Letters*, *34*(3), 164–166.

Parpinelli, R. S., Lopes, H. S. & Freitas, A. A. (2001). An Ant Colony Based System for Data Mining: Applications to Medical Data. In L. Spector, E. D. Goodman, A. Wu, W. Langdon, H.-M. Voigt, M. Gen, S. Sen, M. Dorigo, S. Pezeshk, M. H. Garzon & E. Burke (Hrsg.), *Proceedings of the genetic and evolutionary computation conference (gecco-2001)* (S. 791–797). San Francisco, California, USA: Morgan Kaufmann.

Payne, J. W., Bettman, J. R. & Johnson, E. J. (1988). Adaptive Strategy Selection in Decision Making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(3), 534–552.

Payne, J. W., Bettman, J. R. & Johnson, E. J. (1993). *The adaptive decision maker.* Cambridge: Cambridge University Press.

Pedersen, J. T. & Moult, J. (1997). Ab Initio Protein Folding Simulations With Genetic Algorithms: Simulations on the Complete Sequence of Small Proteins. *PROTEINS: Structure, Function, and Genetics*, *Suppl. 1*, 179–184.

Pedhazur, E. J. & Schmelkin, L. P. (1991). *Measurement, Design, and Analysis: An Integrated Approach.* Hillsdale, NJ: Erlbaum.

Petricoin, E. F., Paweletz, C. P. & Liotta, L. A. (2002). Clinical Applications of Proteomics: Proteomic Pattern Diagnostics. *Journal of Mammary Gland Biology and Neoplasia*, *7*(4), 433–439.

Pfanzagl, J. (1968). *Theory of Measurement.* Würzburg: Physica.

Pinker, S. (2005). So How Does The Mind Work? *Mind & Language*, *20*(1), 1–24.

Ponterosso, P. & Fox, D. S. J. (2000). Optimization of reinforced soil embankments by genetic algorithm. *International Journal for Numerical and Analytical Methods in Geomechanics*, *24*, 425–433.

Pozen, M., D'Agnostino, R. B., Selker, H. P., Sytkowski, P. A. & Hood, W. B. (1984). A predictive instrument to improve coronary-care-unit admission practices in acute ischemic heart disease. *The New England Journal of Medicine*, *310*, 1273–1278.

Prabhu, D., Buckles, B. P. & Petry, F. E. (2000). Genetic Algorithms for Scene Interpretation from Prototypical Semantic Description. *International Journal of Intelligent Systems*, *15*, 901–918.

Prechelt, L. (1995). Some notes on neural learning algorithm benchmarking. *Neurocomputing*, *9*, 343–347.

Priest, G. (2001). Why It's Irrational to Believe in Consistency. In B. Bogaard & B. Smith (Hrsg.), *Rationality and Irrationality: Proceedings of The 23rd Wittgenstein-Symposium, 13th to 19th August 2000, Kirchberg am Wechsel (Austria)* (S. 284–293). Wien: öbv & hpt.

Putnam, H. (1983). Philosopher and human understanding. In H. Putnam (Hrsg.), *Realism and Reason: Philosophical Papers volume 3* (S. 184-204). Cambridge University Press: Cambridge.

Quine, W. V. (1981). Success and Limits of Mathematization. In W. V. Quine (Hrsg.), *Theories and things* (S. 148–155). Cambridge, MA: Harvard University Press.

Quinlan, J. R. & Rivest, R. L. (1989). Inferring decision trees using the minimum description length principle. *Information and Computation*, *80*, 227–248.

Rainer, T. H., Lam, P. K. W., Wong, E. M. C. & Cocks, R. A. (1999). Derivation of a prediction rule for post-traumatic acute lung injury. *Resuscitation*, *42*, 187–196.

Rajendran, C. & Ziegler, H. (2005). Two ant-colony algorithms for minimizing total flowtime in permutation flowshops. *Computers & Industrial Engineering*, *48*.

Rees, J. & Koehler, G. (2002). Evolution in Groups: A Genetic Algorithm Approach to Group Decision Support Systems. *Information Technology and Management*, *3*, 213–227.

Reimer, T. & Katsikopoulos, K. V. (2004). The use of recognition in group decision-making. *Cognitive Science*, *28*, 1009–1029.

Reinhardt, F. & Soeder, H. (1998). *dtv-Atlas Mathematik* (11 Aufl.). München: Deutscher Taschenbuch Verlag.

Reynolds, C. W. (1987). Flocks, Herds, and Schools: A Distributed Behavioral Model. *Computer Graphics*, *21*(4), 25–34.

Richter, T. & Späth, P. (2006). Recognition Is Used as One Cue Among Others in Judgment and Decision Making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(1), 150–162.

Rieskamp, J. (2006). Perspectives of Probabilistic Inferences: Reinforcement Learning and an Adaptive Network Compared. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(6), 1355–1370.

Rieskamp, J. & Hoffrage, U. (1999). When Do People Use Simple Heuristics and How Can We Tell? In G. Gigerenzer, P. M. Todd & the ABC Research Group (Hrsg.), *Simple heuristics that make us smart* (S. 141–167). New York: Oxford University Press.

Rieskamp, J. & Otto, P. E. (2006). SSL: A Theory of How People Learn to Select Strategies. *Journal of Experimental Psychology: General*, *135*(2), 207–236.

Rivera, W. (2001). Scalable Parallel Genetic Algorithms. *Artificial Intelligence Review*, *16*, 153–168.

Rockey, W. M., Laederach, A. & Reilly, P. J. (2000). Automated Docking of $\alpha - (1 \rightarrow 4)$- and $\alpha - (1 \rightarrow 6)$-Linked Glucosyl Trisaccharides and Maltopentaose Into the Soybean β-Amylase Active Site. *PROTEINS: Structure, Function, and Genetics*, *40*, 299–309.

Rorty, R. (2000). Response to Hilary Putnam. In R. B. Brandom (Hrsg.), *Rorty and his Critics* (S. 87–90). Malden, MA: Blackwell.

Roussev, B. (2006). Object-Oriented Modeling in UML2. In L. Liu & B. Roussev (Hrsg.), *Management of the Object-Oriented Development Process* (S. 1–35). Hershey: Idea Group Publishing.

Roy P. Pargas, M. J. H. & Peck, R. R. (1999). Test-Data Generation Using Genetic Algorithms. *Software Testing, Verification and Reliability*, *9*, 263–282.

Russell, B. (1912/1967). *The Problems of Philosophy.* Oxford: Oxford University Press.

Salman, A., Ahmad, I. & Al-Madani, S. (2002). Particle swarm optimization for task assignment problem. *Microprocessors and Microsystems*, *26*, 363–371.

Salzberg, S. (1988). *Exemplar-based learning: Theory and implementation* (Technical Report TR-10-88). 33 Oxford Street, Cambridge, MA 02138: Harvard University, Center for Research in Computing Technology, Aiken Computation Laboratory.

Samrout, M., Yalaoui, F., Châtelet, E. & Chebbo, N. (2005). New methods to minimize the preventive maintenance cost of series-parallel systems using ant colony optimization. *Reliability Engineering and System Safety*, *89*.

Sarges, W. & Wottawa, H. (2004a). *Handbuch wirtschaftspsychologischer Testverfahren* (2. überarb. u. erw. Aufl.). Lengerich: Pabst.

Sarges, W. & Wottawa, H. (2004b). pro facts-Testsystem. In W. Sarges & H. Wottawa (Hrsg.), *Handbuch wirtschaftspsychologischer Testverfahren* (S. 893–904).

Sarges, W. & Wottawa, H. (2004c). ELIGO-System. In W. Sarges & H. Wottawa (Hrsg.), *Handbuch wirtschaftspsychologischer Testverfahren* (S. 851–863).

Sarges, W. & Wottawa, H. (2004d). PERLS. In W. Sarges & H. Wottawa (Hrsg.), *Handbuch wirtschaftspsychologischer Testverfahren* (S. 879–892).

Sarma, J. & De Jong, K. (2000). Generation gap methods. In T. Bäck, D. B. Fogel & Z. Michalewicz (Hrsg.), *Evolutionary Computation 1: Basic Algorithms and Operators* (S. 205–211). Bristol: Institute of Physics Publishing.

Schimank, U. (2005). *Die Entscheidungsgesellschaft: Komplexität und Rationalität der Moderne.* Wiesbaden: VS Verlag für Sozialwissenschaften.

Schlimmer, J. C. (1987). *Concept acquisition through representational adjustment.* Dissertation, University of California, Irvine, CA.

Schmidt, F. L. & Hunter, J. E. (1983). The Future of Criterion-Related Validity. *Personnel Psychology*, *33*, 41–60.

Schmidt, F. L. & Hunter, J. E. (1998). The Validity and Utility of Selection Methods in Personnel Psychology: Practical and Theoretical Implications of 85 Years of Research Findings. *Psychological Bulletin*, *124*(2), 262–274.

Schreier, M. (2003). Rationalität: Individuelle, 'schmale' und soziale, 'breite' Rationalität. In N. Groeben (Hrsg.), *Zur Programmatik einer sozialwissenschaftlichen Psychologie (Band II): Objekttheoretische Perspektiven, 2. Halbband: Situationsbezug, Reflexivität, Rationalität, Theorieintegration* (S. 107–232). Münster: Aschendorffsche Verlagsbuchhandlung.

Schwarb, T. M. (1996). *Die wissenschaftliche konstruktion der personalauswahl.* München: Hampp.

Searle, J. R. (2001). Why It's Irrational to Believe in Consistency. In B. Bogaard & B. Smith (Hrsg.), *Rationality and Irrationality: Proceedings of The 23rd Wittgenstein-Symposium, 13th to 19th August 2000, Kirchberg am Wechsel (Austria)* (S. 311–324). Wien: öbv & hpt.

Sedlmeier, P. (1996). Jenseits des Signifikanztest-Rituals: Ergänzungen und Alternativen. *Methods of Psychological Research Online*, *1*(4), 41–62 [Internet: http://www.pabst-publishers.de/mpr/].

Sedlmeier, P. (1998). Was sind die guten Gründe für Signifikanztests? *Methods of Psychological Research Online*, *3*(1), 39–42 [Internet: http://www.pabst-publishers.de/mpr/].

Shanks, D. R. & Lagnado, D. (2000). Sub-optimal reasons for rejecting optimality. *Behavioral and Brain Sciences*, *23*, 761–762.

Shelokar, P. S., Jayaraman, V. K. & Kulkarni, B. D. (2004). An ant colony classifier system: application to some process engineering problems. *Computers and Chemical Engineering*, *28*, 1577–1584.

Shi, X., Liang, Y., Lee, H., Lu, C. & Wang, L. (2005). An improved GA and a novel PSO-GA-based hybrid algorithm. *Information Processing Letters*, *93*, 255–261.

Sigillito, V. G., Wing, S. P., Hutton, L. V. & Baker, K. B. (1989). Classification of radar returns from the ionosphere using neural networks. *Johns Hopkins APL Technical Digest*, *10*, 262–266.

Silverstein, A. (1976). On Reductionism. In M. H. Marx & F. E. Goodson (Hrsg.), *Theories in Contemporary Psychology* (S. 407-411). New York: Macmillan.

Simon, H. A. (1955). A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics*, *69*(1), 99–118.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, *63*(2), 129–138.

Simon, H. A. (1967). The logic of heuristic decision making. In N. Rescher (Hrsg.), *The logic of decision and action* (S. 1–20). Pittsburgh: The University of Pittsburgh Press.

Simon, H. A. (1978). Rationality as Process and as Product of Thought. *American Economic Review*, *68*(2), 1–16.

Simon, H. A. (1979). Rational Decision Making in Business Organizations. *American Economic Review*, *69*(4), 493–513.

Simon, H. A. (1986). Rationality in Psychology and Economics. *The Journal of Business*, *59*(4), S209–S224.

Simon, H. A. (1990). Invariants of Human Behavior. *Annual Review of Psychology*, *41*, 1–19.

Simon, H. A. (1992). What is an „explanation" of Behavior? *Psychological Science*, *3*(3), 150–161.

Simon, H. A. (1993). *Homo rationalis: Die Vernunft im menschlichen Leben.* Frankfurt a. M.: Campus.

Simon, H. A. (1995). Artificial intelligence: an empirical science. *Artificial Intelligence*, *77*, 95–127.

Simon, H. A. (1996a). Machine as Mind. In P. J. R. Millican & A. Clark (Hrsg.), *Machines and Thought: The Legacy of Alan Turing Vol. I* (S. 81–102). New York: Oxford University Press.

Simon, H. A. (1996b). *The Sciences of the Artificial* (3. Aufl.). Cambridge, MA: MIT Press.

Simon, H. A. (2000). Bounded Rationality in Social Science: Today and Tomorrow. *Mind & Society*, *1*, 25–39.

Simonoff, J. S. (2003). *Analyzing Categorical Data.* New York: Springer.

Simpson, T. W. & D'Souza, B. S. (2004). Assessing Variable Levels of Platform Commonality Within a Product Family Using a Multiobjective Genetic Algorithm. *Concurrent Engineering: Research and Applications*, *12*(2), 119–129.

Smith, A. E. & Coit, D. M. (2000). Penalty functions. In T. Bäck, D. B. Fogel & Z. Michalewicz (Hrsg.), *Evolutionary Computation 2: Advanced Algorithms and Operators* (S. 41–48). Bristol: Institute of Physics Publishing.

Smith, L. & Gilhooly, K. (2006). Regression Versus Fast and Frugal Models of Decision-Making: The Case of Prescribing for Depression. *Applied Cognitive Psychology*, *20*, 265–274.

Smith, M. (2004). Humean Rationality. In A. R. Mele & P. Rawling (Hrsg.), *The Oxford handbook of rationality* (S. 75–92). New York: Oxford University Press.

Sohn, D. (1998). Statistical Significance and Replicability: Why the Former Does Not Presage the Latter. *Theory & Psychology*, *8*(3), 291–311.

Solimanpur, M., Vrat, P. & Shankar, R. (2005). An ant algorithm for the single row layout problem in flexible manufacturing systems. *Computers & Operations Research*, *32*, 583–598.

Solmajer, T. & Zupan, J. (2004). Optimization algorithms and natural computing in drug discovery. *Drug Discovery Today: Technologies*, *1*(3), 346–354.

Sousa, T., Silva, A. & Neves, A. (2004). Particle Swarm based Data Mining Algorithms for classification tasks. *Parallel Computing*, *30*, 767–783.

Standley, D. M., Gunn, J. R., Friesner, R. A. & McDermott, A. E. (1998). Tertiary Structure Prediction of Mixed $\alpha\backslash\beta$ Proteins Via Energy Minimization. *PROTEINS: Structure, Function, and Genetics*, *33*, 240–252.

Steinberg, M. L. & Page, A. B. (1999). Nonlinear adaptive flight control with genetic algorithm design optimization. *International Journal of Robust and Nonlinear Control*, *9*, 1097–1115.

Stinchcombe, A. L. (1990). Reason and Rationality. In K. Schweers Cook & M. Levi (Hrsg.), *The Limits of rationality* (S. 285–317). Chicago: University of Chicago Press.

Stivers, R. (2004). The Role of Statistics in Management Magic. *Bulletin of Science, Technology & Society,*, *24*(2), 99–106.

Suppes, P. & Zinnes, J. L. (1963). Basic Measurement Theory. In R. D. Luce, R. R. Bush & E. H. Galanter (Hrsg.), *Handbook of Mathematical Psychology, vol. 1* (S. 1–76). New York: Wiley.

Tack, W. H. (1991). Über den Status von Wissen und anderen Kognitionen. In K. Grawe, R. Hänni, N. Semmer & F. Tschan (Hrsg.), *Über die richtige Art, Psychologie zu betreiben* (S. 255–269). Göttingen: Hogrefe.

Thanh Hai, N. (1999). *Erkenntnistheoretische und begriffliche Grundlagen der objektorientierten Datenmodellierung.* Unveröffentlichte diplomarbeit, Universität Leipzig.

Thompson, B. (1995). Stepwise regression and stepwise discriminant analysis need not apply here: a guidelines editorial. *Educational and Psychological Measurement*, *55*(4), 525–534.

Thorndike, E. (1949). *Personnel selection: Test and measurement techniques.* New York: Wiley.

Tiampo, K. F., Rundle, J. B., Hopper, P., Sá Martins, J., Gross, S. & McGinnis, S. (2002). Parallelization of a large-scale computational earthquake simulation program. *Concurrency and Computation: Practice and Experience*, *14*, 531–550.

Todd, P. M., Fiddick, L. & Krauss, S. (2000). Ecological rationality and its contents. *Thinking and Reasoning*, *6*(4), 375–384.

Todd, P. M. & Gigerenzer, G. (1999). What We Have Learned (So Far). In G. Gigerenzer, P. M. Todd & the ABC Research Group (Hrsg.), *Simple heuristics that make us smart* (S. 357–365). New York: Oxford University Press.

Todd, P. M. & Gigerenzer, G. (2003). Bounding rationality to the world. *Journal of Economic Psychology*, *24*, 143–165.

Todd, P. M., Gigerenzer, G. & the ABC Research group. (2000). How can we open up the adaptive toolbox? *Behavioral and Brain Sciences*, *23*, 767–778.

Trelea, I. C. (2003). The particle swarm optimization algorithm: convergence analysis and parameter selection. *Information Processing Letters*, *85*, 317–325.

Tuci, E., Quinn, M. & Harvey, I. (2002). An Evolutionary Ecological Approach to the Study of Learning Behavior Using a Robot-Based Model. *Adaptive Behavior*, *10*(3-4), 201–221.

Turner, S. P. (2003). Cause, the Persistence of Teleology, and the Origins of the Philosophy of Social Science. In S. P. Turner & P. A. Roth (Hrsg.), *The Blackwell Guide to the Philosophy of the Social Sciences* (S. 21–41). Malden, MA: Blackwell.

Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review*, *79*, 281–299.

Tversky, A. & Kahneman, D. (1983). Extensional vs. intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, *91*, 293–315.

Tversky, A., Sattath, S. & Slovic, P. (1988). Contingent weighting in judgment and choice. *Psychological Review*, *95*(3), 371–384.

Tversky, A. & Simonson, I. (1993). Context-Dependent Preferences. *Management Science*, *39*(10), 117–85.

Unwin, A. (2000). Using your eyes - making statistics more visible with computers. *Computational Statistics & Data Analysis*, *32*, 303–312.

Van der Merwe, D. W. & Engelbrecht, A. P. (2003). Data clustering using particle swarm optimization. In R. Sarker, R. Reynolds, H. Abbass, K. C. Tan, B. McKay, D. Essam & T. Gedeon (Hrsg.), *Proceedings of the 2003 congress on evolutionary computation cec2003* (S. 215–220). Canberra: IEEE Press.

van Wert, W. F. (1995-1996). Disney world and posthistory. *Cultural Critique*, *32*(Winter), 187–214.

Varetto, F. (1998). Genetic algorithms applications in the analysis of insolvency risk. *Journal of Banking & Finance*, *22*, 1421–1439.

Viotti, P., Polettini, A., Pomi, R. & Innocenti, C. (2003). Genetic algorithms as a promising tool for optimisation of the MSW collection routes. *Waste Management & Research*, *18*, 121–134.

Wedgwood, R. (1999). The A Priori Rules of Rationality. *Philosophy and Phenomenological Research*, *59*(1), 113–131.

Weick, K. E. (1979). *The Social Psychology of Organizing* (2. Aufl.). New York: McGraw-Hill.

Weijters, T. & Paredis, J. (2002). Genetic rule induction at an intermediate level. *Knowledge-Based Systems*, *15*, 85–94.

Weinert, A. B. (1998). *Organisationspyschologie: Ein Lehrbuch* (4., vollst. überarb. ü. erw. Aufl.).

Weizenbaum, J. (1977). *Die Macht der Computer und die Ohnmacht der Vernunft.* Frankfurt a.M.: Suhrkamp.

Westmeyer, H. (2004). Die sogenannte Krise der psychologischen Diagnostik: Erinnerungen an die 70er Jahre des 20. Jahrhunderts. *Diagnostica*, *50*(1), 10–16.

Whitecotton, S. M., Sanders, D. E. & Norris, K. B. (1998). Improving Predictive Accuracy with a Combination of Human Intuition and Mechanical Decision Aids. *Organizational Behavior and Human Decision Processes*, *76*(3), 325–348.

Whitfield, R. I., Duffy, A. H. B., Coates, G. & Hills, W. (2003). Efficient Process Optimization. *Concurrent Engineering: Research and Applications*, *11*(2), 83–92.

Whitley, D. (2000a). Permutations. In T. Bäck, D. B. Fogel & Z. Michalewicz (Hrsg.), *Evolutionary Computation 1: Basic Algorithms and Operators* (S. 139–150). Bristol: Institute of Physics Publishing.

Whitley, D. (2000b). Recombination. Permutations. In T. Bäck, D. B. Fogel & Z. Michalewicz (Hrsg.), *Evolutionary Computation 1: Basic Algorithms and Operators* (S. 274–284). Bristol: Institute of Physics Publishing.

Wild, B., De la Fontaine, A. & Schafsteller, C. (2001). Fishing for Talents: Internet-Recruiting auf neuen Wegen. *Personalführung*, *1*, 66–70.

Wilson, H. T. (1973). Rationality and Decision in Administrative Science. *Canadian Journal of Political Science*, *6*(2), 271–294.

Witten, I. H. & Frank, E. (2001). *Data Mining: Praktische Werkzeuge und Techniken für das maschinelle Lernen.* München: Hanser.

Wittgenstein, L. (1984a). Zettel. In G. E. M. Anscombe & G. H. von Wright (Hrsg.), *Bemerkungen über die Farben, Über Gewißheit, Zettel, Vermischte Bermerkungen: Werkausgabe Band 8* (S. 259–443). Frankfurt a. M.: Suhrkamp.

Wittgenstein, L. (1984b). Über Gewißheit. In G. E. M. Anscombe & G. H. von Wright (Hrsg.), *Bemerkungen über die Farben, Über Gewißheit, Zettel, Vermischte Bermerkungen: Werkausgabe Band 8* (S. 113–257). Frankfurt a. M.: Suhrkamp.

Wittgenstein, L. (1993). Philosophische Untersuchungen. In J. Schulte (Hrsg.), *Tractatus logico-philosophicus Tagebücher 1914-1916, Philosophische Untersuchungen: Werkausgabe Band 1* (9. Aufl., S. 225–618). Frankfurt a. M.: Suhrkamp.

Wittgenstein, L. (1999). Bemerkungen über die Philosophie der Psychologie (Band 2). In J. Schulte (Hrsg.), *Bemerkungen über die Philosophie der Psychologie: Werkausgabe Band 7* (7. Aufl., S. 217–346). Frankfurt a. M.: Suhrkamp.

Woike, J. K. (2000). *Klassische und konnektionistische Modelle zur Kriteriumsprädiktion durch Testskalen.* Unveröffentlichte Diplomarbeit, Ruhr-Universität Bochum.

Woike, J. K. (2003a). Screening. In K. D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 375–377). Weinheim: Beltz.

Woike, J. K. (2003b). Selektionsorientierte Diagnostik. In K. D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 377–380). Weinheim: Beltz.

Woike, J. K. (2003c). Testökonomie. In K. D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 405–407). Weinheim: Beltz.

Wolpert, D. H. (1992). On the connection between in-sample testing and generalization error. *Complex Systems*, *6*, 47–94.

Wolpert, D. H. (1995). The relationship between PAC, the statistical physics framework, the Bayesian framework, and the VC framework. In D. H. Wolpert (Hrsg.), *The mathematics of generalization* (S. 117–214). Reading, MA: Addison-Wesley.

Wolpert, D. H. (1996). The lack of a priori distinctions between learning algorithms. *Neural Computation*, *8*, 1341–1390.

Woodward, J. & Cowie, F. (2004). The Mind is not (just) a System of Modules Shaped (just) by Natural Selection. In C. Hitchcock (Hrsg.), *Contemporary debates in philosophy of science* (S. 312–334). Malden: Blackwell.

Woolery, L., Grzymala-Busse, J., Summers, S. & Budihardjo, A. (1991). The use of machine learning program LERS_LB 2.5 in knowledge acquisition for expert system development in nursing. *Computers in Nursing*, *9*, 227–234.

Wottawa, H. (1980). *Grundriß der Testtheorie.* München: Juventa Verlag.

Wottawa, H. (1987a). Hypotheses Agglutination (HYPAG): A Method for Configuration-based Analysis of Multivariate Data. *Methodika*, *1*, 68–92.

Wottawa, H. (1987b). Konfigurale auswertungsmethoden in der psychotherapieforschung. *Zeitschrift für Klinische Psychologie, Psychopathologie und Psychotherapie*, *25*(2), 100–113.

Wottawa, H. (1989). Angewandte Diagnostik in der Wirtschaft - Alchemie oder der gläserne Mensch. In G. Cisek, U. Schäkel & J. Scholz (Hrsg.), *Instrumente der Personalentwicklung* (S. 13–38). Hamburg: Windmühle.

Wottawa, H. (2000). Umsetzung von situationsdiagnostischen Erkenntnissen in persondiagnostische Überlegungen. In W. Sarges (Hrsg.), *Management-Diagnostik* (S. 175–193). Göttingen: Hogrefe.

Wottawa, H. & Echterhoff, K. (1982). Formalisierung der diagnostischen Urteilsfindung: Ein Vergleich von linearen und auf Psychologenaussagen gestützten konfiguralen Ansätzen. *Zeitschrift für Differentielle und Diagnostische Psychologie*, *13*, 301–309.

Wottawa, H. & Hossiep, R. (1987). *Grundlagen psychologischer Diagnostik.* Göttingen: Hogrefe.

Wottawa, H. & Hossiep, R. (1997). *Anwendungsfelder psychologischer Diagnostik.* Göttingen: Hogrefe.

Wottawa, H., Kirbach, C., Montel, C. & Oenning, S. (2005). Psychologiegestütztes Recruiting im Internet: Stand und Perspektiven. *Report Psychologie*, *4/2005*, 160–166.

Wottawa, H. & Oenning, S. (2002). Von der Anforderungsanalyse zur Eignungsbeurteilung: Wie praktikabel ist die neue DIN 33430 bei der Bewerberauswahl? *Wirtschaftspsychologie-R*, *4*(3), 43–56.

Wottawa, H. & Thierau, H. (2003). *Lehrbuch Evaluation* (3. korr. Aufl.). Bern: Huber.

Wottawa, H. & Woike, J. K. (2002). Internet-Recruiting und -Assessment: Eine Chance, die Wirtschaftspsychologen nutzen sollten! *Wirtschaftspsychologie*, *9*(1), 33–38.

Yamashita, F., Wanchana, S. & Hashida, M. (2002). Quantitative Structure\ Property Relationship Analysis of Caco-2 Permeability. *Journal of Pharmaceutical Sciences*, *91*(10), 2230–2239.

Yoshioka, M. & Omatu, S. (2000). Independent Signal Separation Using Genetic Algorithm. *Electrical Engineering in Japan*, *131*(4), 1739–1744.

Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer*, *3*, 32–35.

Yu, C. & Gao, D. (1997). High-Resolution Radar Image Reconstruction Based on Genetic Algorithms. *Microwave and Optical Technology Letters*, *16*(5), 290–292.

Yu, X., Xiong, X. & Wu, Y. (2004). A PSO-based approach to optimal capacitor placement with harmonic distortion consideration. *Electric Power Systems Research*, *71*, 27–33.

Zhang, H., Li, X., Li, H. & Huang, F. (2005). Particle swarm optimization-based schemes for resource-constrained project scheduling. *Automation in Construction*, *14*, 393–404.

Zhong, M., Xing, W., Weicheng, F., Peide, L. & Baozhi, C. (2003). Airflow Optimizing Control Research Based on Genetic Algorithm During Mine Fire Period. *Journal of Fire Sciences*, *21*, 131–153.

Zhou, Z.-H. (2003). Three perspectives of data mining. *Artificial Intelligence*, *143*, 139–146.

Zhu, H. & Rohwer, R. (1996). No Free Lunch for Cross-Validation. *Neural Computation*, *10*, 1421–1426.

Zhu, Y., Qiu, J. & Tani, J. (2001). Simultaneous Optimization of a Two-Link Flexible Robot Arm. *Journal of Robotic Systems*, *18*(1), 29–38.

Zhuang, H., Wu, J. & Huang, W. (1997). Optimal Planning of Robot Calibration Experiments by Genetic Algorithms. *Journal of Robotic Systems*, *14*(10), 741–752.

Lebenslauf

Lebenslauf

Persönliche Daten

Name: Jan K. Woike
geboren am 15. 02. 1974 in Hattingen

Schulbildung

1980–1984	Grundschule Oberwinzerfeld in Hattingen
1984–1993	Gymnasium Waldstraße in Hattingen, Abiturnote: 1,0

Zivildienst

07/1993–10/1994	Zivildienst in Essen, Pestalozzi-Schule für Geistigbehinderte

Studium

10/1994–09/1997	Studium der Psychologie, Ruhr-Universität Bochum
10/1996	Vordiplom in Psychologie
10/1997–05/1998	Studium als Special Student (psychology/philosophy/dramatic arts) an der Harvard University (Cambridge, MA, USA)
10/1998–03/2001	Studium der Psychologie sowie der Philosophie/Germanistik/Anglistik an der Ruhr-Universität Bochum
03/2001	Diplom in Psychologie an der Ruhr-Universität Bochum, Gesamtnote: „mit Auszeichnung“
11/2001	Abschluss der Ausbildung zum Theaterpädagogen in Unna

Berufliche Laufbahn

05/2001–01/2003	Ruhr-Universität Bochum, Wissenschaftlicher Mitarbeiter an der Fakultät für Psychologie, Lehrstuhl für Methodenlehre, Diagnostik und Evaluation
02/2003–05/2003	Arizona State University (Tempe, AZ, USA), Kurzzeitdozentur für den German Drama Workshop im German Department gefördert durch den DAAD
04/2005–12/2005	Max-Planck-Institut für Bildungsforschung, Berlin, Mitarbeiter in der Gruppe „Adaptive Behavior and Cognition“
06/2005–12/2006	Universität Erfurt, Wissenschaftlicher Mitarbeiter in der Forschungsgruppe Bildungsmanagement des Zentrums für Lehr-/Lern- und Bildungsforschung (ZLB)
seit 09/2007	Université de Lausanne (Schweiz), Wissenschaftlicher Mitarbeiter an der Faculté des Hautes Etudes Commerciales (HEC), Département de management (DEMA)

Stipendien

04/1997–04/2001	Stipendiat des Cusanuswerks (Grundförderung)
06/2002–05/2005	Stipendiat des Cusanuswerks (Graduiertenförderung)

Fremdsprachen

Englisch, Französisch, Latinum, Graecum

Lausanne, 04. Oktober 2007 Jan K. Woike

Anhang zur Dissertation
- Die Paradoxe Rationalität Einfacher Heuristiken -

Anhang

- A - Ergänzungen zu Partikelschwärmen: Analyse der Dynamik und Erweiterungen
- B - Ergänzungen zu genetischen Algorithmen: TSP und Inselalgorithmen
- C - Entwicklung und Struktur der begleitenden Software
- D - Vollständige Ergebnisse der *Fast and Frugal Trees*

An algorithm has to be seen to be believed.
Donald E. Knuth

A

Ergänzungen zu Partikelschwärmen: Analyse der Dynamik und Erweiterungen

In diesem Anhang soll ein Überblick über Untersuchungen zur Dynamik von Partikelschwärmen gegeben werden. Zudem werden einige Modifikationsmöglichkeiten beschrieben.

A.1 Die Dynamik von Partikelschwärmen in eindimensionalen Suchräumen

Zur Analyse der Dynamik werden einzelne Partikel in einfachen Umgebungen untersucht. Dazu kommt das Softwaremodul *gier.swarmset* zum Einsatz, das die Simulation und Dokumentation der Simulation ermöglicht. Dabei beginnt die Betrachtung mit eindimensionalen Suchräumen mit einem einzelnen gespeicherten und nicht-dynamischen Optimum. Anschließend werden zweidimensionale Suchräume, dann auch mit zwei gespeicherten Positionen untersucht. Diese reduzierte Form des Partikelschwarms kann die Auswirkungen von Parametersetzungen besser demonstrieren als der volle Schwarm.

Im ersten Teil sollen eindimensionale Suchräume betrachtet werden. Beispiele für eindimensionale Suchräume sind etwa die möglichen Werte eines einzelnen Parameters bei der Funktionsoptimierung. Immer dann, wenn die Verteilung der Parameterwerte einen einzigen Freiheitsgrad aufweist, d.h. aus der Festlegung eines Parameters alle notwendigen Variablen definiert werden können, werden eindimensionale Suchräume betrachtet.

A.1.1 Das vereinfachte System

Um das Verhalten der Partikel zu veranschaulichen, soll zunächst ein stark vereinfachtes System betrachtet werden. Der Suchraum wird dabei auf eine einzige Dimension eingeschränkt, und es soll davon ausgegangen werden, dass ein einziger Partikel sich isoliert bewegt, und dementsprechend nur ein Bezugswert für die Bewegung berücksichtigt werden muss. Zusätzlich wird dieser Bezugswert als statisch angenommen. Die Anfangsgeschwindigkeit des Partikels wird auf 0 gesetzt. Das Verhalten des Partikels über mehrere Zeitschritte soll in ein Diagramm überführt werden, in das den Zeitschritten zugeordneten Positionswerte als Punkte eingetragen werden, die aus Gründen der Verdeutlichung von Bewegungstendenzen mit Linien verbunden werden sollen. Die Bewegung bleibt aufgrund der Modellgleichungen diskret, die Zwischenzustände sind lediglich linear extrapoliert, haben für das Modell und den Algorithmus keine Bedeutung. Das Optimum wird auf den Nullpunkt der Dimension gesetzt, so dass gleich große Abweichungen in beide Richtungen den gleichen Betrag haben. Um zu verdeutlichen, welche Teile des Suchraums geprüft werden, wird eine vertikale Linie für jede Prüfstelle eingezeichnet. Diese Partikelbewegung entspricht nicht der Partikelbewegung in Partikelschwärmen, da hier zum einen zwei Optima betrachtet werden, zum anderen die Position dieser Optima dynamisch ist. Mit diesem einfachen Modell soll aber eine erste Analyse vorgenommen werden, um das Verhalten von Schwärmen bei verschiedenen Ausgangsbedingungen zu erfassen.

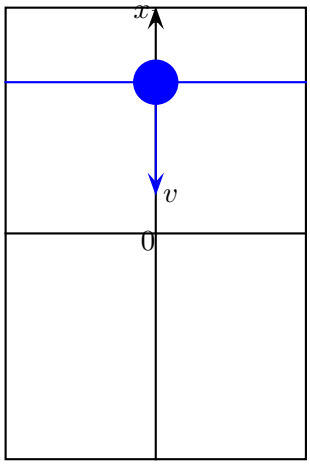


Abbildung A.1: Darstellung des eindimensionalen Simulationssystems

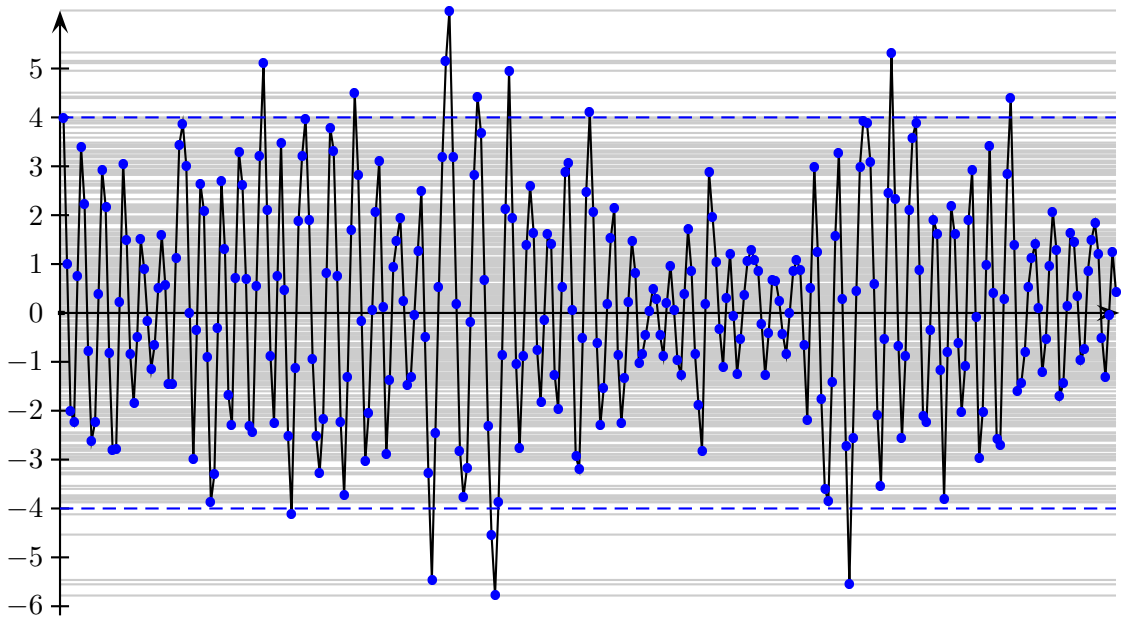


Abbildung A.2: Ergebnis der Schwarmsimulation im eindimensionalen Suchraum ($\varphi_{max} = 2$, $V_{max} = 3$, Startpunkt: 4, Gespeichertes Optimum: 0, Anfangsgeschwindigkeit: 0, Zeitschritte: 300)

Die Simulationen wurden in Delphi programmiert. Das Modul *gier.swarmtest* beinhaltet eine eindimensionale und eine zweidimensionale Simulation mit Visualisierung der Partikelbewegung. Die Positionsverlaufsdiagramme sind in das LaTeX-Format exportierbar. Die Parameter für die eindimensionale Version sind: φ_{max}, V_{max} (s.u.), Position des Optimums, Ausgangsposition, sowie Darstellungsoptionen für die Diagramme. Abbildung A.2 zeigt ein Beispiel für das Diagramm einer Partikelbewegung (Ausgangsposition: 4, Optimum: 0, $\varphi_{max} = 2$, $V_{max} = 3$) für 300 Zeitschritte. Es ist zu erkennen, dass der Partikel den optimalen Punkt in beide Richtungen überschreitet, dabei braucht es zuweilen mehrere Zeitschritte bis sich die Bewegungsrichtung umkehrt. Da in die Berechnung der Beschleunigung der Abstand vom Optimum einfließt, der mit einer Zufallszahl multipliziert wird, ist die Bewegung nicht-deterministisch. Durch die Beschränkung der Geschwindigkeit nach oben ist aber zu erwarten, dass sich der Partikel im größten Teil der Zeitschritte sich in einem engen, aber nicht kleiner werdenden Intervall um 0 befindet, die Dichte der Linien wird gegen 0 hin größer, nur wenige Punkte liegen außerhalb des Intervalls $[-4; 4]$.

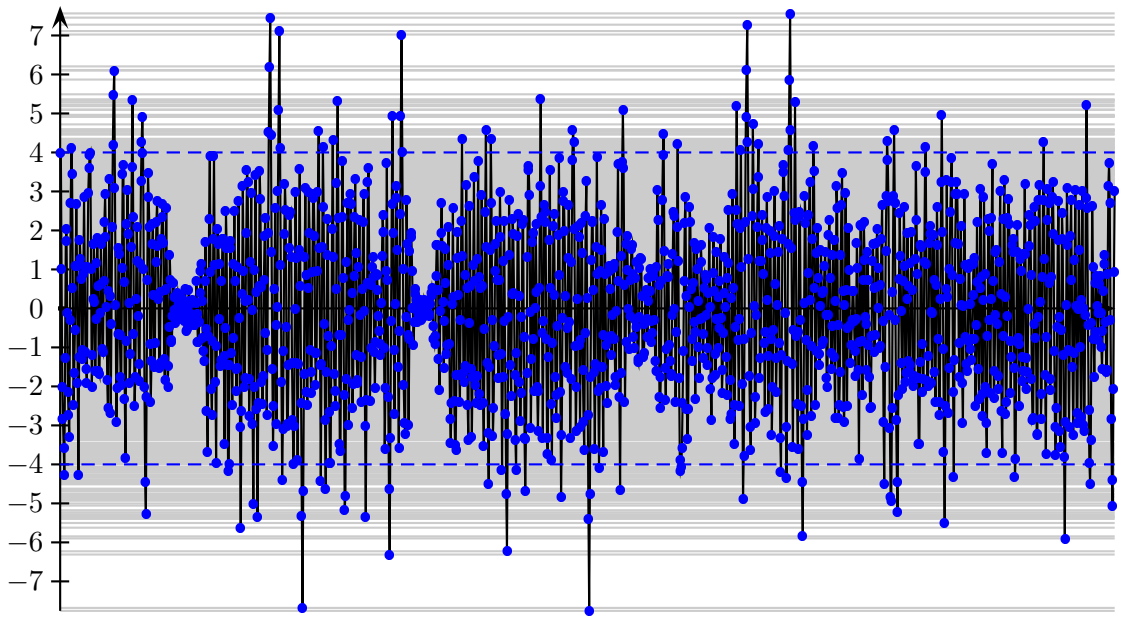


Abbildung A.3: Ergebnis der Schwarmsimulation im eindimensionalen Suchraum ($\varphi_{max} = 2$, $V_{max} = 3$, Startpunkt: 4, Gespeichertes Optimum: 0, Anfangsgeschwindigkeit: 0, Zeitschritte: 1500)

In Abbildung A.3 ist eine Simulation mit gleichen Parameterwerten über 1500 Zeitschritte abgebildet. Das Verhalten des Partikels ist in jedem Zeitsegment ähnlich, durch die Variationen ist aber zu erwarten, dass eine immer größere Anzahl verschiedener Punkte um 0 herum untersucht wird.

A.1.2 Auswirkungen der Wahl von φ_{max} und V_{max} bei randomisierter Beschleunigung

Um einen Eindruck zu geben von der Empfindlichkeit des Partikelverhaltens bei Änderung der Parameterwerte sind zunächst zwei Variationen zu Abbildung A.2 simuliert worden. Abbildung A.4 zeigt die Konsequenzen einer Setzung von V_{max} auf den Wert $0,8$. Der Intuition entsprechend werden die Bewegungsschritte des Partikels durchschnittlich kleiner, die meisten Werte liegen jetzt im Intervall [-1,1].

In einer zweiten Variation wird der Parameter φ_{max} auf den Wert $0,1$ gesetzt, während V_{max} erneut den Wert 3 annimmt. Abbildung A.5 demonstriert die Konsequenzen dieser Wahl an einem Beispiel. Der niedrige Wert von φ_{max} verhindert eine große Beschleunigung bei geringer Distanz zu 0. Aus diesem Grund entfernen sich die Partikel langsam, aber über mehrere Zeitschritte hinweg von 0 und werden dabei abgebremst, bis sich die Bewegungsrichtung umkehren kann, und eine neue langsame Beschleunigung zum Nullpunkt beginnt. In der Konsequenz wird ein größeres Intervall abgesucht, als in den Simulationen zuvor, die Dichte der Suchpunkte nimmt dementsprechend aber ab. Da der Richtungswechsel der Parameter in vielen Bewegungsschritten erfolgt, die keine große Positionsveränderung aufweisen, und diese Richtungswechsel in ähnlicher Entfernung vom Nullpunkt passieren, gibt es zudem eine höhere Suchdichte als zuvor in zwei vom Optimum entfernten Bereichen.

Die Notwendigkeit für die Setzung von V_{max} demonstriert schließlich AbbildungA.6. Wird die Geschwin-

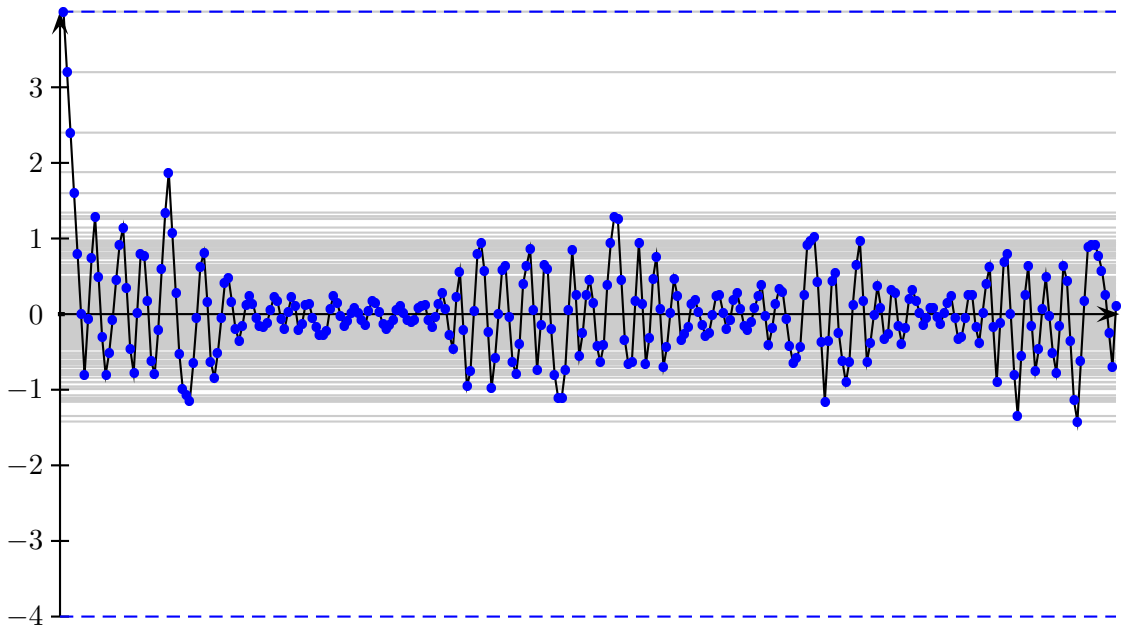


Abbildung A.4: Ergebnis der Schwarmsimulation im eindimensionalen Suchraum ($\varphi_{max} = 2$, $V_{max} = 0,8$, Startpunkt: 4, Gespeichertes Optimum: 0, Anfangsgeschwindigkeit: 0, Zeitschritte: 300)

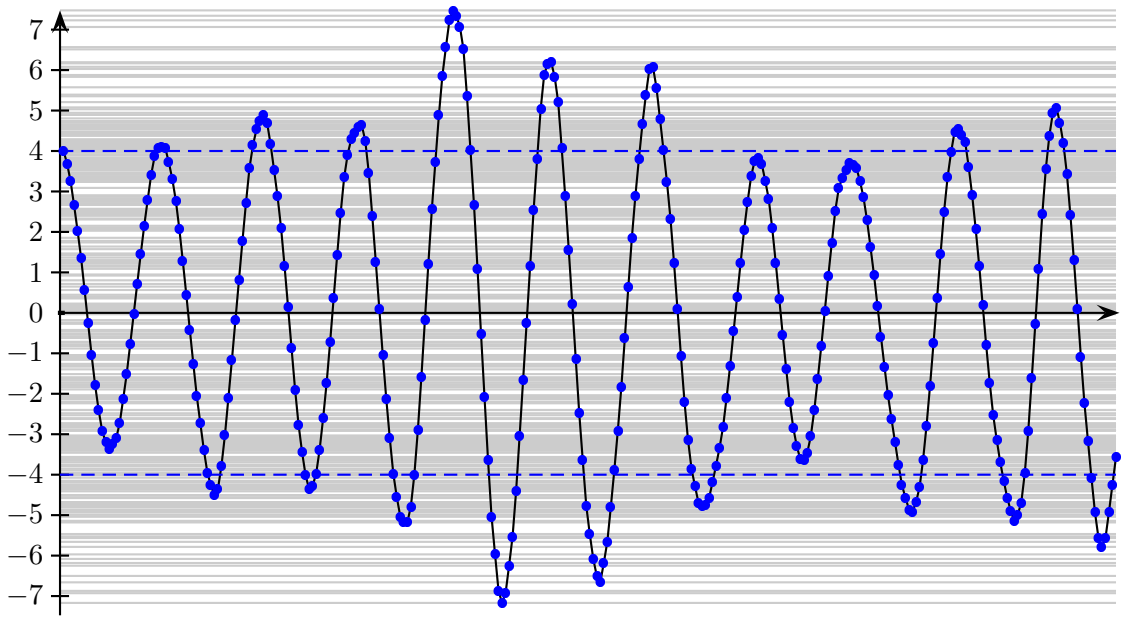


Abbildung A.5: Ergebnis der Schwarmsimulation im eindimensionalen Suchraum ($\varphi_{max} = 0,1$, $V_{max} = 3$, Startpunkt: 4, Gespeichertes Optimum: 0, Anfangsgeschwindigkeit: 0, Zeitschritte: 300)

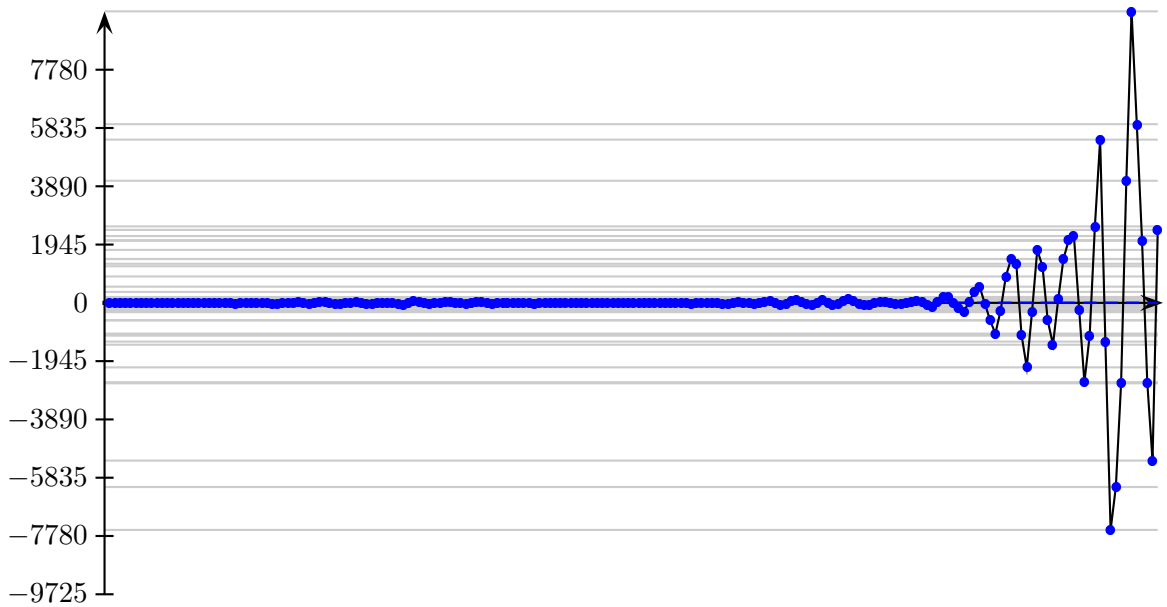


Abbildung A.6: Ergebnis der Schwarmsimulation im eindimensionalen Suchraum ($\varphi_{max} = 2$, V_{max} nicht festgelegt, Startpunkt: 4, Gespeichertes Optimum: 0, Anfangsgeschwindigkeit: 0, Zeitschritte: 200)

digkeit der Partikel nicht reguliert, so steigt die Geschwindigkeit der Partikelbewegung kontinuierlich an, was durch die Randomisierung der Beschleunigung erklärt werden kann, die Partikel entfernen sich immer weiter vom optimalen Punkt, und werden nach Umkehr der Bewegungsrichtung über mehr Zeitschritte als zuvor beschleunigt. Die Simulation wurde auf 200 Zeitschritte beschränkt, der erste Teil des Diagramms wirkt nur deshalb flach, weil die Positionsdistanzen im Laufe der Zeit sehr stark ansteigen.

Um einen Eindruck zu gewinnen von der Interaktion zwischen φ_{max} und V_{max} in Bezug auf die Charakteristika der Verlaufsdiagramme, wurden die beiden Parameter kovariiert. Abbildung A.7 zeigt eine Matrix von Diagrammen für die Kombinationen aus drei verschiedenen Werten für V_{max} in den Spalten und fünf verschiedene Werte von φ_{max} in den Zeilen. Bei sehr niedrigen Werten von φ ist die Beschleunigung der Partikel sehr gering, die Geschwindigkeit wächst langsam an und wird ebenso langsam abgebremst, die wachsende Entfernung vom Optimum stellt aber die Bewegungsumkehr auf lange Sicht sicher. V_{max} muss sehr klein gewählt werden, um das Verhalten der Partikel überhaupt zu beeinflussen, da keine großen Gesamtgeschwindigkeiten erreicht werden. Bei höheren Werten von φ_{max} spielt V_{max} eine größere Rolle. Ein Wert von $0,5$ führt zu einer Einengung der Streubreite der Partikel auf einen Bereich um das Optimum herum. Allein bei sehr kleinen Werten von φ_{max} tritt dieser Effekt nicht auf, die Entfernung wird durch die lange Dauer der Geschwindigkeitsverringerung erreicht. Je höher φ_{max} gesetzt wird, desto größer wird die Wahrscheinlichkeit, dass durch die Beschleunigung der maximale Geschwindigkeitswert überschritten wird und die Partikel bewegen sich in mehr und mehr Zeitschritten bei maximal zulässiger Geschwindigkeit. Da die Seiten hierbei alternieren wird eine große Entfernung vom gespeicherten Optimum seltener und seltener erreicht. Geht φ_{max} gegen unendlich, dann würde der Partikel je nach Ausgangsposition nur noch zwei oder drei Stellen periodisch ansteuern.

Hohe Werte von V_{max} erlauben dem Partikel größere Entfernungen vom Bezugspunkt. Da φ_{max} sowohl die positive als auch die negative Beschleunigung beeinflusst, ist der Wechselwirkung zwischen φ_{max} und V_{max} komplexer als intuitiv zu erwarten wäre. Es gilt nicht, dass höhere Werte von φ_{max} automatisch zu einer höheren Wahrscheinlichkeit großer Entfernungen vom Bezugspunkt führen.

A.1.3 Auswirkungen der Wahl von φ_{max} und V_{max} bei nichtrandomisierter Beschleunigung

Die bisher betrachteten Diagramme variieren bei jeder Simulation unter Umständen stark, da die in die Beschleunigung eingehende Zufallsvariable den Verlauf der Bewegung mitbestimmt. Aus diesem Grund ist die Simulation noch einmal wiederholt worden, wobei für jeden Zeitschritt φ auf φ_{max} gesetzt wurde. Abbildung A.8 demonstriert die Auswirkung dieser Veränderung an einem Beispiel. Der Verlauf der Bewe-

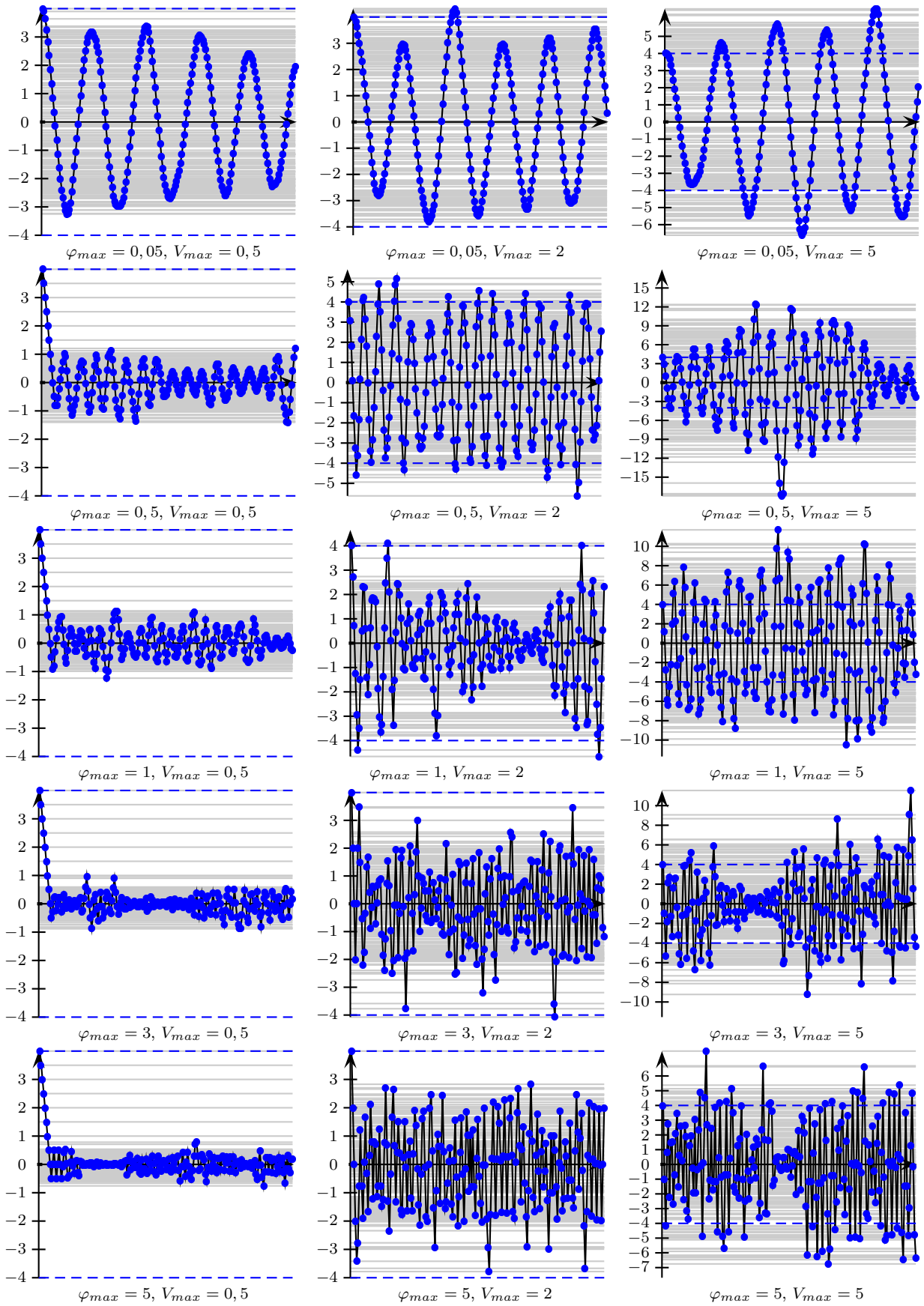


Abbildung A.7: Diagramm der Schwarmsimulation im eindimensionalen Suchraum für fünfzehn verschiedene Kombinationen von φ_{max} und V_{max} (t=200, Optimum: 0, Startposition:3, Anfangsgeschwindigkeit: 0)

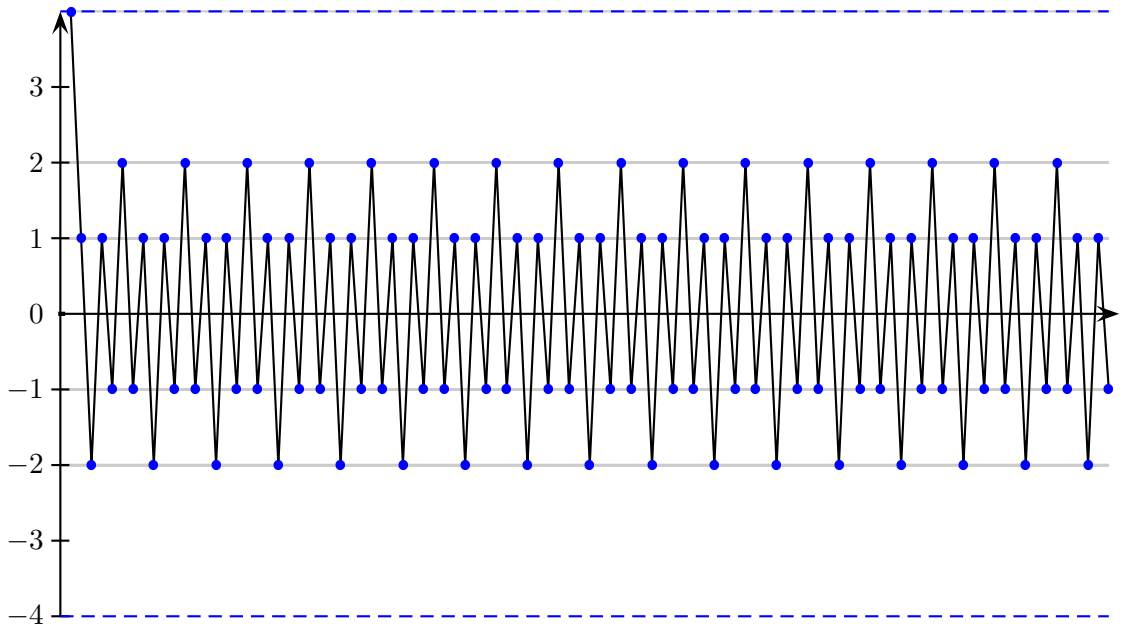


Abbildung A.8: Ergebnis einer nichtrandomisierten Schwarmsimulation im eindimensionalen Suchraum ($\varphi = 5$, $V_{max} = 3$, Startpunkt: 4, Gespeichertes Optimum: 0, Anfangsgeschwindigkeit: 0, Zeitschritte: 100)

gung ist periodisch ab dem zweiten Punkt. Der erst Schritt geschieht mit Maximalgeschwindigkeit, gefolgt von einer maximalen Abbremsung und Beschleunigung auf Maximalgeschwindigkeit in Gegenrichtung. An dieser Stelle ist der Partikel zu nah am Optimum, um die maximale Gegengeschwindigkeit erreichen zu können, dies gelingt erst beim nächsten Schritt, weil die Ausgangsgeschwindigkeit hier niedriger ist. Die nächsten Schritte sind symmetrisch zu den vorangegangenen. Es zeigt sich, dass die Regelmäßigkeit der Kurve auf einem deterministischen, aber komplexen Prozess beruht. In gleicher Weise wurden die Zufallseinflüsse aus den Simulationen entfernt, deren Diagramme in Abbildung A.7 abgebildet sind.

Die aus der Umstellung resultierenden neuen Diagramme sind in Abbildung A.9 enthalten. Es wird deutlich, dass auch der Startpunkt bei dieser Form der Simulation eine Rolle spielt, da das periodische Verhalten entscheidend von der Entfernung vom gespeicherten Optimum abhängig ist. Aus diesem Grund lassen sich die Verläufe in Diagrammen, die eine periodisch wiederkehrende Bewegung zeigen mit zwei oder drei wiederkehrenden Punkten, nur vorsichtig generalisieren. Bei einigen Diagrammen wird aufgrund der Teilbarkeit des Ausgangsabstandes durch den Parameter der Nullpunkt erreicht, so dass die Periode drei verschiedene Punkte aufweist. Eine geringfügige Veränderung des Startpunktes würde die Anzahl der Punkte verändern.

A.2 Surfende Partikel

Ozcan und Mohan (1999) analysieren die Trajektorien von Partikeln in Schwärmen. Sie kommen zu dem Ergebnis, dass sich Partikel auf sinusförmigen Bahnen bewegen, deren Form durch den Wert des Parameters φ bestimmt wird. Da dieser Wert zufallsabhängig ist, ändert sich die Bahn des Partikels mit jedem Schritt, so dass in Analogie zu einem Surfer, der Wellen zur Bewegung nutzt, aber die Wellen kontinuierlich wechselt, davon gesprochen werden kann, dass die Partikel den Suchraum durchsurfen, wobei Sinuswellen ihre Bewegung formen.

> We have shown that in the general case, a particle does not „fly“ in the search space, but rather „surfs“ it on sine waves. An optimal location seeking particle attempts to „catch“ another wave randomly, by manipulating its frequency and amplitude. A wave with a high frequency or amplitude is not necessarily a good wave for a particle. Limiting the velocity (V_{max}) seems to help the particle to „jump“ onto another wave. Velocity cut-off occurs only in regions where the amplitude of the sine wave is large. (Ozcan & Mohan, 1999, S. 1943)

Die Begrenzung der Maximalgeschwindigkit führt zu einem Bruch im Muster der Beschleunigungen und die Bewegung kann einer neuen Form folgen, eine Verhinderung dieses Bruchs etwa durch die Wahl eines Inertia-Gewichts sehen die Autoren dementsprechend eher als potenziell schädlich an. Um einen Eindruck zu geben von der idealen Form der Wellen, die Partikel nutzen können, sind in Abbildung A.10 einige Formen für verschiedene Werte von φ zusammengestellt.

Eine radikale Veränderung der Kurven tritt bei $\varphi = 4$ auf. Ab diesem Wert entfernen sich die Partikel immer weiter von der Optimalachse. Für sehr kleine Werte von φ ist die Frequenz der Periode niedrig. Aus diesem Grund führt die Wahl von $\varphi = 4$ zu einem Bewegungsverhalten, das zwischen den abgebildeten Varianten changiert, was dem gewünschten Verhalten für die Suche entspricht. Ozcan und Mohan (1999) schlagen Erweiterungen des Algorithmus vor, in denen Partikel mit einem virtuellen Sensor ausgestattet werden, um ungünstige Regionen des Suchraums kontrolliert verlassen zu können. Dies wäre der Übergang von einem zufälligen Wechsel der Wellen zu einem kontrollierteren zielgebundenen Wechseln.

A.3 Die Dynamik von Partikelschwärmen in zweidimensionalen Suchräumen

Die Erweiterung der Simulation um eine zweite Dimension lässt keine gravierende Veränderung der Dynamiken erwarten, da jede Komponente von Geschwindigkeits- und Positionsvektoren unabhängig voneinander berechnet werden.

A.3.1 Das erweiterte vereinfachte System

Das zweidimensionale System unterscheidet sich von dem zuvor betrachteten System lediglich in der Erweiterung des Suchraums um eine zweite Dimension. Damit können Suchprobleme abgebildet werden, deren mögliche Parameterverteilungen zwei Freiheitsgrade aufweisen, also zum Beispiel Funkionsoptimierungen mit zwei unabhängig voneinander zu bestimmenden Parametern. Damit haben alle Positions- und Geschwindigkeitsvektoren zwei Komponenten.

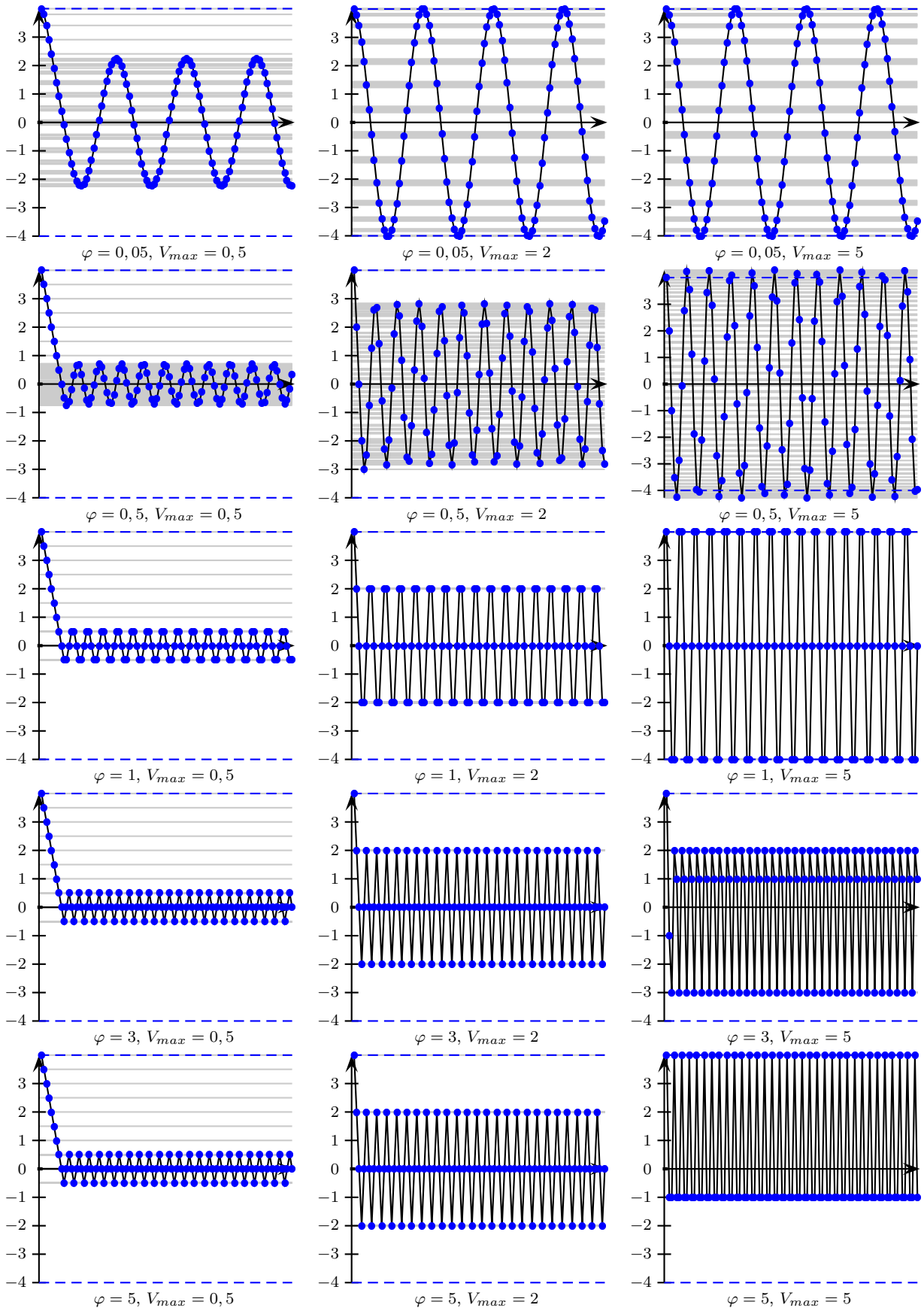


Abbildung A.9: Diagramme der nichtrandomisierten Schwarmsimulation im eindimensionalen Suchraum für fünfzehn Kombinationen von $\varphi = \varphi_{max}$ und V_{max} (t=100, Optimum: 0, Startposition:3, Anfangsgeschwindigkeit: 0)

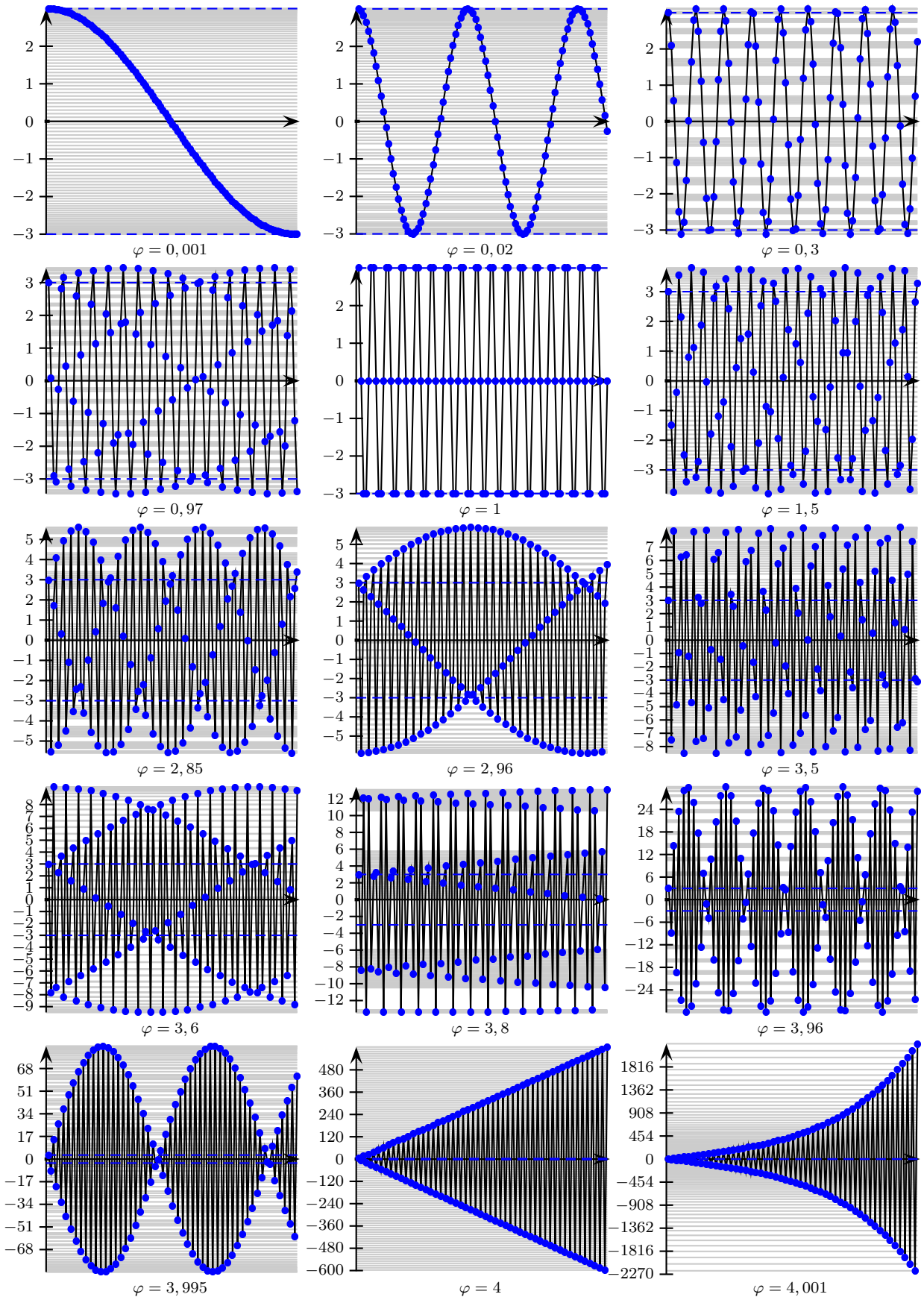


Abbildung A.10: Diagramme der nichtrandomisierten Schwarmsimulation im eindimensionalen Suchraum für fünfzehn verschiedene Werte von $\varphi = \varphi_{max}$ ohne Festlegung von V_{max} (t=100, Optimum: 0, Startposition:3, V_{max} nicht festgelegt, Anfangsgeschwindigkeit: 0)

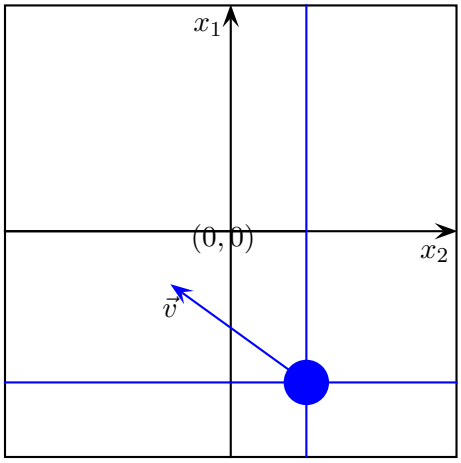


Abbildung A.11: Darstellung eines Partikels im zweidimensionalen Suchraum: Der Vektor zeigt die Geschwindigkeit des Partikels an, die horizontale bzw. vertikale Linie markiert die x_1-Position bzw. x_2-Position

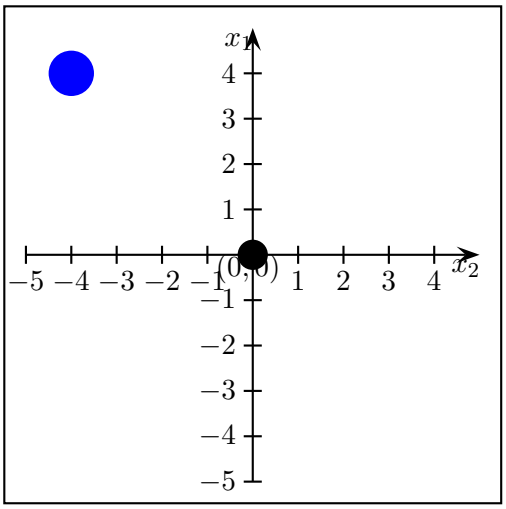


Abbildung A.12: Ausgangsbedingung für den Partikel im zweidimensionalen Suchraum mit einem Optimum

A.3.2 Dynamisches Verhalten bei einer gespeicherten Optimalposition

Zunächst soll das Verhalten eine Schwarms in einem System mit einem einzigen gespeicherten Optimum bzw. mit zwei deckungsgleichen gespeicherten Positionen betrachtet werden. Ein solches System ist in Abbildung A.11 abgebildet. Ausgangspunkt der Bewegung ist der Punkt $(-4; 4)$, die Optimalstelle liegt bei $(0; 0)$, die Anfangsgeschwindigkeit soll $\vec{v} = \binom{0}{0}$ sein. Abbildung A.13 zeigt das Ergebnis der Simulation, aufgeschlüsselt für die Bewegung in der ersten und der zweiten Dimension sowie der Bewegungslinie im zweidimensionalen Raum. Da der Ausgangspunkt in beiden Dimensionen exakt gleich weit vom gespeicherten Optimum entfernt ist, wird der Partikel auch in beiden Dimensionen in identischer Weise beschleunigt, so dass es in der gesamten Simulation keinen Aufenthaltspunkt mit unterschiedlichen Werten in den beiden Dimensionen geben kann. Dies verdeutlicht, dass durch φ zwar die Beschleunigung randomisiert wird, sie aber trotzdem in jedem Zeitschritt in gleicher Weise in allen Dimensionen zum Tragen kommt.

Die Verhältnisse verändern sich, wenn der Ausgangspunkt der Simulation versetzt wird (s. Abbildung A.14). Wäre kein Wert für V_{max} vergeben, so würde sich der Partikel nicht von der Gerade durch Ausgangspunkt und gespeicherte Stelle entfernen. Statt dessen wird der Partikel aber bereits im ersten Schritt auf einer der beiden Dimensionen abgebremst. Aus diesem Grund entsteht ein verändertes Bewegungsverhalten in der Simulation. Auch wenn die Tendenz zur Ausrichtung an Linien durch die Speicherstelle bestehen bleibt, sorgt das Ausbremsen für eine weitere Exploration des Suchraums.

Wird auch im zweidimensionalen Fall die Zufallskomponente beseitigt, indem $\varphi = \varphi_{max}$ gesetzt wird,

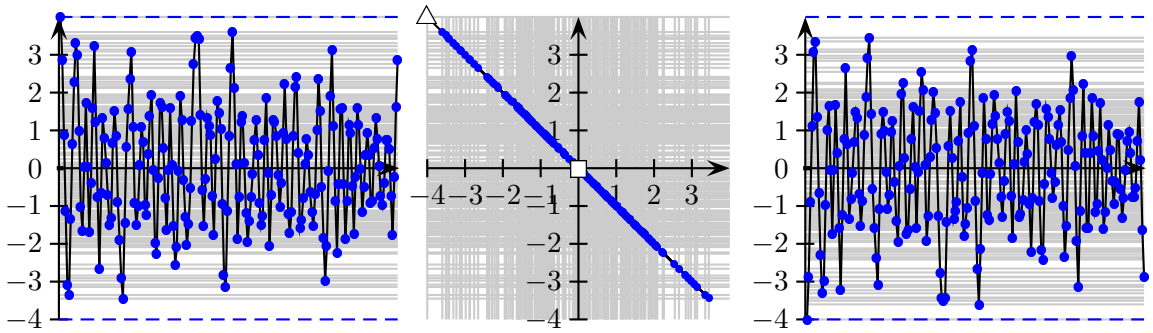


Abbildung A.13: Ergebnis der Schwarmsimulation im zweidimensionalen Suchraum ($\varphi_{max} = 2$, $V_{max} = 2$, Startpunkt: $(-4; 4)$, Gespeichertes Optimum: $(0; 0)$, Anfangsgeschwindigkeit: $(0; 0)$, Zeitschritte: 200)

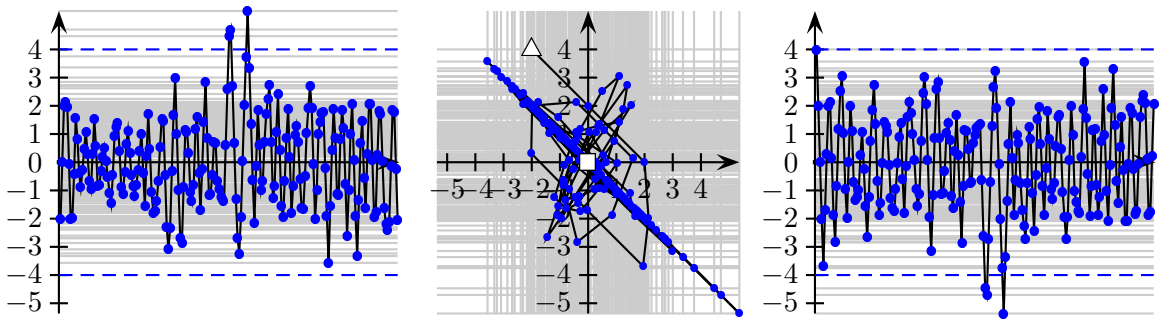


Abbildung A.14: Ergebnis der Schwarmsimulation im zweidimensionalen Suchraum ($\varphi_{max} = 2$, $V_{max} = 2$, Startpunkt: $(-2; 4)$, Gespeichertes Optimum: $(0; 0)$, Anfangsgeschwindigkeit: $(0; 0)$, Zeitschritte: 200)

ergeben sich interessante Variationen der Trajektorien für verschiedene Werte von φ. In diesem Fall wurde V_{max} verwendet, damit die Trajektorien mehr als eindimensional verlaufen können.

A.3.3 Untersuchungen mit zwei gespeicherten Optimalpositionen

Die Erweiterung besteht in der Verwendung zweier Speicherstellen. Dabei werden $\varphi_{b_{max}}$ und $\varphi_{n_{max}}$ jeweils auf den Wert $\frac{\varphi_{max}}{2}$ gesetzt. Die beiden Optimalstellen liegen nun bei $(-4;-1)$ und $(4;1)$, die Startposition und -geschwindigkeit bleiben unverändert (s. Abbildung A.16).

Bei randomisierter Beschleunigung bewegen sich die Partikel auf den einzelnen Dimensionen unabhängig voneinander. Dies ergibt sich aus der separaten Beschleunigung und der separaten Überprüfung der Maximalgeschwindigkeit für jede Dimension. Die entstehenden Diagramme sind also Zusammensetzungen aus mehreren Einzelbewegungen, im zweidimensionalen Fall die Zusammensetzung der Bewegung auf der x_1-Achse und der x_2-Achse. Die entstehende Dynamik ist vergleichbar mit der Dynamik im Fall mit einer einzigen Speicherstelle, mit dem Unterschied, dass durch die randomisierte Gewichtung der Bezugspunkt immer zwischen den beiden gespeicherten Stellen liegt. Dementsprechend sind die entstehenden Diagramme je nach Entfernung der beiden Punkte unruhiger (eine größere Entfernung lässt eine größere Varianz der Positionswerte erwarten). Die Empfindlichkeit für verschiedene Werte von φ_{max} und V_{max} ist hingegen ähnlich (s. Abbildung A.17).

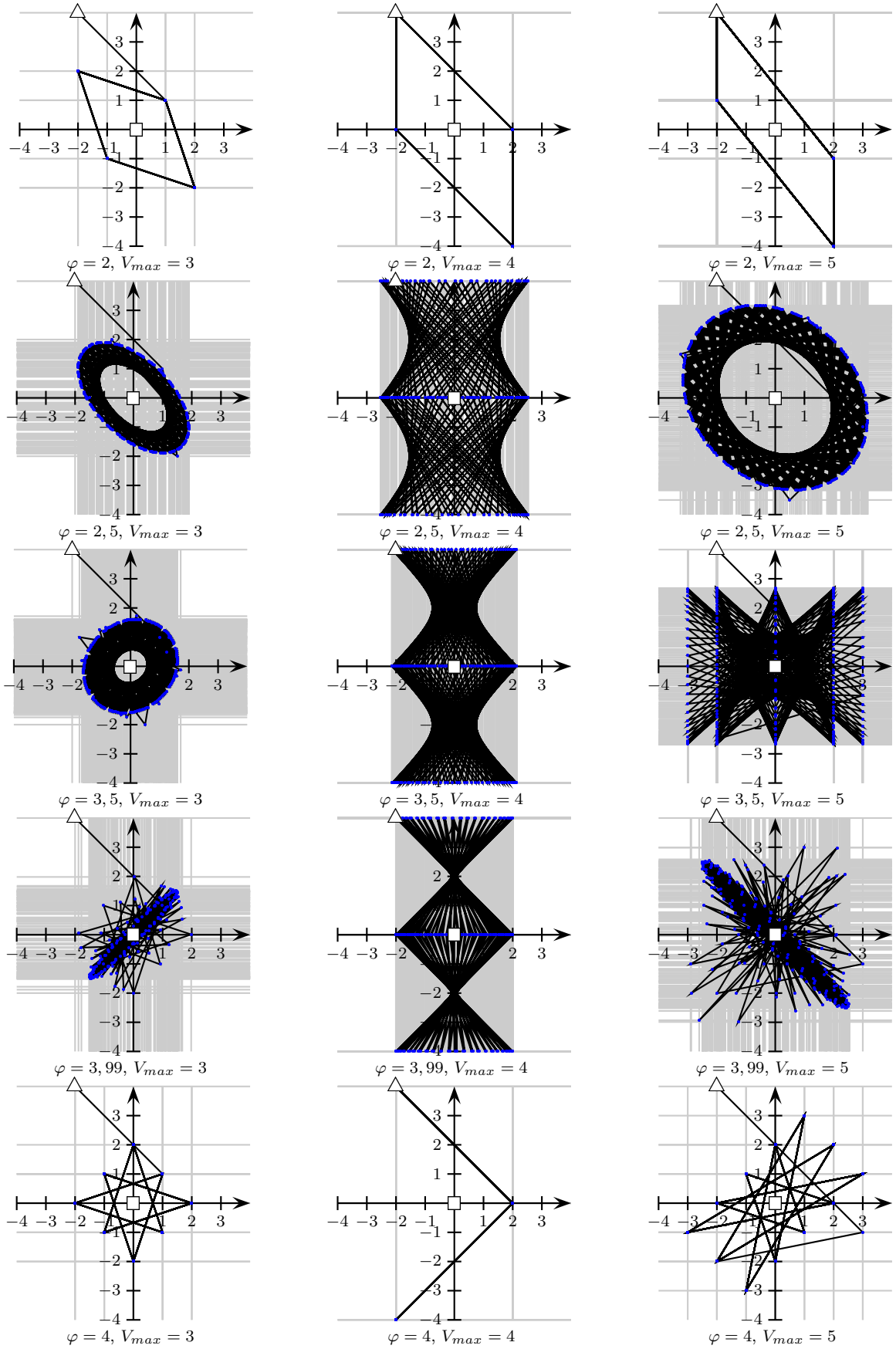


Abbildung A.15: Diagramme der nichtrandomisierten Schwarmsimulation im zweidimensionalen Suchraum für fünfzehn verschiedene Wertekombinationen von $\varphi = \varphi_{max}$ und V_{max} (Startpunkt: $(-2; 4)$, Gespeichertes Optimum: $(0; 0)$, Anfangsgeschwindigkeit: $(0; 0)$, Zeitschritte: 200)

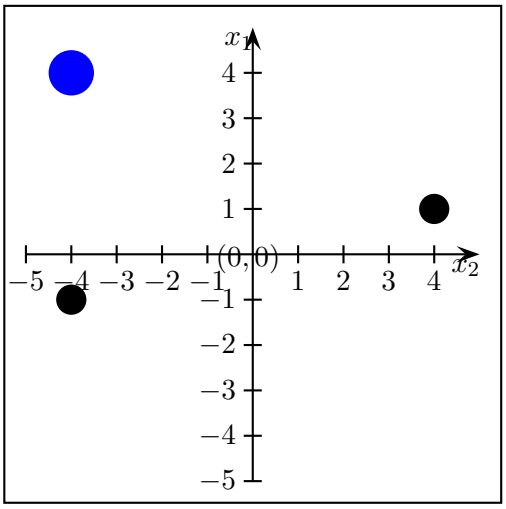


Abbildung A.16: Ausgangsbedingung für den Partikel im zweidimensionalen Suchraum mit zwei Speicherstellen

Werden die Beschleunigungen deterministisch festgelegt, indem die für die Zufallsvariablen die Maximalwerte eingesetzt werden, fällt auf, dass sich die Trajektorien nicht von den Trajektorien bei nur einer Speicherstelle unterscheiden (s. Abb. A.18).

Dieses Phänomen ist leicht zu erklären: Die unabhängige Bestimmung von Zufallszahlen in Bezug auf die beiden Speicherstellen für Partikel i zum Zeitpunkt t ließe sich ebenfalls als Bechleunigung in Bezug auf einen Punkt p_{z_i} darstellen, so dass

$$\vec{p}_{z_i}(t) = \frac{\varphi_b \vec{p}_{b_i}(t) + \varphi_n \vec{p}_{n_i}(t)}{\varphi_b + \varphi_n}. \tag{A.1}$$

Im Fall der Randomisierung kann der Punkt also an beliebiger Stelle zwischen den beiden Speicherstellen liegen. Werden aber $\varphi_b = \varphi_{b_{max}}$ und $\varphi_n = \varphi_{n_{max}}$ gesetzt, bleibt der Punkt an gleicher Stelle . In der Simulation oben wurden $\varphi_b = 2$ und $\varphi_n = 2$ gesetzt, dementsprechend gilt:

$$\vec{p}_{z_i}(t) = \frac{2\vec{p}_{b_i}(t) + 2\vec{p}_{n_i}(t)}{2+2} = \frac{1}{2}\vec{p}_{b_i}(t) + \frac{1}{2}\vec{p}_{n_i}(t) \tag{A.2}$$

.

Da sich die beiden Speicherstellen in der Simulation nicht verändern gilt weiterhin

$$\vec{p}_{z_i}(t) = \frac{1}{2}\begin{pmatrix} -4 \\ -1 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 4 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{A.3}$$

.

Dementsprechend ist die Darstellung in Abbildung A.18 identisch mit der Darstellung in Abbildung A.15.

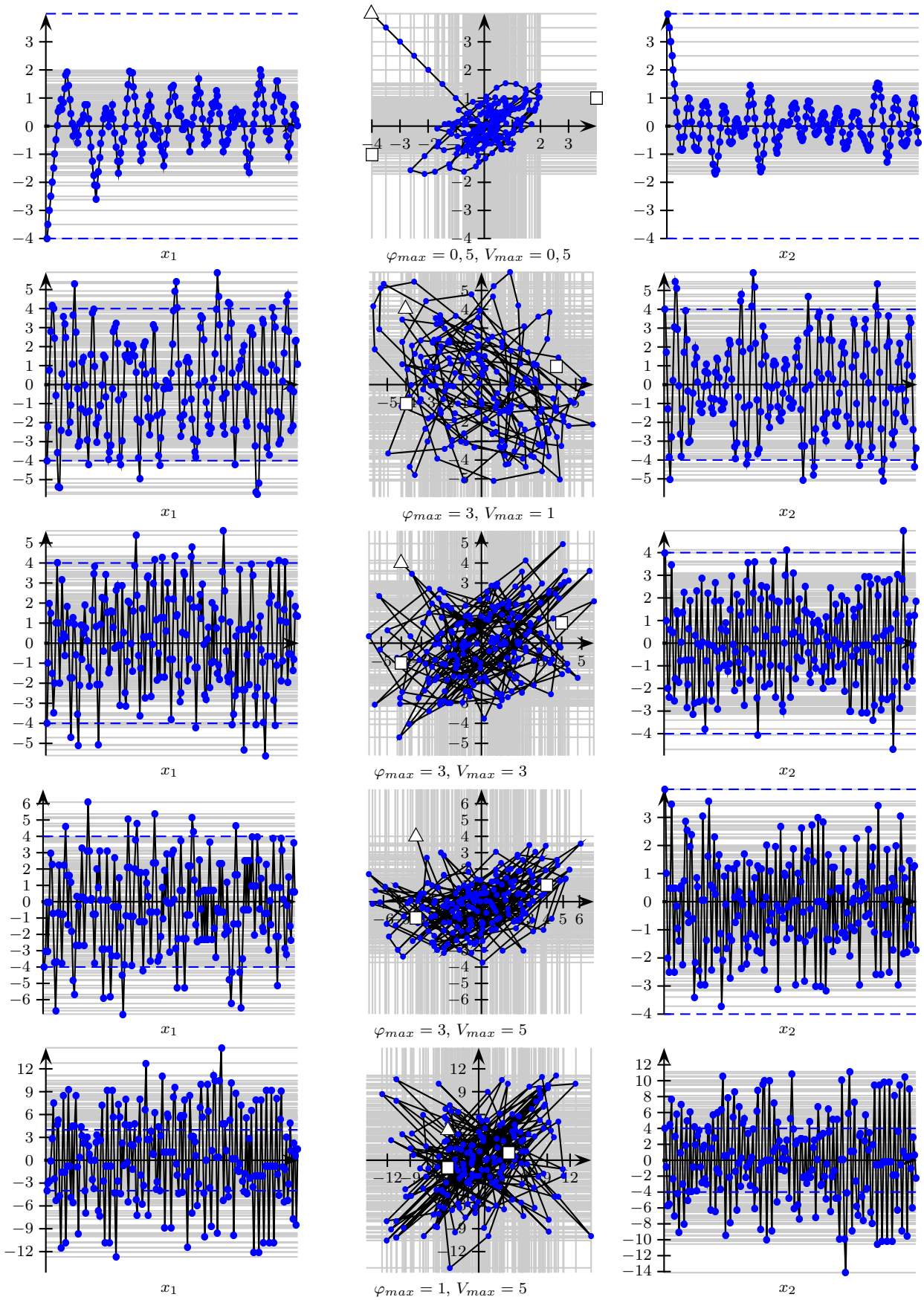


Abbildung A.17: Zweidimensionale und eindimensionale Diagramme der Schwarmsimulation im zweidimensionalen Suchraum für fünf verschiedene Kombinationen von φ_{max} und V_{max} (t=200, erste Speicherposition: (-4,-1), zweite Speicherposition (4,1), Startposition: (-4,4), Anfangsgeschwindigkeit: (0,0))

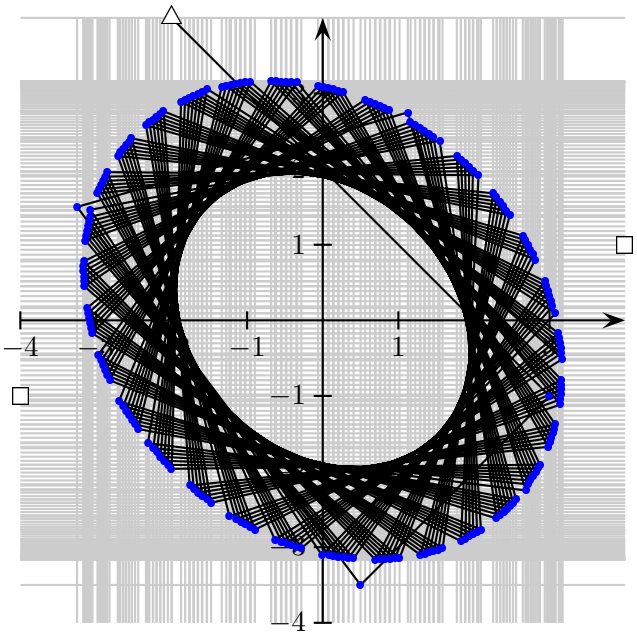


Abbildung A.18: Ergebnis der nichtrandomisierten Schwarmsimulation im zweidimensionalen Suchraum mit zwei Speicherstellen ($\varphi_{max} = 2,5$; $V_{max} = 5$; Startpunkt: $(-2;4)$; Gespeicherte Speicherstelle 1: $(-4;-1)$; Gespeicherte Speicherstelle 2: $(4;1)$ Anfangsgeschwindigkeit: $(0;0)$, Zeitschritte: 200)

A.4 Mögliche Erweiterungen des Partikelschwarmalgorithmus

A.4.1 Constriction

Als Alternative zur Verwendung von Maximalgeschwindigkeiten wurde die Verwendung von Gewichtungsfaktoren empfohlen. Kennedy und Eberhart (2001) diskutieren zwei Typen von Koeffizienten, die im Rahmen eines Modells von Clerc entwickelt worden sind. Die einfachste Art der Gewichtung wird durch den *Simplest Constriction Coefficient gebildet.*

Die Bewegung wird in diesem Modell determiniert durch die Gleichungen:

$$\vec{v}_i(t) = \chi(\vec{v}_i(t-1) + \varphi_b(\vec{p}_{b_i}(t-1) - \vec{x}_i(t-1)) + \varphi_n(\vec{p}_{n_i}(t-1) - \vec{x}_i(t-1))) \tag{A.4}$$

und

$$\vec{x}_i(t) = \vec{x}_i(t-1) + \vec{v}_i(t) \tag{A.5}$$

mit $\varphi_{max} > 4,0$.

Der einzige Unterschied zum ursprünglichen Modell ist die Einführung eines Koeffizienten χ, mit dem der neue Geschwindigkeitswert des alten Modells multipliziert wird. Dies hat eine Konvergenz des Partikels zur Folge, wenn die beiden gespeicherten Optima nah beieinander liegen.

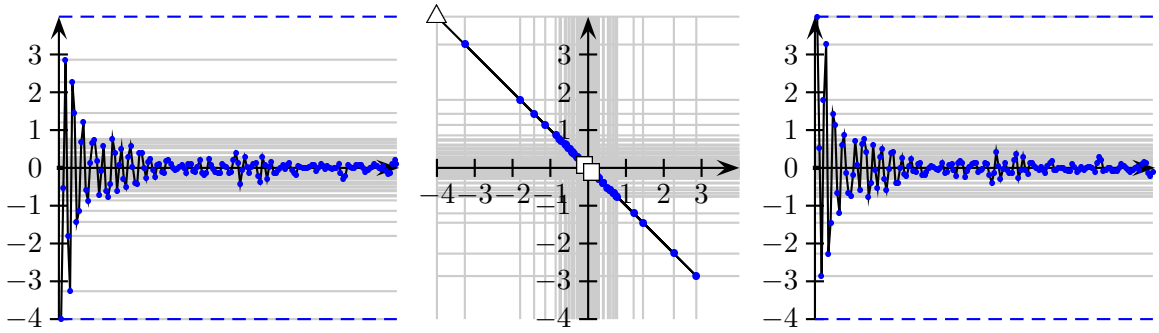


Abbildung A.19: Ergebnis der Schwarmsimulation im zweidimensionalen Suchraum mit Simplest Constriction ($\varphi_{max} = 4,1$; V_{max} nicht verwendet; $\kappa = 1,0$; Startpunkt: $(-4;4)$; Gespeichertes Optimum 1: $(-0,1;-0,1)$; Gespeichertes Optimum 2: $(0,1;0,1)$ für 150 Zeitschritte; Anfangsgeschwindigkeit: $(0;0)$)

Zur Standardisierung wird die Festlegung von χ indirekt durch Festlegung einen Parameter κ mit $\kappa \in [0;1]$ geleistet, mit

$$\chi = \frac{2\kappa}{\left|2 - \varphi_{max} - \sqrt{\varphi_{max}^2 - 4\varphi_{max}}\right|} \tag{A.6}$$

.

Eine Simulation für zwei nahe beieinander liegende Optima und $\kappa = 1$ mit $\varphi_{max} = 4,1$ zeigt Abbildung A.19. Die Auswirkungen der Setzung von $\kappa = 0,2$ zeigt Abbildung A.20. Es ist deutlich zu sehen, dass der Partikel früher konvergiert und dementsprechend weniger exploriert. Eine typische Wertekombination ist etwa $\varphi_{max} = 4,1$ und $\kappa = 1$ wie in Abbildung A.19. Vorteilhaft ist zudem die Eigenschaft des Partikels, bei einer Veränderung einer Speicherposition zu größeren Schrittweiten zurückzukehren um dann erneut zu konvergieren.

Abbildung A.21 demonstriert die Bewegung des Partikels, wenn sich nach 100 Zeitschritten eine Speicherstelle in einer Dimension verändert. In diesem Fall kommt es in den folgenden Zeitschritten zu keiner Konvergenz, da sich die Distanz zwischen den gespeicherten Stellen vergrößert hat. Es muss darauf hingewiesen werden, dass dieser Umstand auf die Künstlichkeit der Simulation zurückzuführen ist, in der sich die Speicherstellen nicht dynamisch verändern.

A.4.2 Inertia

Eine Alternative zum *Simplest Constriction Coefficient* stellt das *Inertia Weight* da. Dieses Gewicht wirkt nicht auf den gesamten Term der Geschwindigkeitsberechnung ein, sondern ausschließlich auf die

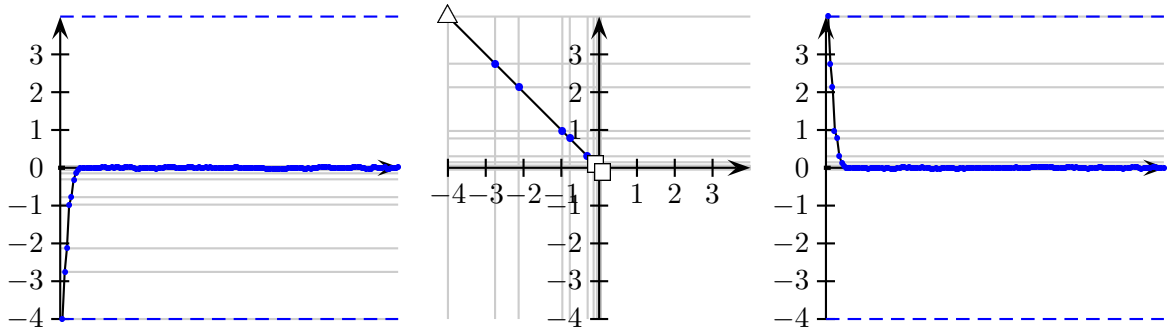


Abbildung A.20: Ergebnis der Schwarmsimulation im zweidimensionalen Suchraum mit Simplest Constriction ($\varphi_{max} = 4,1$; V_{max} nicht verwendet; $\kappa = 0,2$; Startpunkt: $(-4;4)$; Gespeichertes Optimum 1: $(-0,1;-0,1)$; Gespeichertes Optimum 2: $(0,1;0,1)$ für 150 Zeitschritte; Anfangsgeschwindigkeit: $(0;0)$)

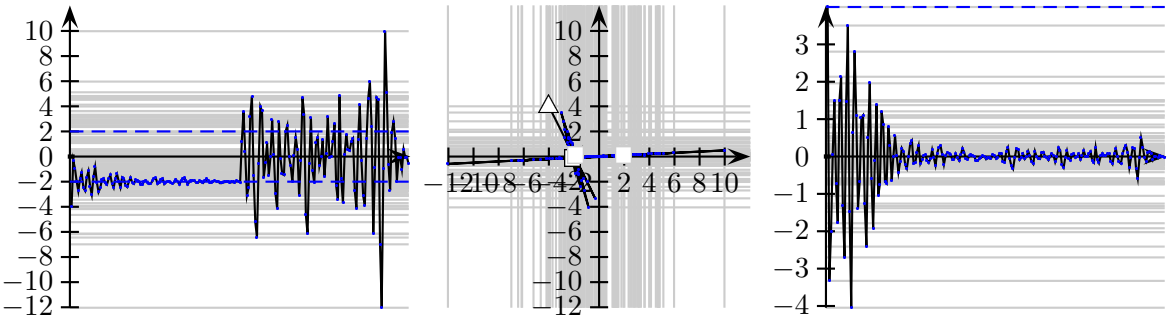


Abbildung A.21: Ergebnis der Schwarmsimulation im zweidimensionalen Suchraum mit Simplest Constriction ($\varphi_{max} = 4,1$; V_{max} nicht verwendet; $\kappa = 1$; Startpunkt: $(-4;4)$; Gespeichertes Optimum 1: $(-0,1;-0,1)$; Gespeichertes Optimum 2: $(0,1;0,1)$ für 100 Zeitschritte; $(2;0,1)$ für die folgenden 100 Zeitschritte; Anfangsgeschwindigkeit: $(0;0)$)

alte Geschwindigkeit. Insofern sind die Partikel also träge, als dass sie ohne neue Beschleunigung langsam über mehrere Zeitschritte sich der Nullgeschwindigkeit annähern würden. Wird dieser Koeffizient mit α bezeichnet, so verändern sich die Gleichungen dementsprechend zu

$$\vec{v}_i(t) = \alpha(\vec{v}_i(t-1)) + \varphi_b(\vec{p}_{b_i}(t-1) - \vec{x}_i(t-1)) + \varphi_n(\vec{p}_{n_i}(t-1) - \vec{x}_i(t-1)) \tag{A.7}$$

und

$$\vec{x}_i(t) = \vec{x}_i(t-1) + \vec{v}_i(t) \tag{A.8}$$

.

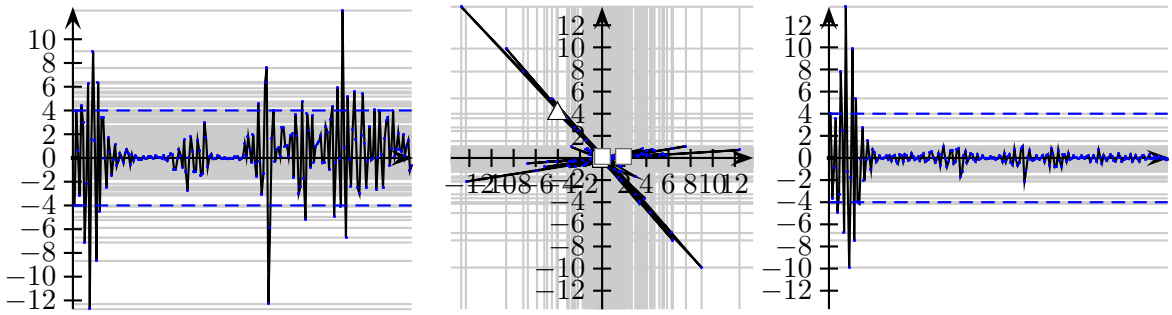


Abbildung A.22: Ergebnis der Schwarmsimulation im zweidimensionalen Suchraum mit Inertia Weight ($\varphi_{max} = 4,1$; $V_{max} = 4$; $\alpha = 0,7$; Startpunkt: $(-4;4)$; Gespeichertes Optimum 1: $(-0,1;-0,1)$; Gespeichertes Optimum 2: $(0,1;0,1)$ für 100 Zeitschritte; $(2;0,1)$ für die folgenden 100 Zeitschritte; Anfangsgeschwindigkeit: $(0;0)$)

Konvergenz wird erreicht wenn $0 < \varphi \leq 2\alpha + 2 - \epsilon$ und $0 < \alpha < 1$. In diesem Fall ist die Verwendung von V_{max} zur Verhinderung von Geschwindigkeitseskalationen nicht notwendig, aber auch nicht ausgeschlossen. Es ist zudem möglich, den Wert des Koeffizienten über die Zeitschritte hinweg kleiner werden zu lassen. Kennedy und Eberhart (2001) nennen bspw. das Intervall von $0,9$ bis $0,4$ bei $\varphi_{max} = 4$ als sinnvolle Setzungen unter Zuhilfenahme eines an die Geschwindigkeitsentwicklung angepassten V_{max}.

Chuanwen und Bompard (2005) schlagen eine Bestimmung des Gewichtungsfaktors für den Zeitpunkt t in der Form

$$\alpha(t) = \alpha(0)e^{-\rho\frac{t}{T_{max}}} \tag{A.9}$$

vor, wobei ρ einen Kontrollfaktor, $\alpha(0)$ den Anfangswert von α und T_{max} die maximale Zahl an Zeitschritten darstellt. Für $\rho = 1$, $\alpha(0) = 1$ und $T_{max} = 100$ sähe die Abfolge der Werte aus wie in Abbildung A.23 dargestellt.

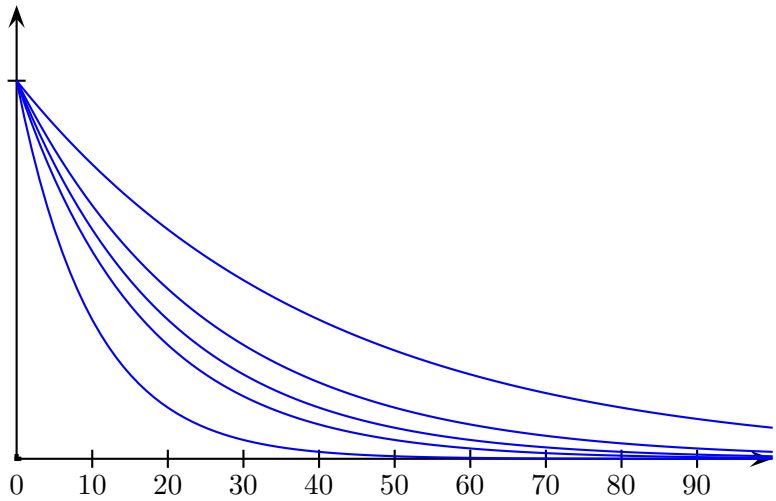


Abbildung A.23: Funktionen für die Bestimmung des Inertia-Faktors nach Chuanwen und Bompard (2005): Funktionen für die Parameterwerte $\rho = 1$ bei $\alpha(0) = 1$ und $T_{max} = 100$

Chatterjee und Siarry (2004) schlagen die Gleichung

$$\alpha(t) = \frac{(T_{max} - t)^n}{T_{max}^n}(\alpha(0) - \alpha(T_{max})) + \alpha(T_{max}) \tag{A.10}$$

vor. In dieser Gleichung wird n als nonlinearer Modulationsindex bezeichnet, dessen Werte im Intervall $[0,9; 1,3[$ als optimal eingeschätzt werden. Vorgeschlagen wird $n = 1,2$. Die Autoren betrachten dabei die Setzung von $\alpha(0) = 0,2$ und $\alpha(T_{max}) = -0,3$. Zur Veranschaulichung wird der Verlauf der Werte von $\alpha(t)$ zunächst einmal für $\alpha(0) = 1$, $\alpha(T_{max}) = 0,1$ und verschiedene Werte von n in Abbildung A.24 gezeigt.

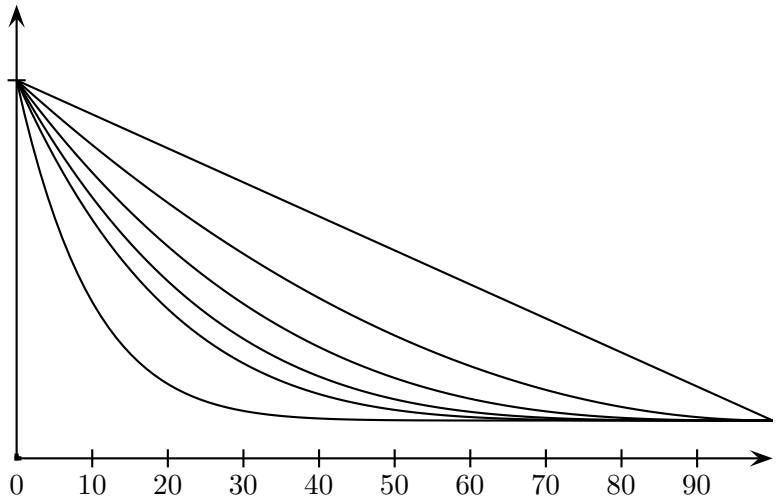


Abbildung A.24: Funktionen für die Bestimmung des Inertia-Faktors nach Chatterjee und Siarry (2004): Funktionen für die Parameterwerte $n = 1, n = 2, n = 3, n = 4, n = 5$ und $n = 10$ bei $\alpha(0) = 1$ und $\alpha(T_{max}) = 0,1$

Der nach Ansicht der Autoren ideale Verlauf ist inklusive der Grenzen in Abbildung A.25 zu sehen.

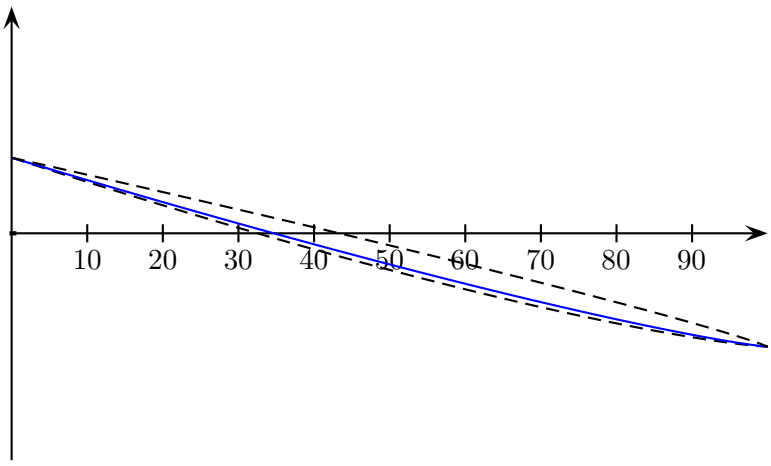


Abbildung A.25: Funktionen für die Bestimmung des Inertia-Faktors nach Chatterjee und Siarry (2004): Funktionen für die Parameterwerte $n = 0,9$ sowie $n = 1,2$ und $n = 1,3$ bei $\alpha(0) = 0,2$ und $\alpha(T_{max}) = -0,3$

Eine weitere Alternative liefern Liu, Wang, Jin, Tang und Huang (2005) mit dem *Adaptive Inertia Weight Factor* (AIWF). Dieser Faktor berechnet sich überführt in die hier verwendete Notation als

$$\alpha_i(t) = \begin{cases} \alpha_{min} + (\alpha_{max} - \alpha_{min})\frac{G_{max}(t) - G(\vec{p}_i(t))}{G_{max}(t) - \bar{G}(t)} & G(\vec{p}_i(t)) \geq \bar{G}(t) \\ \alpha_{max} & G(\vec{p}_i(t)) < \bar{G}(t) \end{cases} \tag{A.11}$$

Damit wird das Gewicht für einen Partikel angepasst an die Qualität der seiner Position entsprechenden Lösung im Vergleich zum Durchschnitt aller aktuellen Lösungen. Dabei ist der Faktor für unterdurchschnittliche Lösungen maximal und nimmt für überdurchschnittliche Lösungen ab, wobei für die beste gefundene Lösung der Minimalwert vergeben wird.

A.4.3 Zusätzliche Entwicklungsideen für Schwarmalgorithmen

Inspiriert durch Beobachtungen von Fischschwärmen schlagen He, Wu, Wen, Saunders und Paton (2004) ein Modell mit passives Kongregation vor. In diesem Modell werden Partikel zusätzlich auf in Richtung zufällig ausgewählter weiterer Partikel des Schwarms beschleunigt:

> Schooling fish are generally considered a „selfish herd“ [...], in that each individual attempts to take the sweeping generalization advantage from group living, independent of the fates of neighbors [...]. In these congregations, information may be transferred passively rather than actively [...].(He et al., 2004, S. 137)

Mostaghim und Teich (2003) stellen eine Erweiterung des Algorithmus für multiobjektive Optimierungsprobleme vor, in denen Probleme mit mehr als einem Zielkriterium angegangen werden können.

Hybride Modelle, in denen Algorithmen Eigenschaften von Schwarmalgorithmen mit denen anderer Algorithmenklassen verbinden, sind ebenfalls in vielfältiger Form entwickelt worden. Lovbjerg, Rasmussen und Krink (2001) sowie Shi, Liang, Lee, Lu und Wang (2005) stellen ein hybrides Modell vor, in denen ein Partikelschwarm mit einer Populationsentwicklung im Sinne Genetischer Algorithmen verknüpft wird. Conradie, Miikkulainen und Aldrich (2002) entwickeln ein Modell mit einer Verbindung zwischen Partikelschwärmen und Neuronalen Netzwerken und einer Anwendung auf die Kontrolle von Reaktoren. (Kennedy & Eberhart, 1995) diskutieren weiterhin die Möglichkeit, Interaktionen zwischen Partikeln über Entfernungen zu modellieren, um eines bessere Exploration des Suchraums zu ermöglichen.

Eine Vielzahl von Erweiterungen ist denkbar, diese kurze Darstellung soll an dieser Stelle aber genügen, um einen Eindruck zu geben von möglichen Ansatzpunkten und Formen der Modifikationen.

[...] we with our lives are like islands in the sea or like trees in the forest. The maple and the pine may whisper to each other with their leaves and Conanicut and Newport hear each other's fog horns, but the trees also co-mingle their roots in the darkness underground, and the islands also hang together through the ocean's bottom [...].

William James, Final Impressions of a Psychical Researcher (1909)

B

Ergänzungen zu genetischen Algorithmen: TSP und Inselalgorithmen

In diesem Anhang sollen zwei Erweitungen von genetischen Algorithmen vorgestellt werden: Eine Anwendung von modifizierten genetischen Algorithmen auf eine andere Problemklasse und eine Klasse paralleler genetischer Algorithmen.

B.1 Genetische Algorithmen für das TSP-Problem

In diesem Abschnitt soll eine Anwendung genetischer Algorithmen auf das Problem des fahrenden Handlungsreisenden vorgestellt werden. Der Weg eines Händlers, der eine Menge von Städten jeweils einmal besuchen will, um dann zum Ausgangspunkt zurückzukehren, soll minimiert werden. Dazu sind andere Repräsentationen, Fitnessfunktionen und Operatoren notwendig.

B.1.1 Kodierung von Permutationen und Sequenzen

Es soll an dieser Stelle eine Einsatzmöglichkeit für genetische Algorithmen diskutiert werden, die von der Vorgehensweise abweicht Kodierungen als Kodierungen von Parameterwerten zu benutzen. Dabei hat sich die Klasse der Permutationsprobleme als Herausforderung erwiesen (Whitley, 2000a): Der Suchraum bei Permutationsproblemen wächst mit der Objektanzahl, bei n Objekten gibt es $n!$ Ordnungen dieser Objekte mit $0! = 1$ und $n! = n \cdot (n-1)!$. Permutationen sind relevant für Probleme der Reihenfolgebildung, etwa bei der Planung von Reiserouten. Sie können Zyklen oder Ketten abbilden, womit sich das Problem ergibt, dass bei einem Zyklus verschiedene Reihenfolgen von Objekten aufgrund der Geschlossenheit des Zyklus äquivalent werden: $(1, 2, 3, 4)$ und $(4, 3, 2, 1)$ sind äquivalent, wenn es um die beschriebene Aufeinanderfolge in einem Zyklus geht.

Eine Möglichkeit der Repräsentation von Permutationen der Länge k ist eine geordnete Menge von k natürlichen Zahlen, die jede Zahl zwischen 1 und k enthält. Kürzer lässt sich die Menge der möglichen Permutationen auch auf natürliche Zahlen im Intervall $[0; k! - 1]$ abbilden. Die Vorgehensweise für die Enkodierung könnte man folgendermaßen skizzieren:

1. Bring die Objekte in eine festgelegte Ordnung und vergib feste Rangplätze für jedes Objekt.

2. Setze den Wert s auf 0.

3. Führe den folgenden Schritt für jede Position von 1 bis k aus.

4. Für eine Position j in der Permutation sieh nach, wie viele Objekte, die Positionen nach der betrachteten Position einnehmen einen niedrigeren Rangplatz haben als das Objekt an der betrachteten Position und multipliziere diese Zahl mit $(k - j)!$ und addiere das Ergebnis zu s.

5. Nimm das Ergebnis s als Kodierung für die gegebene Permutation.

Als Beispiel würde die Sequenz mit den Rangnummern $(5-4-1-6-3-2)$ zu dem Ergebnis

$$\begin{aligned} s =& 4\cdot(6-1)! + 3\cdot(6-2)! + 0\cdot(6-3)! + 2\cdot(6-4)! + 1\cdot(6-5)! + 0\cdot(6-6)! \\ =& 4\cdot 120 + 3\cdot 24 + 0\cdot 6 + 2\cdot 2 + 1\cdot 1 + 0\cdot 1 = 480 + 72 + 0 + 4 + 1 + 0 \\ =& 557 \end{aligned} \tag{B.1}$$

führen, der Sequenz $(1-2-3-4-5-6)$ würde $s=0$ zugeordnet und der Sequenz $(6-5-4-3-2-1)$ der Wert $s = 719 = k! - 1$. Die Eindeutigkeit der Lösung folgt aus

$$t! > t-1\cdot(t-1)! + t-2\cdot(t-2)! + \ldots + 0\cdot 0! \tag{B.2}$$

, denn

$$\begin{aligned} t! &= (t-1)\cdot(t-1)! + \cdot(t-1)! = (t-1)\cdot(t-1)! + (t-2)\cdot(t-2)! + (t-2)! \\ &= \ldots = t-1\cdot(t-1)! + t-2\cdot(t-2)! + \ldots + 0\cdot 0! + 0! \end{aligned} \tag{B.3}$$

Damit ist jeder Summand aus der Summe rekonstruierbar, und somit kann aus der Summe die ursprüngliche Ordnung rekonstruiert werden, da die Anzahl der folgenden Positionen mit höherem Rangplatz jede Position sukzessive determiniert.

Die Anzahl möglicher Zyklen ergibt sich bei Irrelevanz der Richtung und des Startpunkts eines Zyklus bei k Objekten als $\frac{k!}{2k} = \frac{1}{2}(k-1)!$. Bei einem Zyklus reduziert sich damit die relevante Information auf die Frage, mit welchen zwei anderen Objekten ein Objekt verbunden ist, da Antworten auf die Frage, welche relative Position das Tripel zum Startpunkt sowie die Frage, welches der zwei angrenzenden Objekte vor dem betreffenden Objekt positioniert ist keine Differenzierung zwischen Zyklen ermöglichen. Nach Whitley (2000a) bietet sich deshalb eine Matrixrepräsentation an mit k Zeilen und k Spalten. Stellt $\mathcal{A}$ diese Matrix mit den Elementen a_{ij} dar, dann gilt für zwei benachbarte Objekte i und j, dass $a_{ij} = 1$ und für alle anderen Paare dass $a_{ij} = 0$. Überführt man den Zyklus in eine zyklische Graphdarstellung, so würde jeder Kante zwischen zwei Objekten i und j eine 1 in der Matrix an der Stelle a_{ij} entsprechen.

Der Zyklus $(5-4-1-6-3-2)$, der sich alternativ auch als $(6-3-2-5-4-1)$ oder auch als $(4-5-2-3-6-1)$ schreiben lässt, würde damit in jedem Fall repräsentiert durch die Matrix

$$\mathcal{A} = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \end{pmatrix} \tag{B.4}$$

.

Diese Darstellungsform ermöglicht zudem eine leichte Vergleichbarkeit von verschiedenen Lösungen, da gleiche Komponenten zweier Matrizen direkt auf strukturelle Ähnlichkeiten verweisen.

Prinzipiell ist jede symbolisch kodierbare Form der Repräsentation geeignet, um in einen genetischen Code überführt zu werden, sofern eine Abstimmung erfolgt zwischen Bewertungsfunktion, Variationsoperatoren und Selektionsmethode (Fogel & Angeline, 2000). Die Bewertungsfunktion muss auf der gewählten Repräsentation aufsetzen, der Effekt von Variationsoperatoren in Bezug auf die Lösungsvariation hängt von der Beschaffenheit der Kodierung ab. Als Richtlinie lässt sich vielleicht nur festhalten:

> The most natural choice for representation is a subjective choice, and it will differ across investigators, although, like a suitable scientific model, a suitable representation should be as complex as necessary (and no more so) and should 'explain' the phenomena investigated, which here means that the resulting search should be visualizable or imaginable to some extent. (Fogel & Angeline, 2000, S. 162)

B.1.2 Fitnessfunktionen für das TSP

Für das Travelling-Salesman-Problem ergibt sich die Konstruktion der Fitnessfunktion aus der Aufgabenstellung. Die Summe der Längen aller zurückgelegten Wege soll minimiert werden, dementsprechend lässt sich die Bewertungsfunktion als negierte Summe der Längen einer spezifischen Lösung bestimmen. Als Eingangsdaten werden räumliche Koordinaten von Objekten, oder anschaulicher von Städten benötigt, die eine beliebige Dimensionalität haben können. Die Koordinaten müssen eine gültige Grundlage für die Etablierung einer Distanzmetrik darstellen. Eine gültige Lösung besteht aus einer Permutation der Städte $P = (p_1, \ldots, p_k)$ in der die natürlichen Zahlen von 1 bis k genau einmal enthalten sind. Die Bewertungsfunktion ergibt sich als:

$$o(P_c) = -\sum_{i=1}^{k-1} D_{p_{c_i}, p_{c_{i+1}}} - D_{p_{c_{i+1}}, p_{c_1}} \tag{B.5}$$

Dabei ist $D_{i,j}$ ein Distanzmaß für den Abstand zwischen zwei Objekten i und j, das unterschiedlich festgelegt werden kann, je nach Problemstellung. Geht es um den „direkten“ Weg in zwei- oder dreidimensionalen Problemen bietet sich als natürliches Maß die euklidische Distanz an. Ist x_{qr} die q-te Koordinate von Objekt r ergibt sich die euklidische Distanz bei m Dimensionen als:

$$D_{ij} = \sqrt{\sum_{f=1}^{m} (x_{fi} - x_{fj})^2}. \tag{B.6}$$

Wird die Distanz hingegen als aus den einzelnen Variablenabständen zusammengesetzt konzipiert, wäre die City-Block-Metrik angemessener, also

$$D_{ij} = \sum_{f=1}^{m} |x_{fi} - x_{fj}|. \tag{B.7}$$

Beide Metriken gehören zur Klasse der Minkowski-Metriken mit der durch k parametrisierten Form

$$D_{k_{ij}} = \sqrt[k]{\sum_{f=1}^{m} |x_{fi} - x_{fj}|^k}. \tag{B.8}$$

Grabmeier und Rudolph (2002) sprechen von der sogenannten L_r-Norm und bezeichnen den Fall $r = 1$ auch als „Manhattan“-Distanz. Mit jedem dieser Distanzmaße könnte dementsprechend eine mögliche Bewertungsfunktion konstruiert werden, die Angemessenheit des Distanzmaßes muss durch das Zielproblem vorgegeben werden.

Neben den beiden beschriebenen Problemklassen und reinen Funktionsoptimierungen gibt es einen große Anzahl anderer Aufgabenstrukturen für Genetische Algorithmen. Das Konstruktionsprinzip für eine spezifische objektive Funktion besteht aber auch hier in einer Zuordnung von Zahlenwerten aufgrund der eingeschätzten Qualität von Lösungen. Dabei wird im allgemeinen eine Messung dieser Qualität auf einem möglichst hohen Skalenniveau angestrebt, wenn nicht ein bestimmtes Skalenniveau durch die Auswahl der Operatoren im Restalgorithmus entweder gefordert oder ausreichend ist.

Dabei kann es sich um Zeitmessungen, Streckenmessungen, Messungen von Materialverbrauch, Messung der Kosten eines Vorgehens, Messungen der Anzahl von Unfällen oder Störungen, Messungen von Erträgen, Stromflüssen, Wassermengen handeln, ohne dass der Algorithmus modifiziert werden müsste. Der Genetische Algorithmus ist blind für das empirische Relativ das mit dem numerischen Relativ durch die Fitnessfunktion verbunden ist.

B.1.3 Mutation für das TSP-Problem

Andere Formen der Mutation existieren für verschiedene Repräsentationen und Kodierungen. Sind die Werte bspw. nicht binär kodiert, sondern mit reellen Zahlen, so würde der Ersetzungsprozess einer stochastischen Abbildung folgen müssen. Etwas kreativer muss die Aufgabe gelöst werden etwa für die Kodierung von Reihenfolgen wie für das TSP. Eine lokale Veränderung ist unmöglich, da Elemente nicht

einfach dupliziert werden können, und eine sinnvolle Lösung alle Elemente enthalten muss. Eine Variante besteht hier in folgendem Algorithmus:

```
MUTATIONSOPERATOR_REIHENFOLGE(population)
1  for a ← 1 to population.size
2  do pos ← POSITIONSANZAHL(population.individuum[a].genom)
3      for i ← 1 to pos
4      do r ← ZUFALLSZAHL(0, 1)
5          if r < mutationsrate
6             then s ← i
7                  while s = i
8                  do s ← ZUFALLSZAHL_GANZZAHLIG(1, pos)
9                      AUSTAUSCHEN(population.individuum[a].genom, i, s)
```

Listing B.1: Pseudocode für den Mutationsoperator bei kodierten Reihenfolgen

Eine Mutation wird durchgeführt, indem das betroffene Element mit einem zufällig ausgewählten anderen Element des Genoms ausgetauscht wird. Für das TSP-Problem bedeutet ein Austausch von zwei Positionen eine Veränderung der Reihenfolge, in der Städte besucht werden. In Abbildung B.1 ist die Konsequenz einer derartigen Veränderung visualisiert worden. Durch den Austausch zweier Elemente

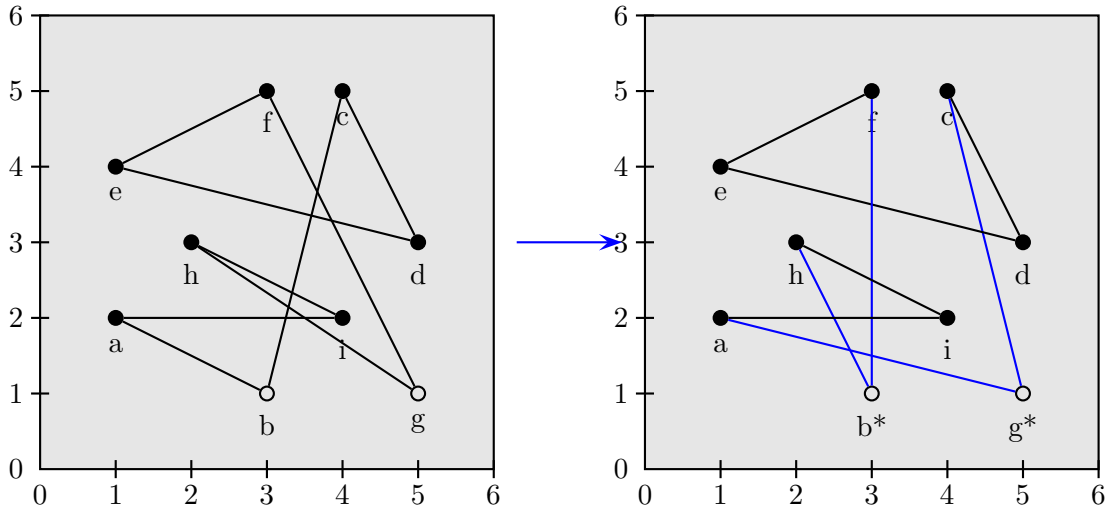


Abbildung B.1: Mutation für eine TSP-Lösung: Durch das Austauschen der Elemente b und g haben sich vier Verbindungen verändert

werden maximal vier (wie im Beispiel) und minimal zwei Verbindungen verändert.

Natürlich wären auch radikalere Mutationen denkbar, bei denen Gruppen von Elementen mit anderen Gruppen ausgetauscht würden. Ein solcher Operator könnte bspw. im Fall einer Mutation zunächst zufällig eine Zahl von Elementen n_{el} kleiner der halbierten Gesamtzahl bestimmen, dann einen Ersetzungsort, an den die Gruppe aus dem Startelement und $n_{el} - 1$ folgenden Elementen versetzt würde. Das Beispiel aus Abbildung B.2 zeigt den Fall, dass n_{el} den Wert 3 angenommen hat, Startelement das Element b und Zielelement das Element g ist.

Theoretisch ließe sich diese Strategie auch wieder zurück auf den binären Fall übertragen. Dort kann sie aber zu unerwünscht disruptiven Effekten führen. Im Normalfall werden Merkmale durch mehr als eine Stringposition kodiert, so dass Punktmutationen in der Lage sind, unterschiedliche große und hinreichend vielfältige Effekte zu erzeugen, je nachdem an welcher Stelle mutiert wird. Sinnvoller wäre der Austausch funktioneller Einheiten, wobei hier mehr über das Problem und die Kodierung bekannt sein muss, um einen solchen Operator zu konstruieren. Im Fall des TSP-Problems bietet der erweiterte Mutationsoperator eine verbesserte Chance, lokalen Minima im Suchraum zu entgehen, was natürlich auch durch eine entsprechend hoch gewählte Mutationsrate im einfachen Fall erreicht werden kann.

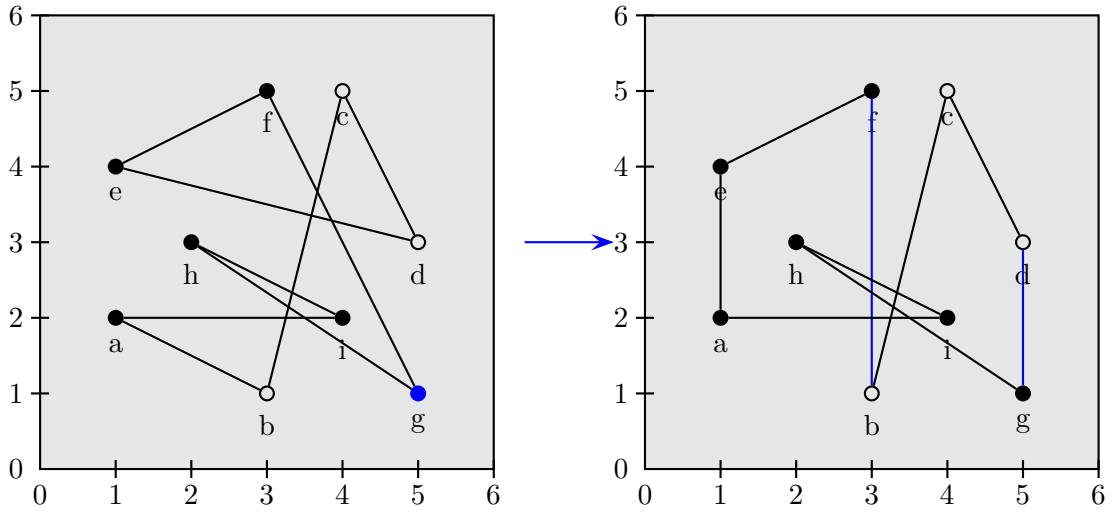


Abbildung B.2: Alternative Mutation für die TSP-Lösung: Durch das Versetzen der Gruppe der Elemente b, c und d an die Position von g haben sich zwei Verbindungen verändert

In den Anfangstagen des *Evolutionary Computing* wurde auch mit dem *Inversions*-Operator experimentiert, der einen Teil des Genoms, oder auch das gesamte Genom in dem Sinne invertiert, dass das erste Element zum letzten und das letzte Element zum ersten wird und sich die Reihenfolge der Kodierungen ändert. Damit sollte eine maximale Variation erzeugt werden zur Vergrößerung des Suchraums.

B.1.4 Rekombination für das TSP-Problem

Die Wahl des Rekombinationsoperators hängt ab von der Form der Kodierung. Solange die Information an Stellen ohne Einschränkung verteilt werden kann, d.h. solange keine Abhängigkeiten zwischen Belegungen bestehen, sind die beschriebenen Operatoren unter Umständen leicht modifiziert einsetzbar. Probleme gibt es aber in den Fällen, in denen bestimmte Eigenschaften, die das gesamte Genom betreffen eingehalten werden müssen, wie etwa bei der Kodierung von Reihenfolgen.

Würde hier ein Genom frei zusammengesetzt aus den Elterngenomen, würden Positionen unter Umständen dupliziert und andere Positionen in den Nachkommen fehlen. Eine der Lösungen, wie trotzdem ein Rekombinationsoperator zum Einsatz kommen kann, ist der so genannte *Order Cross-Over* Operator, oder auch *OX*-Operator (Michalewicz & Fogel, 2000). Dieser soll hier exemplarisch vorgestellt werden:

Gegeben seien zwei Reihenfolgen der Länge k. Der OX-Operator bestimmt zufällig zwei Trennstellen zwischen 1 und $k-1$. Der Bereich zwischen diesen Trennstellen wird nun jeweils einem der Nachkommen übergeben, so dass der Mittelteil eines Elternteils unverändert bei einem der beiden Nachkommen vorliegt. Nun werden die Informationen des jeweils anderen Elternteils auf die freien Stellen in folgender Weise verteilt: Beginnend nach der zweiten Trennstelle werden die Werte des anderen Elternteils kopiert unter Auslassung der Werte, die bereits im Mittelteil vorhanden sind. Wird das Ende eines der Strings beim Lesen oder Schreiben erreicht, wir die Operation am Anfang des Strings fortgesetzt (s. Abbildung B.3 für ein Beispiel, eine Veranschaulichung für das TSP-Problem zeigt Abbildung B.4). Der OX-Operator rekombiniert die in den Reihenfolgen enthaltene ordinale Information der Elternteile und ist damit ein gültiger Rekombinationsoperator.

Michalewicz und Fogel (2000) schlagen für Reihenfolgekodierungen weiterhin die Operatoren *Partially-Mapped Cross-Over* (PMX), *Cycle Cross-Over* (CX), *Edge Recombination* (ER), *Union*, *Matrix Cross-Over* (MX) sowie Variationen dieser Verfahren vor, eine Beschreibung des *Maximal Preservative Corss-Overs* (MPX) und des *Merge Cross-Overs* (MX1) gibt Whitley (2000b). Für die Kodierung mit reellen Zahlen bieten sich neben den Methoden, die für binärkodierte Strings vorgeschlagen worden sind auch Methoden an, in denen Werte der Nachkommen als arithmetisches Mittel der Werte der Elternteile oder als zufällig bestimmte Linearkombination im Bereich zwischen den Elternwerten bestimmt werden (Fogel, 2000c).

Mit den vorgestellten Modifikationen kann ein GA für das TSP-Problem entwickelt werden, bspw. wie in Anhang C berichtet.

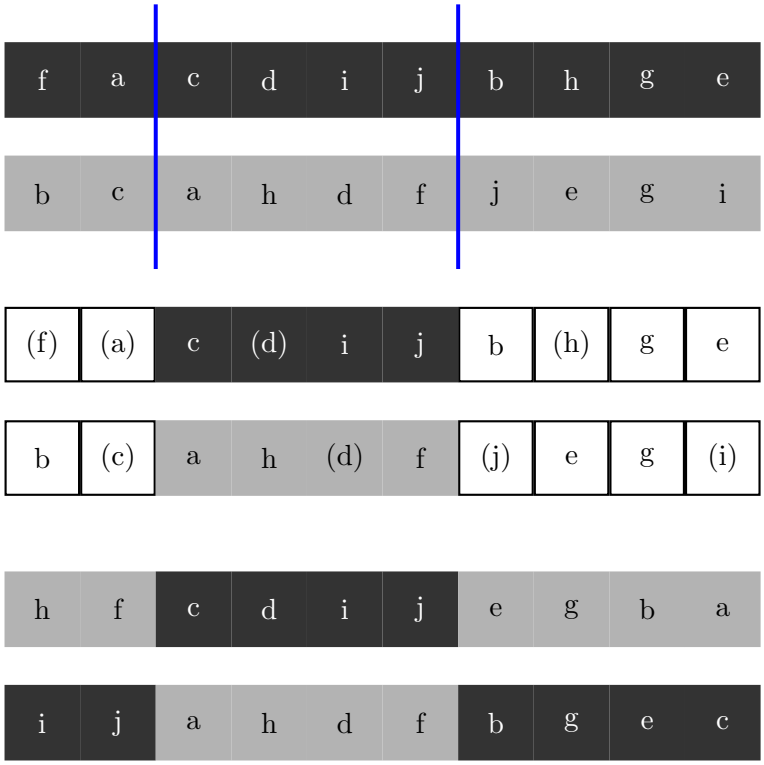


Abbildung B.3: Beispiel für die Anwendung des OX-Cross-Overs in zwei Schritten: Im ersten Schritt werden die Mittelteile übertragen, die im Genom des anderen Nachkommen vorhandenen Positionen sind eingeklammert, im zweiten Schritt werden die nicht-eingeklammerten Positionen jeweils von der zweiten Trennungsstelle ausgehend übertragen.

B.2 Inselalgorithmen

B.2.1 Inselalgorithmen zur Klassifikation

Die in diesem Abschnitt vorgestellten Algorithmen sind Meta-Algorithmen, sie sind auf die Implementierung anderer Algorithmen angewiesen, für die sie eine Einbettung liefern. Die Metapher, die zur Darstellung gewählt wurde, ist die Metapher der Inselgruppe (s. Abbildung B.5). Diese Illustration bezieht sich auf die Eigenschaften evolutionärer Prozesse, die auf Inselgruppen beobachtet werden, wie etwa auf den Galapagos-Inseln durch Charles Darwin. .

Die vorgestellten Algorithmen sind eine Variante paralleler Algorithmen, einzelne Inseln können etwa bei Nutzung einer parallelen Rechnerarchitektur oder bei Verwendung eines Computernetzwerkes durch einzelne Rechner oder Prozessoren simuliert werden. Dabei werden auf jeder Insel parallel Suchalgorithmen gestartet, die über den Austausch von Zwischenergebnissen koordiniert werden.

Rivera (2001) unterscheidet vier Arten der Implementierung Paralleler Genetischer Algorithemn:

1. *Global Parallelization*: Lediglich die Fitnessberechnung wird auf mehrere Prozessoren verteilt, dies ist kein Inselalgorithmus im eigentlichen Sinn.

2. *Coarse-Grained Parallelization*: Die Population wird in Subpopulationen oder *Demes* aufgeteilt, dies ist das vorgestellte Inselmodell.

3. *Fine-Grained Parallelization*: Jedes Individuum wird auf einen eigenen Prozessor verteilt, und die Verbindungen zwischen den Prozessoren bestimmen die Funktionsweise der Genetischen Operatoren, Cross-Over-Operatoren werden beispielsweise nur auf Individuen angewandt, die auf benachbarten Prozessoren situiert sind.

4. *Hybrid Parallelization*: Mehrere Parallelisierungsansätze werden kombiniert.

An dieser Stelle sollen die Algorithmen der zweiten Klasse näher beschrieben werden.

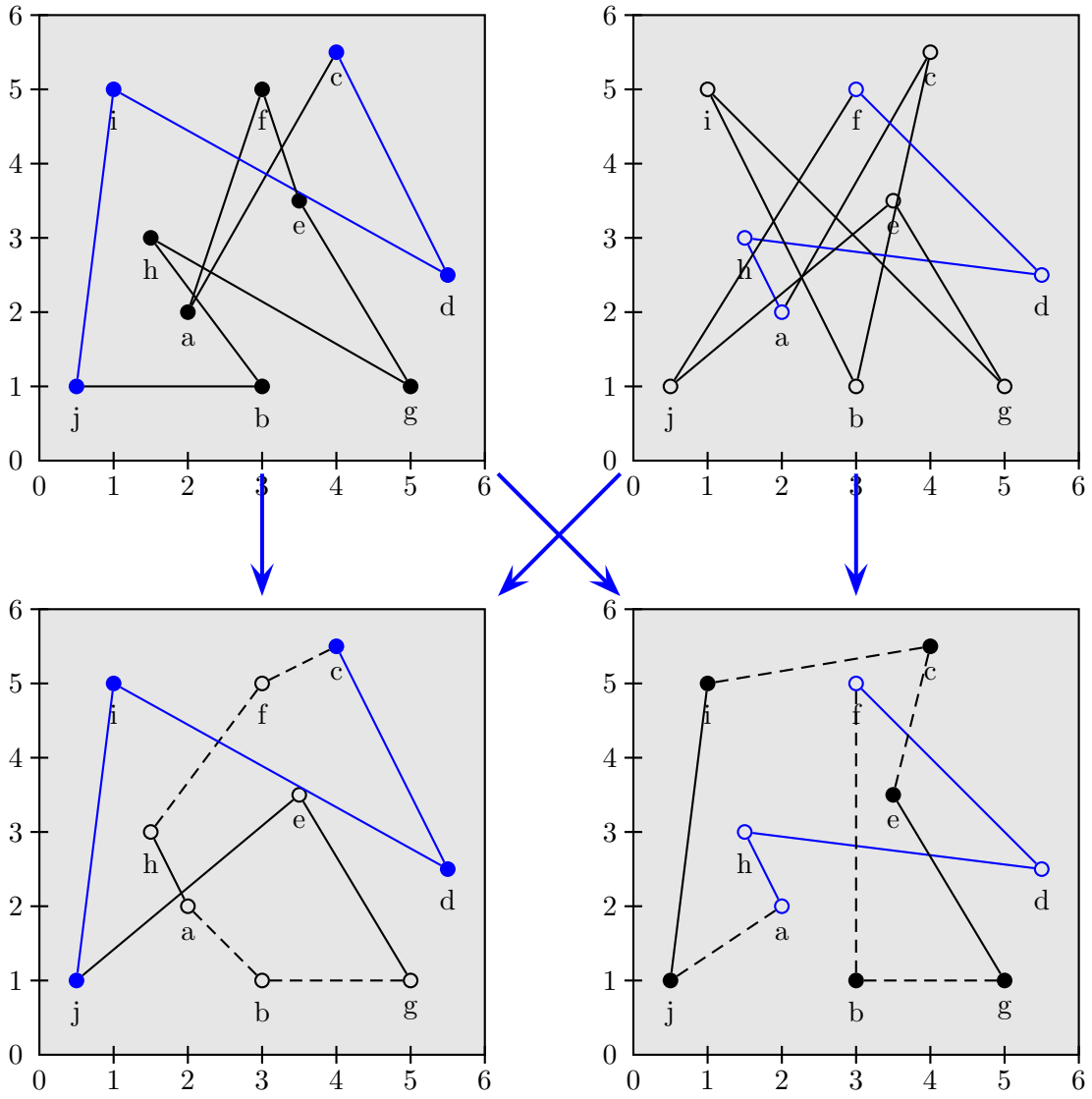


Abbildung B.4: Veranschaulichung von Abbildung B.3 für das TSP-Problem: Die komplett übertragenen Knoten und Verbindungen sind farbig markiert, Verbindungen, die beim Kopieren identisch geblieben sind, sind durchgängig gezeichnet.

B.2.2 Grundalgorithmus

Der Ablauf des Inselalgorithmus ist wenig komplex (s. Abbildung B.6). Zu Beginn wird auf jeder Insel eine Anfangspopulation nach den Regeln des dort implementierten Algorithmus erstellt. Dann werden in einer Schleife zunächst die einzelnen Inseln einer Evolution unterzogen in dem Sinne, dass auf jeder Insel der implementierte Suchalgorithmus vollständig durchgeführt wird. In einem zweiten Schritt der Schleife kommt es zur Migration zwischen den Inseln, in dem Sinne, dass Individuen, und damit auch Lösungen, in andere Inselpopulationen kopiert werden. Die Schleife wird mit einem Integrationsschritt beendet, in dem die besten Lösungen aller Einzelalgorithmen in einer Meta-Population gesammelt werden und die beste Lösung der gesamten Inselgruppe identifiziert wird. Ist die Abbruchbedingung an dieser Stelle erfüllt, endet der Lauf, ansonsten beginnt die Schleife von Neuem. Der Pseudo-Code für Inselalgorithmen , der sich nahe liegend aus dem Ablaufschema ergibt, ist in Listing B.2 dargestellt.

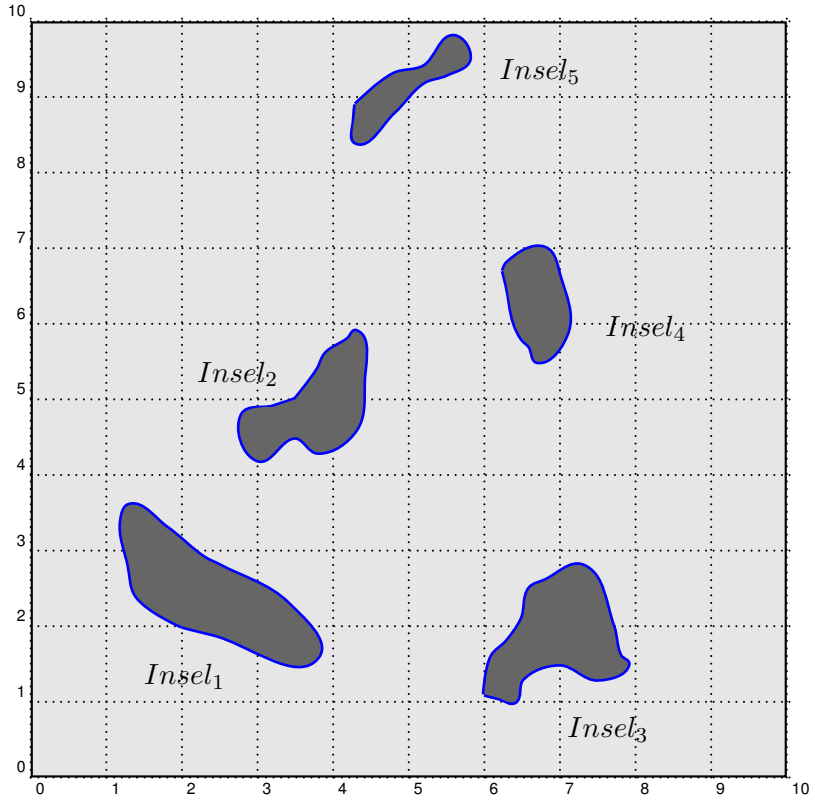


Abbildung B.5: Darstellung einer Inselgruppe

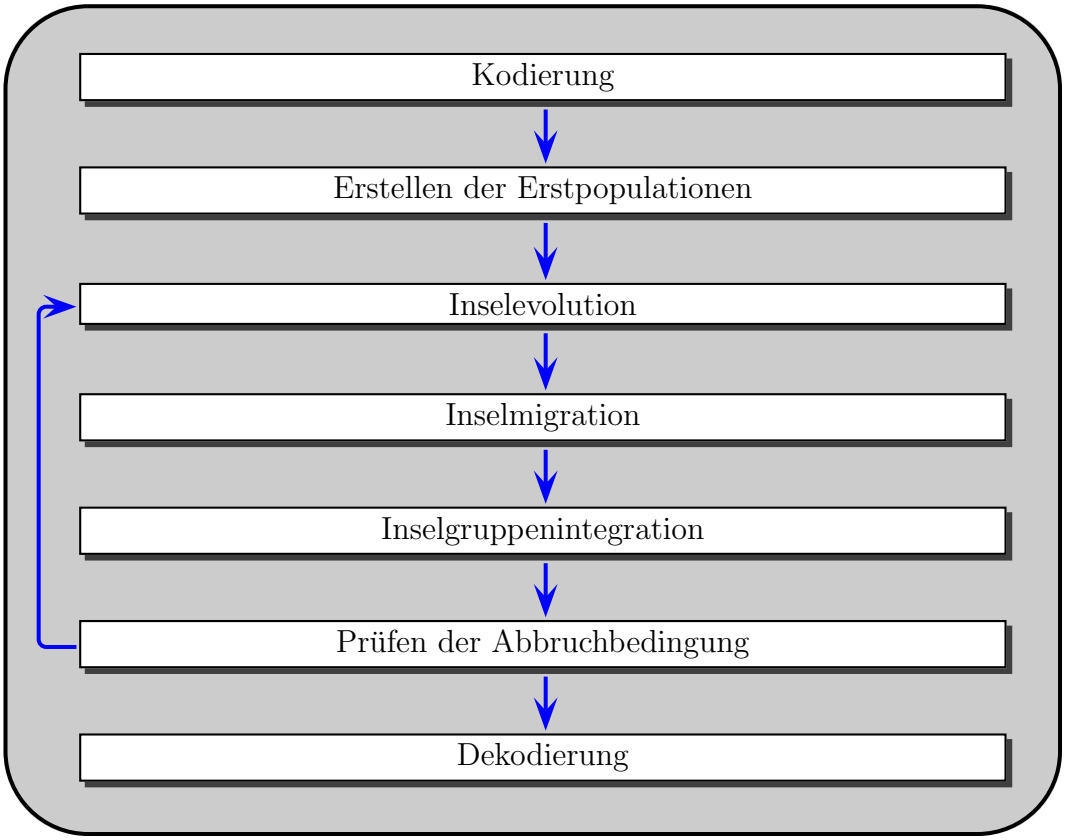


Abbildung B.6: Schritte eines Inselalgorithmus

```
METAISLAND(inselgruppe)
1  for i ← 1 to anz_inseln
2  do ERZEUGE_ ERSTE_ POPULATION(inselgruppe[i])
3  repeat
4          EPOCHE(insgrp)
5    until EPOCHEN_ ABBRUCHBEDINGUNG(insgrp) = True
6  b ← suche_bestes_individuum(insgrp)
7  return b

EPOCHE(insgrp)
1  for i ← 1 to insgrp.anz_inseln
2  do repeat
3              EVOLUTION(insgrp.insel[i])
4              INDIVIDUEN_ BEWERTEN(insgrp.insel[i])
5              for j ← 1 to insgrp.insel[i].anz_nachbarn
6              do MIGRATION(insgrp.insel[i], insgrp.insel[insgrp.insel[i].nachbarn[j]])
7        until GENERATIONEN_ ABBRUCHBEDINGUNG(insgrp.insel[i]) = True
8      BESTES_ INDIVIDUUM_ BESTIMMEN(insgrp.insel[i])
9
```

Listing B.2: Pseudo-Code für den Inselalgorithmus

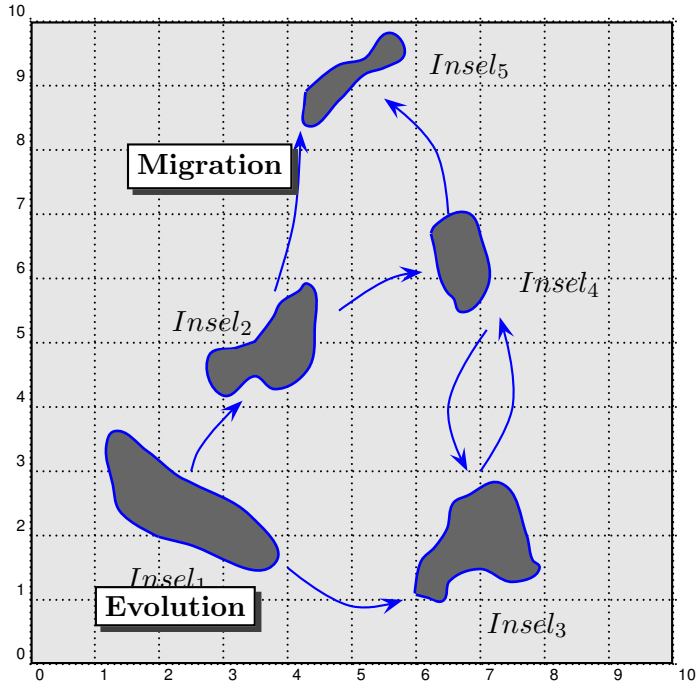


Abbildung B.7: Darstellung von Evolution und Migration am Beispiel einer Inselgruppe

Abbildung B.7 veranschaulicht die ersten beiden Schritte der Schleife. Auf jeder einzelnen Insel wird ein Evolutionszyklus, der auch Evolutionsepoche genannt wird, durchgeführt. Anschließend kommt es zu einem Kopieren von Algorithmen anhand einer vorher festgelegten Verbindungstopologie. Die abschließende Integration der Inselergebnisse zeigt Abbildung B.8. Damit gehören zur Komplettierung der Definition eines Inselalgorithmus die Bestimmung der zu implementierenden Einzelalgorithmen, die Bestimmung des Migrationsprinzips, die Wahl der Inselgruppentopologie und die Festlegung des Abbruchkriteriums. Diese Entscheidungen sollen der Reihe nach erläutert werden.

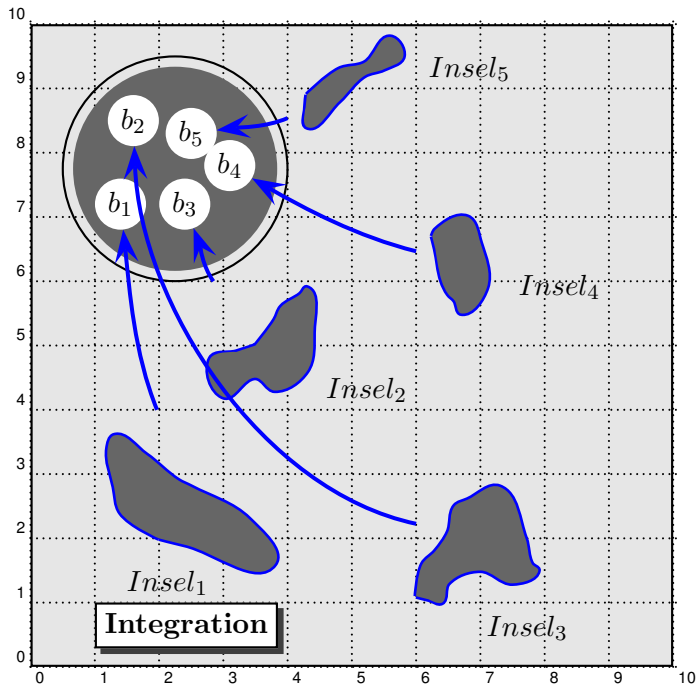


Abbildung B.8: Integration der Ergebnisse: Die besten Einzellösungen werden in eine Meta-Population kopiert

B.2.3 Einzelalgorithmen

Der Optimierungsschritt auf den Inseln wurde als „Evolution“ bezeichnet. Diese Sprechweise verweist auf den Ursprung der Inselmodelle, nämlich auf den Versuch einer Parallelisierung von Genetischen Algorithmen. Die Anwendungsmöglichkeiten sind aber nicht auf diese Algorithmenklasse beschränkt. Jeder Algorithmus, der in der Lage ist, eine Population aus Einzellösungen als Input zu verwenden, und der ebenso in der Lage ist, eine Population von Einzellösungen als Output auszugeben, kann in das Modell integriert werden, unabhängig davon, ob Zwischenschritte oder zusätzliche Algorithmen herangezogen werden müssen, um die Eingliederung zu ermöglichen. Es ist problemlos möglich, dass sich die Algorithmen zwischen den einzelnen Inseln unterscheiden, dass sogar unterschiedliche Algorithmenklassen gleichzeitig zum Einsatz kommen können.

Was gewährleistet werden muss, ist aber die Kompatibilität der Populationen. Eine unterschiedliche Kodierung kann durch einen Umkodierungsalgorithmus problemlos aufgehoben werden, wird hingegen Unterschiedliches repräsentiert, sind die Populationen inkompatibel. So kann eine Population, in der Einzelregeln kodiert werden, nicht ohne weiteres mit einer Population in Interaktion treten, in der Regelmengen kodiert sind. Könnte die zweite Population zwar Elemente der ersten Population noch als Regelmengen mit einem Element auffassen, besteht in der umgekehrten Richtung keine Möglichkeit zur Eingliederung der Individuen der zweiten Population. Solange es aber möglich ist, dass Populationen kompatibel sind, und die Algorithmen in der Lage sind, an einer bestehenden Population anzusetzen und eine Population mit gleichen Rahmenparametern auszugeben, können die Algorithmen zusammen in einem Inselmodell eingesetzt werden.

B.2.4 Migrationsformen

Nachdem auf jeder Insel der Optimierungsalgorithmus durchlaufen worden ist, werden Individuen zwischen den Inseln kopiert. Dabei werden Individuen auf jeder Insel zum Transfer ausgewählt, und es werden jene Individuen ausgewählt, die durch Kopien fremder Inseln ersetzt werden sollen. Dementsprechend ist der Begriff „Migration“ irreführend, da er zwar auf den Transfer von Individuen zwischen Inseln verweist, aber gleichzeitig eine materielle Bewegung impliziert.

Für die Darstellung der Prinzipien soll jede Lösung durch ihren Qualitätswert repräsentiert werden. Abbildung B.9 zeigt den Zustand auf einer einzelnen Insel vor der Migration. Zur Auswahl der Individuen, die auf andere Inseln transferiert werden sollen, bieten sich vor allem zwei Prinzipien an:

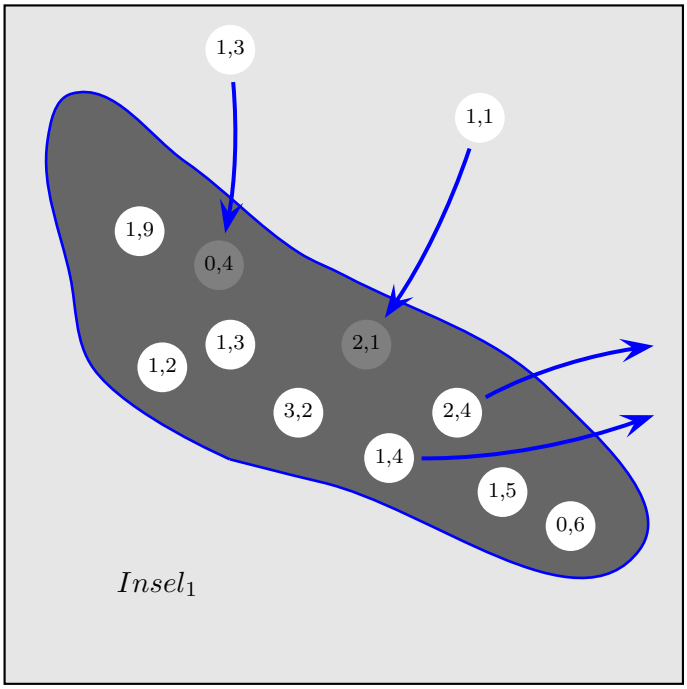


Abbildung B.9: Darstellung einer einzelnen Insel vor der Migration: Individuen der Population sind durch den Qualitätswert der Lösung repräsentiert

- Die besten Individuen werden gewählt (entspricht einer deterministischen Bestselektion).
- Die Individuen werden zufällig ausgewählt (entspricht einer Roulette-Wheel-Selektion mit Gleichwahrscheinlichkeit).
- Die Individuen werden durch einen anderen Selektionsoperator ausgewählt (z.B. Tournament Selection).

Zudem muss die Frage beantwortet werden, ob Mehrfachauswahlen eines Elements zugelassen werden, und wie diesbezüglich in Situationen verfahren wird, in denen eine Insel Individuen zu mehreren anderen Inseln schickt. Sind die „Migranten“ aus jeder Population in diesem Sinne festgelegt worden, muss bestimmt werden, welche Individuen durch die neuen Kopien ersetzt werden sollen. Auch hier gibt es eine Reihe von Alternativen:

- Die schlechtesten Individuen werden gewählt.
- Die Individuen werden zufällig ausgewählt.
- Die Individuen werden durch einen modifizierten Selektionsoperator ausgewählt (z.B. Tournament Selection mit Verliererauswahl).

Zwei mögliche Kombinationen von Auswahlkriterien sind in den Abbildungen B.10 und B.11 für zwei Inseln mit gegenseitiger Migration (z.B. die Inseln 3 und 4 im Beispiel) exemplarisch umgesetzt. Abbildung B.10 demonstriert die Kombination aus Bestenauswahl und Schlechtestenauswahl für Emigration und Immigration, Abbildung B.11 die Kombination aus zwei Zufallsauswahlen.

Die Wahl des Verfahrens erfolgt nach folgendem Kalkül: Hauptaufgabe eines Inselalgorithmus bzw. der Nutzen gegenüber der Konzentration auf ein einziges Suchverfahren ist die Möglichkeit, lokalen Minima auszuweichen. Diese Aufgabe wird behindert, wenn alle Inseln zu früh auf eine einzige beste Lösung oder wenige beste Lösungen konvergieren. Durch eine Zufallsauswahl wird demgegenüber erreicht, dass die Verfahren eher Hinweise auf andere Regionen des Suchraums erhalten, für die sich eine Überprüfung lohnt. So kann ein Schwarmpartikel in diese Region gesetzt werden, zum anderen kann in einem Genetischen Algorithmus neue genetische Information zur Verfügung gestellt werden. Der Vorteil gegenüber einer wiederholten Vorgabe der Einzelalgorithmen unabhängig voneinander besteht also in der möglichen Kombination aus gefundenen besten Lösungen mit fremden Lösungen.

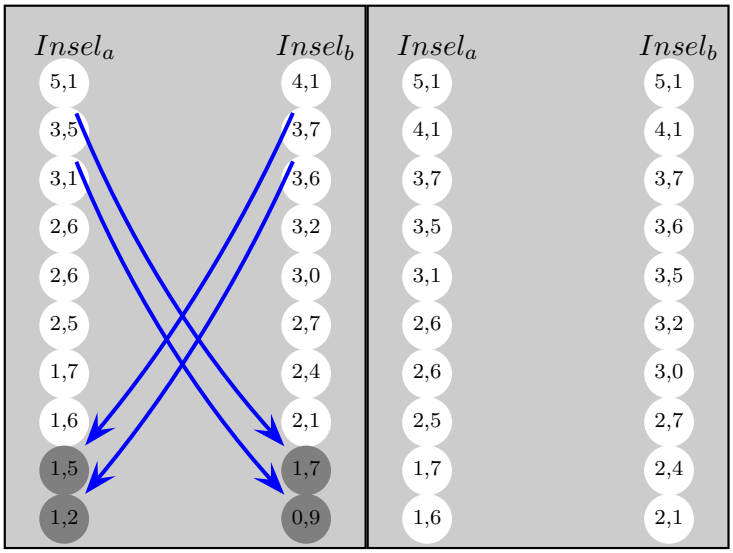


Abbildung B.10: Migrationsbeispiel 1: Bestenauswahl für Emigration und Schlechtestenauswahl für Immigration

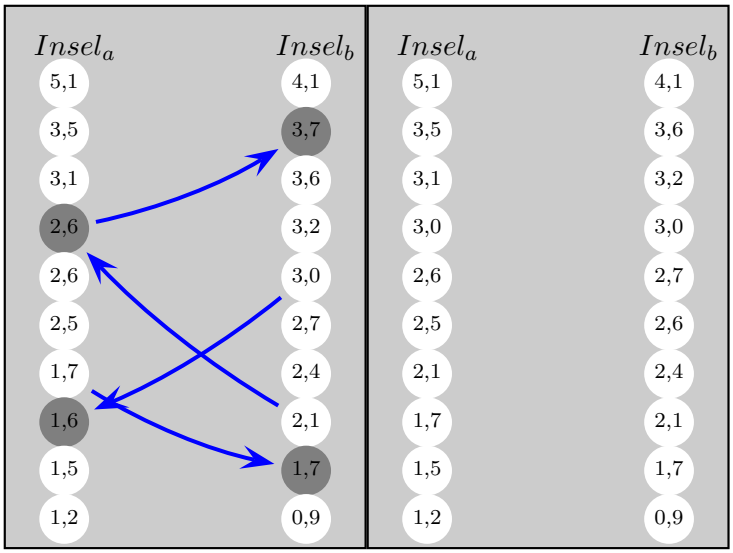


Abbildung B.11: Migrationsbeispiel 2: Zufallsauswahl für Emigration und Immigration

Es kann aber deshalb nicht gefordert werden, nur die Zufallsauswahl als sinnvolle Methode zuzulassen, da sie gegenübereiner Bestenauswahl unter Umständen die Laufzeit der Algorithmen verlängern kann. Dementsprechend muss die konkrete Wahl vor dem Hintergrund der Problemschwierigkeit, der Hardwareausstattung bzw. Prozessorvefügbarkeit und anderen pragmatischen Erfordernissen erfolgen.

B.2.5 Abbruchkriterium

Bei der Wahl eines Abbruchkriteriums für einen Inselalgorithmus sollte berücksichtigt werden, dass es sich bei dem Verfahren um einen Meta-Algorithmus handelt. Sollte das gleiche Kriterium angelegt werden wie für die Einzelalgorithmen, wäre die Suche nach einer Epoche abgeschlossen. Es stehen eine Reihe möglicher Abbruchkriterien zur Auswahl:

- Mindestqualität für die beste gefundene Lösung: In diesem Fall würde ein gegenüber den Einzelverfahren verschärftes Qualitätskriterium angelegt
- Feste Anzahl von Epochen: Der Algorithmus endet nach einer festgelegten Anzahl von Optimierungsschleifen unabhängig von der Art der gefundenen Lösung
- Feste Laufzeit des Algorithmus: Gerade bei begrenzten Ressourcen bietet sich eine Begrenzung der Laufzeit an, der Algorithmus endet nach einer festgelegten Zeit.
- Konvergenz der Lösungen auf den Einzelinseln: Der Algorithmus endet, wenn sich entweder die Lösungsqualität der besten Lösungen auf Einzelinseln hinreichend angenähert haben oder die Lösungen selbst bei Nutzung eines Ähnlichkeitsmaßes als hinreichend ähnlich klassifiziert werden können. Dabei kann die Bedingung insofern verschärft werden, dass dieser Zustand über eine Anzahl zusammenhängender Epochen beobachtet werden kann.
- Stagnation des Fortschritts auf den Einzelinseln: Der Algorithmus endet, wenn sich durch Migration und Evolution die besten Lösungen auf den Einzelinseln für eine Anzahl von Epochen nicht mehr verändern.

Die Wahl des Kriteriums obliegt erneut dem Anwender unter Berücksichtigung von Problemschwierigkeit, verfügbaren Ressourcen und Vorerfahrungen mit Datensatz und Algorithmen.

B.2.6 Inseltopologien

Neben der Art und Weise, wie Individuen zwischen Inseln ausgetauscht werden, muss zusätzlich bestimmt werden, zwischen welchen Inseln Individuen ausgetauscht werden sollen. Dabei bieten sich Migrationstopologien analog zu den Kommunikationstopologien von Partikelschwärmen an. Beziehungen können reziprok sein, jede Insel kann mit mehreren Inseln in Kontakt stehen, sowohl als Quelle als auch als Ziel für Migrationsbewegungen. Die folgenden Topologien befinden sich im Einsatz und können einen Einblick in die Möglichkeiten der Algorithmengestaltung geben.

Stepping Stones

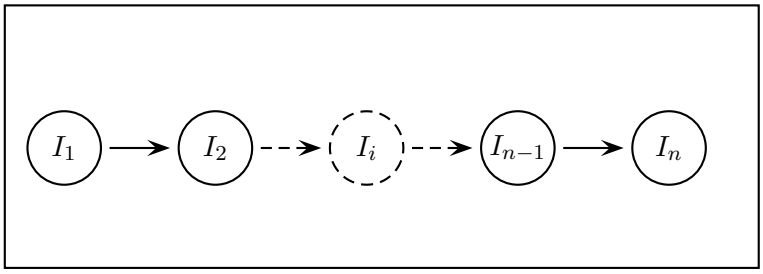


Abbildung B.12: Einseitige Stepping-Stones-Topologie

Eine Stepping-Stones-Topologie ist eine nicht-geschlossene Kette von Inseln mit reziproker oder einseitiger Migrationsbeziehung.Dementsprechend gibt es zwei äußere Inseln, die nur mit maximal einer anderen

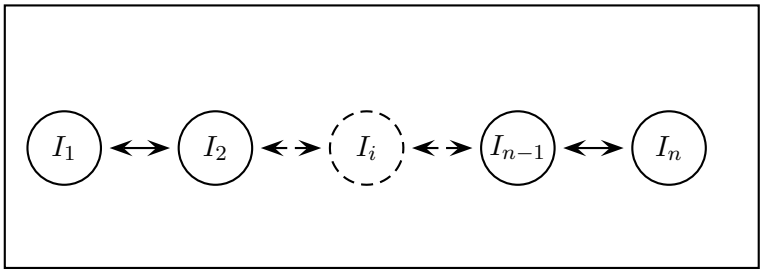


Abbildung B.13: Reziproke Stepping-Stones-Topologie

Insel verbunden sind, und die voneinander maximal entfernt sind. Demenstprechend kann die Information, die an den beiden Seiten der Reihe vorhanden ist, sehr unterschiedlich sein. Abbildung B.12zeigt eine einseitige, Abbildung B.13 eine reziproe Stepping-Stones-Topologie.

Kreise

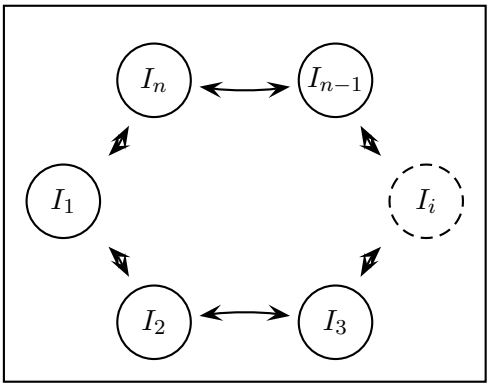


Abbildung B.14: Reziproke Kreistopologie

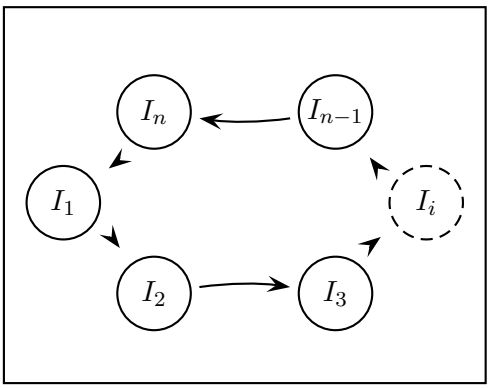


Abbildung B.15: Einseitige Kreistopologie

Wird die Kette einer Stepping-Stones-Topologie geschlossen, dann ist das Resultat eine Kreistopologie, die ebenfalls wahlweise einseitig oder reziprok gestaltet sein kann (s. Abbildung B.15 und Abbildung B.14). In diesen Topologien sind im Kreis einander gegenüberliegende Inseln maximal weit entfernt, die Information kann in zwei Richtungen schneller verteilt werden als zuvor, da die durchschnittliche Entfernung zwischen zwei Inseln kleiner ist.

Werden mehrere Kreise gebildet, so dass Elemente in inneren Kreisen mit Elemente in äußeren Kreisen verbunden werden (s. AbbildungB.16), gibt es weitere Verbreitungsmöglichkeiten.

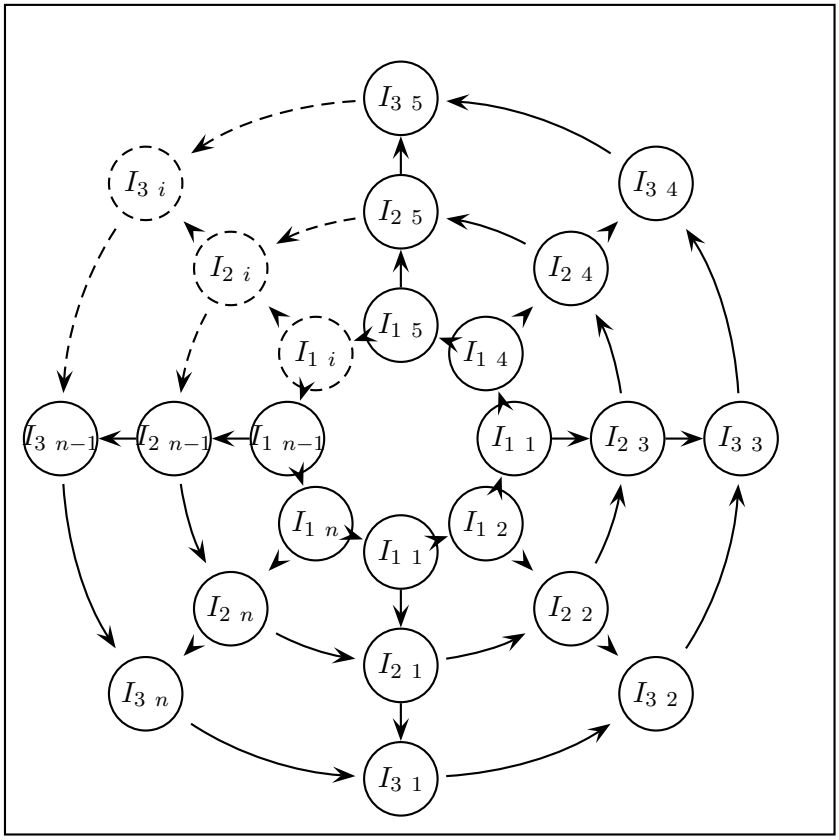


Abbildung B.16: Mehrfachkreistopologie

Gitter

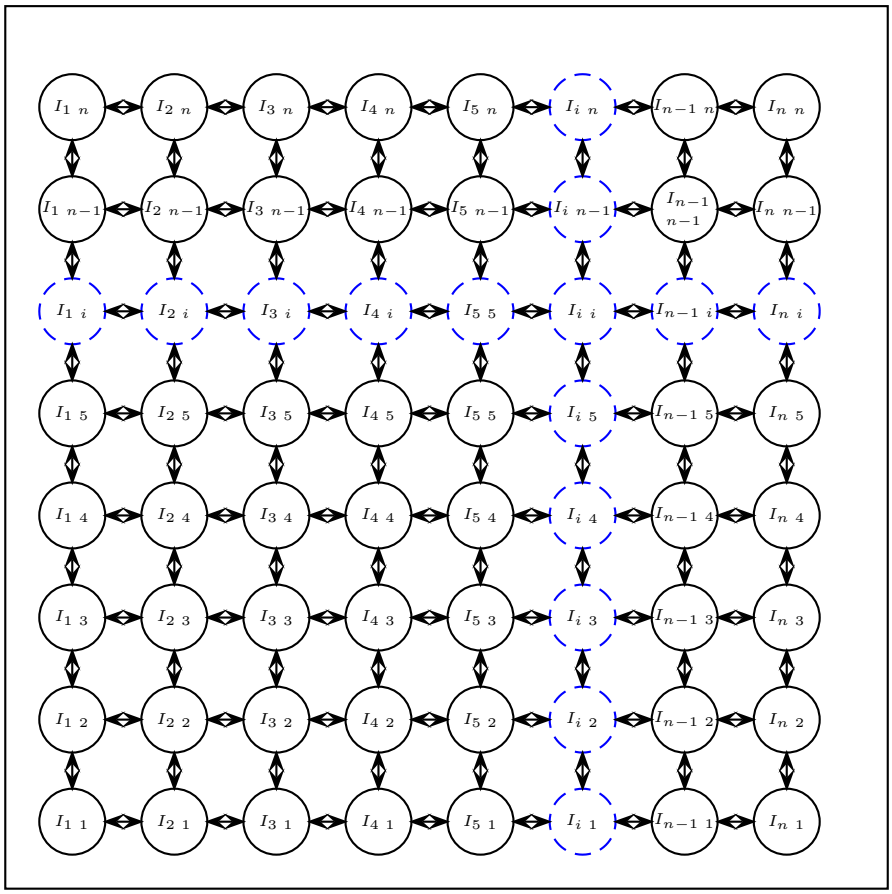


Abbildung B.17: 2D-Gitter-Topologie

In einem 2D-Gitter ist jeder Agent mit zwei bis vier anderen Agenten verbunden, ähnlich wie Schachbrettfelder auf einem Schachbrett miteinander geteilte Grenzlinien besitzen (s. Abbildung B.17). Information ist lokal verteilt, breitet sich aber in mehrere Richtungen aus, zwei gegenüberliegende Eckfelder sind maximal voneinander entfernt. Naheliegend ist die Erweiterung um eine dritte Dimension, so dass sich die Inseln als Würfel in einem Kubus darstellen ließen.

Torusformen

Verbindet man die gegenüberliegenden Seiten eines 2D-Gitters miteinander, so erhält man eine Torusstruktur (s. Abbildung B.18. Im Unterschied zum Gitter, aber auch zur Mehrkreis-Topologie gibt es in einem Torus keine Ränder, so dass Information immer in vier Richtungen weiterfließen kann. Diese Topologie ist eine beliebte Topologie für Simulationen, da keine Bewegungsgrenzen berücksichtigt werden müssen.

Pyramiden

Eine besondere Topologie stellt die Pyramiden-Topologie dar. Hierbei erhalten höhere Ebenen gebündelte Informationen von niedrigeren Ebenen, so dass an der Spitze der Pyramide die gesamte in der Pyramide verteilte Information zusammenfließen kann. Dabei kann Information auch horizontal auf gleicher Ebenenstufe ausgetauscht werden.

Mit dieser Topologie soll der Überblick abgeschlossen werden, da die wesentlichen Möglichkeiten in den beschriebenen Modellen enthalten sind.

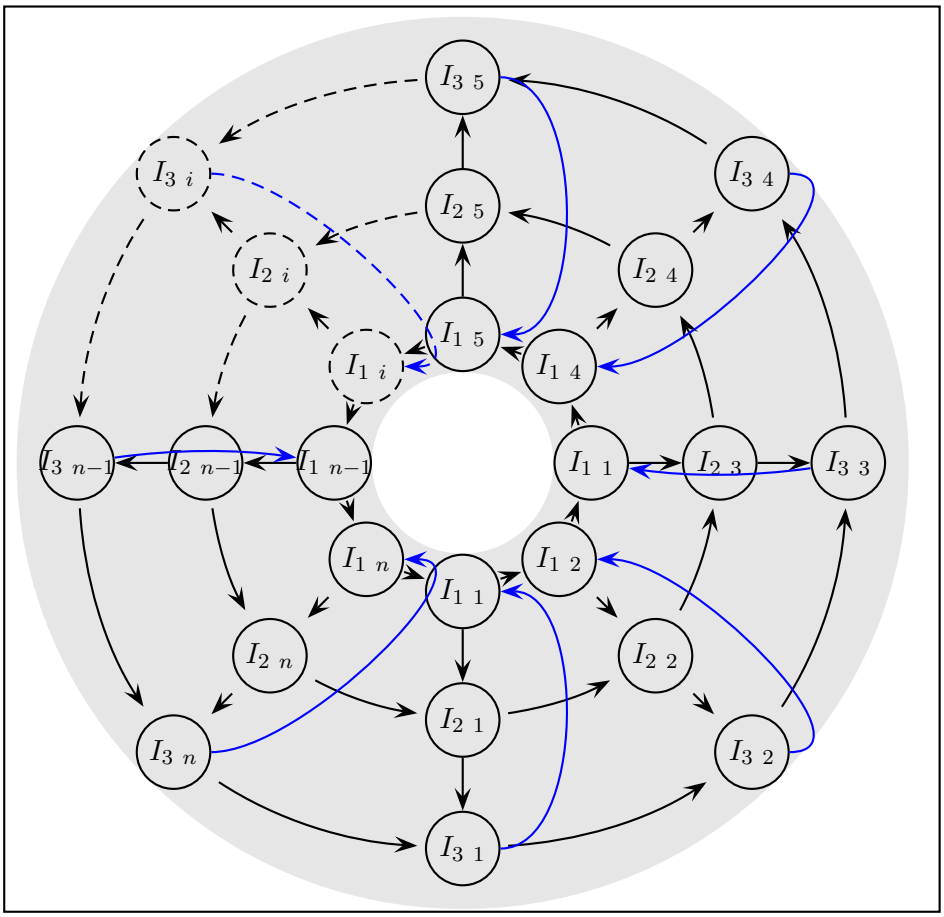


Abbildung B.18: Torustopologie

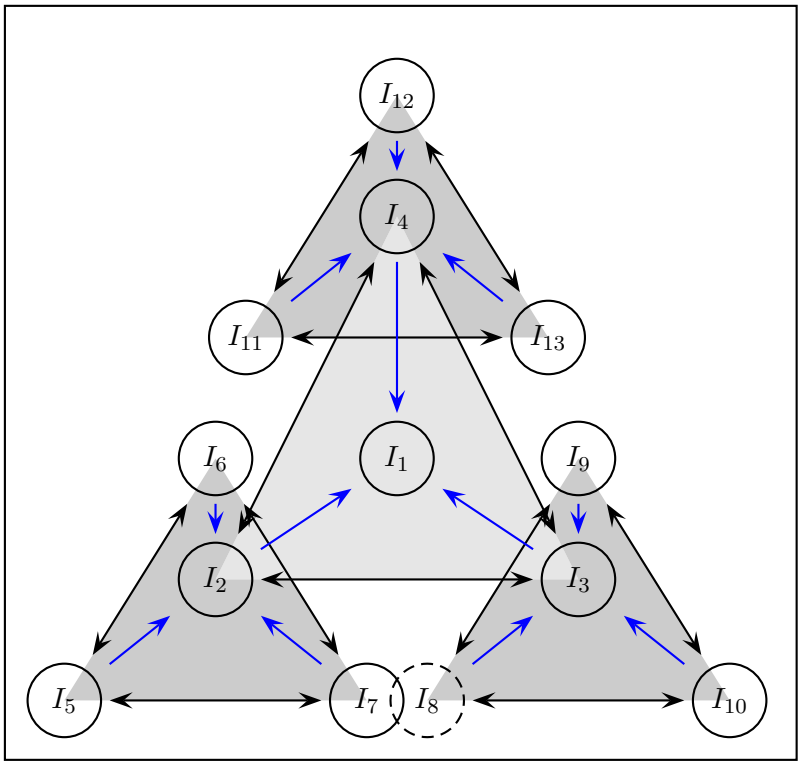


Abbildung B.19: Pyramidentopologie

B.2.7 Anwendungen von Inselalgorithmen

Exemplarisch soll hier nur auf wenige Anwendungen von Parallelen Genetischen Algorithmen in der Form von Inselalgorithmen eingegangen werden: Liu, Iba und Ishizuka (2001) experimentieren mit einem Inselalgorithmus Auswahl informativer Gene in der Gewebeklassifikation, stellen aber keine empirischen Vergleiche mit anderen Agorithmen an. Sie verwenden dabei zwei Algorithmen, die auf zwei unterschiedlichen Prozessoren gestartet werden.

Herrera, Lozano und Moraga (1999) stellen einen hierarchischen Inselalgorithmus vor, bei dem Ringe von genetischen Algorithmen eine untere Ebene bilden, auf der lokale Migration untereinander erfolgt und zusätzlich Individuen zwischen Ringen ausgetauscht werden. Bei der Zusammenstellung der Einzelalgorithmen berücksichtigen die Autoren die Möglichkeit heterogene Algorithmen mit unterschiedlichem Explorationsgrad in verschiedenen Kombinationen zusammenzustellen. Im Vergleich zu einfachen Algorithmen stellen die Autoren fest, dass Reliabilität und Genauigkeit der Algorithmen höher sind, dass zum anderen auch die Unterschiedlichkeit der untersuchten Lösungen größer wird.

Eby et al. (1997) nutzen für das Design von *Flywheels* ebenfalls mit einem hierarchischen Ansatz die Möglichkeit, Teilprobleme bzw. eine schnelle Voroptimierung auf niedrigen Stufen mit schnell optimierenden und eher oberflächlich suchenden Genetischen Algorithmen zu vollziehen, um auf höheren Stufen dann gründlicher Analysen durchzuführen. Der oben vorgestellte Pyramidenansatz wäre für diesen Einsatz ebenfalls geeignet. Zu einer positiven Einschätzung der Leistungsfähigkeit von Inselalgorithmen kommen auch Alba und Troya (2002) bei Anwendung einer Testsuite, bestehend aus natürlichen und künstlichen Datensätzen.

Tip 30: You Can't Write Perfect Software
Hunt und Thomas (2000)

C
Entwicklung und Struktur der begleitenden Software

Die Programmierung einer Software wurde unter anderem als Mittel zur Verhinderung software-getriebenen Denkens in der Datenanalyse vorangetrieben. Diesen zunächst kontraintuitiven Zusammenhang verdeutlichen Pedhazur und Schmelkin (1991):

> Procedures included in popular software packages exert undue (we are tempted to say almost total) influence on the kind of analyses that are being carried out. Users, even the more sophisticated ones, tend to limit their thinking about choices of analytic techniques in terms of software availability. This is attested to, among other things, by sudden spurts in the use of specific types of analyses following their incorporation in some statistical package. (S. 345)

Naheliegende Analysefehler entstehen nach Ansicht der Autoren durch Dateneingabefehler, durch die Verletzung von Modellvoraussetzungen oder Bugs in Software. Wenige Programme setzen deutliche Warnsignale, wenn eine dieser Fehlerarten auftritt - ein Umstand, der nicht für alle Anwender selbstverständlich ist. Bugs treten auch in kommerzieller Standardsoftware auf, vor allem nach der Revision bestehender Code-Teile. Ein Lesen der Update-Information zeigt oft die Schwere von wohlgemerkt gefundenen Bugs. Für namhafte Produkte am Markt ist es deshalb nicht ausgeschlossen, dass

- unsichtbar für den Anwender falsche Algorithmen gestartet werden.
- unsichtbar für den Anwender Prozeduren falsche Ergebnisse liefern
- Programmteile reproduzierbar zu Systemfehlern führen.

Der letztgenannte Fehler ist dabei der harmloseste Fehler, da keine fehlerbehafteten Ergebnisse erzeugt werden. Die vorliegende Software ist nahezu mit Sicherheit nicht fehlerfrei, teilt diese Eigenschaft allerdings mit nahezu jedem Beispiel von verfügbarer Software.

Die Software wurde vollständig in Object Pascal (Borland GmbH, 2001) mit *Borland Delphi 6.0* für Windows-Systeme programmiert. Diese Sprache stellt die Möglichkeit der objektorientieren Programmierung bereit und verfügt gleichzeitig über die Möglichkeit grafische Komponenten zu nutzen, die leicht in die Windows-API eingebettet werden können. Damit können Fenster und Steuerelemente eingesetzt werden, die dem Windows-Benutzer vertraut sind. Der objektorientierten Programmierung soll ein eigener Abschnitt gewidmet werden. Zunächst werden neu entwickelte Algorithmenklassen und Erweiterungen vorgestellt, die nicht im Rahmen der Arbeit bereits theoretisch eingeführt worden sind, um dann Design- und Implementierungsentscheidungen erläutern zu können. Es folgt eine Darstellung von Objektorientierung und Visualisierung, damit dann die implementierten Algorithmen im Detail präsentiert werden können. Abschließend wird ein Überblick über zusätzliche Komponenten von GIER$_{[1.0]}$gegeben. Neben GIER$_{[1.0]}$wurde im Rahmen der Arbeit auch der Werkzeugkasten GIER$_{[1.0]}$.simple_heuristics entwickelt. Dieser zweite Werkzeugkasten sowie eine letzte Komponente von GIER$_{[1.0]}$sollen im Anschluss beschrieben werden.

C.1 Design- und Implementierungsentscheidungen

C.1.1 Modularisierung der Arbeitsschritte

Trennung von Algorithmenentwurf und Anwendung der Algorithmen

Um zu unterstreichen, dass der Entwurf von Algorithmen unabhängig ist von der konkreten Anwendung des Algorithmus, wurde die Algorithmendefinition in GIER$_{[1.0]}$von der Verknüpfung mit Datensätzen oder der Spezifikation von Variablen- oder Fallzahlen getrennt. Die Angemessenheit eines entworfenen Algorithmus für ein konkretes Problem ist von den Merkmalen des Datensatzes abhängig, aber für die reine Spezifikation der Algorithmenform ist der Datensatz verzichtbar. Die Anzahl und Art der Operatoren, die Setzung der Laufparameter, die Bestimmung von Abbruchkriterien erfolgen allesamt ohne die Notwendigkeit, Fall- oder Variablenzahl zu kennen. Der gleiche Entwurf ist anwendbar auf eine Vielzahl von Datensätzen. Dementsprechend wurde hier entgegen der Konvention von Statistikpaketen entschieden, keine Menüs zu implementieren, in denen Eigenschaften des Algorithmus und Eigenschaften der Stichprobe gleichzeitig festgelegt werden.

Trennung von Algorithmenentwurf und Datensatz

Die Weitergabe von Datensätzen unterliegt den Bestimmungen des Datenschutzes, es gehört zudem zur Politik vieler Institutionen, dass keine Kopien der intern gespeicherten Datensätze herausgegeben werden. Sind aber Statistikanwender und Datenbank räumlich voneinander getrennt, wird die statistische Auswertung stark erschwert. Die Automatisierungsmöglichkeiten vieler Statistikpakete bieten eine Möglichkeit der Trennung von Algorithmenentwurf und -applikation am Datensatz. Oft ist diese Trennung aber in dem Sinne unvollständig, dass Datenstruktur oder Variablennamen präzise übermittelt werden müssen, um automatisierte Anweisungen zu erstellen. Die vorliegende Lösung benötigt diese Informationen nicht, sondern bezieht sich statt dessen auf eine festgelegte Struktur von Datensätzen. Damit wird ermöglicht, dass Algorithmen ohne Einsicht in die Daten entwickelt werden und auch Auswertungsprojekte einsatzbereit zur Verfügung werden können, um dann am Datensatz ohne großen Aufwand oder Kenntnisse über die Theorie der Algorithmen zum Einsatz gebracht zu werden.

Trennung von Algorithmenentwurf und Script

Der Scriptentwurf beinhaltet die Kombination von Datensätzen und Algorithmen, definiert Ausgabeoptionen, Klassifikationsaufgaben und Ausgabeoptionen. Diese Facetten sind logisch getrennt von der Erstellung der Algorithmen. Da sich mit Ausnahme der Sondermodule alle Algorithmen in Input und Output gleichen, ist die Trennung der Entwürfe von Algorithmus und Script naheliegend. Anders als in vielen Statistikpaketen kann so ein entworfener Algorithmus in verschiedenen Scripten und mit verschiedenen Ausgabeoptionen zum Einsatz gebracht werden, ohne dass Änderungen in der Prototypendatei des Algorithmus erfolgen müssen. Gleichzeitig kann in der Festlegung neuer Scripte auf die Prototypen ohne Modifikation zurückgegriffen werden.

C.1.2 Die Script-Sprache gier.script

Besonderes Augenmerk beim Design der Software wurde auf eine funktionsfähige Organisationsmöglichkeit für Auswertungsvorgänge gelegt. Dementsprechend wurde eine Entscheidung getroffen gegen die durchaus übliche Implementierung einer Möglichkeit, nach Eingabe von Parametern in Fenstern direkt Algorithmen ansteuern zu können. Der Vorteil dieser getroffenen Entscheidung liegt zum einen in einer verbesserten Dokumentation von Arbeitsschritten und der Reproduzierbarkeit von Entscheidungen. Zu diesem Zweck wurde eine differenzierte Scriptsprache erstellt, die als Mittler zwischen Wünschen des Softwareanwenders und der Umsetzung dieser Wünsche durch die Auswahl von Programmroutinen fungiert. Die Scriptsprache hat die Aufgabe, alle Spezifikationen zu kodieren, wobei ein Parser in der Lage sein muss, das in der Sprache Kodierte auch wieder zu entschlüsseln. Das macht die Definition von Konventionen notwendig, da die Scriptsprache eine dekodierbare Struktur aufweisen muss. Um die modulare Entwicklung zu gewährleisten und die Weiderverwendbarkeit des Entworfenen wahrscheinlicher zu machen, teilen sich die Scriptdateien in vier Typen auf:

- **Prototypendateien** enthalten die genauen Spezifizierungen für Algorithmen zusammen mit allen Designentscheidungen und Parametersetzungen.

- **Taskbeschreibungen** enthalten Angaben über Ausgabeoptionen, verwendete Datensätze und Prototypendateien sowie über die Art der gewünschten Auswertungen und **Subtasks**.

- **Scriptdateien** enthalten neben Entscheidungen über Ausgabe- und Protokolloptionen Angaben über die auszufühenden Tasks.

- **Meta-Scriptdateien** enthalten Angaben über auszuführende Scripte.

Dabei treten Taskbeschreibungen nur innerhalb von Scriptdateien auf. Zum Ausführen von Algorithmen sind Script-Dateien erforderlich, Meta-Scriptdateien sind ein optionaler Dateityp, mit dem mehrere Scriptdateien nacheinander ausgeführt werden können.

Prototypen in Gier$_{[1.0]}$.script

Jeder Algorithmenentwurf kann in einer Prototypendatei gespeichert werden, die alle Informationen enthält, die den Algorithmus definieren. Damit muss jeder Algorithmemtyp nur einmal über das Menü definiert werden, und steht in der Folge als Datei zur Verfügung. Die Abspeicherung der Algorithmen erfolgt über einen Parser, der jeden Algorithmus entsprechend seiner Objektstruktur in einen Text überführt. Jeder Bestandteil wird dabei einzeln in Textzeilen übersetzt, während umgekehrt aus diesen Textzeilen wiederum das entsprechende Objekt rekreiert werden kann. Jede Prototypendatei beginnt mit einer Kopfzeile, die den Namen des Prototypen gefolgt von einem Slash („/“) enthält. In der zweiten Zeile steht der Typ des Algorithmus in Klammern, gefolgt mit einer Zeile, die generelle Parameter und Eigenschaften des Algorithmus enthält. Dabei bestehen Angaben jeweils aus einer Bezeichnung, einem Gleichheitszeichen und der Wertzuweisung und werden voneinander erneut durch einen Slash getrennt. Nachfolgende Zeilen enthalten Angaben über die Objektklassen, aus denen die Komponenten des Algorithmus bestehen. Zeilen beginnen mit dem Namen der Klasse, gefolgt von zwei Sternchen („**“), optional gefolgt von der Liste der Parameter für diese Objekte analog zu den Eigenschaften des Algorithmus.

```
/Beispiel-Prototyp
(gier.ga_script)
Mind.-n(Regel)=50/ Restf.=False/ Pop.gr.=20/ Neu-Pop.gr.=20/ Generationen=0/
                 Datens.=False/ Regels.=False/ Aktivg=True/ abd.v.reg.=True/
                 Max.Var.=3/ Elit.=False/
    t_creation_rule**  U.repair=False/ R.repair=True/ Aktivg=True/ Max.var=3/
                                                                  Snowbd=False/
        t_creation_decision_all_x**  x=2/
        no_snowboard**
    select_all**
    t_objective_tq**
    t_fitness_notransformation**
    t_scalefitness_none**
        t_partition_size**
    Operatoren**  Anzahl=3/ Mutationen=1/ Cross-Overs=0/ Mut. Aktivgen=1/
                              Cr.-Ov. Aktivgen=1/ mu/ 0/ ma/ 0/ ca/ 0/
    t_mutation_standard**  p=0,3/
    t_mutation_active**  p=0,1/
    t_cross_over_active**  cross_two=False/ species_rule=False/
                                      cross_probability=0,3/
          t_cross_detail_uniform_50**  probability=0,3/
    t_reparatur_randomrepair**  One-Scale=False/
    t_bonus_fitness_no_bonus**
    t_penalty_fitness_no_penalty**
    t_selection_wheel**
        t_partition_size**
    t_stopcheck_generationen**  mind.Generationen=0/
     hoechst.Generationen=10000000/ Generationen=20/
```

Listing C.1: Beispiel-Prototyp in $\text{GIER}_{[1.0]}$.script

Listing C.1 ist ein Beispiel für einen in $\text{GIER}_{[1.0]}$.script formulierten Algorithmus. Die Bestandteile der Algorithmenklassen werden in Abschnitt C.3 ausführlich dargestellt, so dass die Detailebene der Prototypen-Programmierung dem Programmcode entnommen werden kann, die Beschreibung hier aber nicht weiter vertieft werden soll. Jede Prototypenklasse wird mit einer eigenen Endung gespeichert, so dass die Art des Algorithmus am Dateinamen erkannt werden kann.

Tasks, Scripte und Meta-Scripte

Durch *Scriptdateien* wird die Ausführung zeitintensiver Rechenoperationen vereinfacht, da eine Liste von Auswertungen der Reihe nach abgearbeitet werden kann, ohne dass zwischenzeitlich Eingaben des Anwenders notwendig sind[1]. Zudem kann gleichzeitig das gewünschte Ausgabeformat festgelegt und die Art der zu produzierenden Grafiken und Diagramme bestimmt werden. Ein Script besteht aus einer Spezifikation von Ausgabeoptionen und einer Liste von Tasks. In Tasks werden die Prototypen von Algorithmen mit Datensätzen kombiniert, zudem wird die Art der Auswertung vorgegeben. Zum einen gibt es die Möglichkeit eine Menge von Datensätzen mit einer Menge von Algorithmen zu kombinieren, wobei im Ergebnis die Fitting-Ergebnisse nebeneinander gestellt werden können.

Es stehen aber neben dem reinen Fitting auch Möglichkeiten der Kreuzvalidierung zur Verfügung, zum einen Trainingsset/Testset-Splits zur Prüfung der Generalisierung und die Option der k-fachen Kreuzvalidierung. Ergebnisse können wahlweise zusätzlich auch als Pseudo-Code oder als SPSS-Syntax zum Klassifizieren von Objekten mit gleicher Variablenstruktur ausgegeben werden und auch unter Angabe der zum Einsatz gekommenen Regel erstellt werden. Weiterhin können Variablenstatistiken und eine Klassifikationstabelle angegeben werden. Die Ergebnismodelle der Algorithmen können zudem an zusätzlichen Datensätzen geprüft werden, die auch unter Angabe der Regelnummer für Objekte klassifiziert werden können. Als Diagrammoptionen stehen deskriptive Variablendiagramme, 2D-Scatter-Plots, 2D-Shading-Plots, Stichprobenüberblicksdiagramme, Regelliniendiagramme und Varianten von Entscheidungsbäumen zur Verfügung, deren Einstellungen und Parameter über eine Parameterdatei spezifiziert werden. Ein Script

[1] Diese Erleichterung kann solange genutzt werden, wie genügend Arbeitsspeicher für die Speicherung der Ergebnisse zur Verfügung steht

wird ebenfalls über einen Parser in einer Textdatei gespeichert und kann durch Rückgriff auf die Textdatei gestartet werden.

```
beispiel_script/
(gier.script)
    Auswahl=2/ Ausw_aktuell=0/
    pr_pseudo=False/ pr_syntx=True/ pr_inter=True/ nr_pseudo=False/ nr_syntx=True/
                                                    prot=True/ desk_prot=False/
    klass=True/ klass_nr=True/ and_daten=True/ and_nrx=True/ anz_and_daten=2/
    Diag_Einst=diagramme_test.gdg/ Diag_regel=True/ Diag_dtstr=True/ Diag_dtemp=True/
    Diag_deskr=True/ Diag_2Dsc=True/ Diag_2Dsh=True/ Diag_pop=True/
    daten_zeigen=False/ display1=False/ display2=False/
        ** Auswertungsliste
           [3 x 2]  (swarmset1.gss, downhill1.gdh, downhill2.gdh)
                       (daten1.dat, testdatei mit namen.dat) Wbw
           [CV x 4]  (ga1.gga.gdh, gaset1.ggs, gaset2.ggs, ga2.gga)
                                   (daten2.dat) [k 6] Wbw vll
        ** Klassifikationsdatensätze
            daten3.dat
            daten4.dat
```

Listing C.2: Beispiel-Script in GIER$_{[1.0]}$.script

Listing C.2 ist ein Beispiel für eine Script-Datei, in der ein Algorithmenvergleich und eine 6-fache Kreuzvalidierung spezifiziert sind. Zusätzlich sollen weitere Datensätze mit den Ergebnisregeln klassifiziert werden.

Metascripte sind Scripte, die mehrere Scriptdateien der Reihe nach ausführen. Metascripte sind dann nützlich, wenn sich etwa die Ausgabeeinstellungen für verschiedene Auswertungen ändern oder mehrere Methodenvergleiche der Reihe nach ausgeführt werden sollen. Auch Metascripte werden über einen Parser in Textdateien gespeichert und werden aus diesen Dateien gestartet. Für Metascripte wird in der Ausgabe ein eigenes Verzeichnis angelegt, in das die Ergebnisse der einzelnen Scripte als Unterverzeichnisse geschrieben werden.

```
** gier.scriptliste **
5
script1.gsc
script2.gsc
script3.gsc
script4.gsc
script5.gsc
```

Listing C.3: Beispiel-Meta-Script in GIER$_{[1.0]}$.script

Da die Parameterwerte in den aufgerufenen Scriptdateien gespeichert sind, sind Meta-Scriptdateien sehr einfach strukturiert. Listing C.3 ist ein Beispiel für den Inhalt einer Meta-Scriptdatei. Die Software nutzt diese Dateien, um die Scripte der Reihe nach abzuarbeiten.

C.1.3 Dateiformate

Eine Liste mit den verwendeten Dateitypen ist Tabelle C.1 zu entnehmen. Es soll hier kurz auf die Beschreibung der Dateiarten eingegangen werden:

Alle Prototypendateien und alle Scriptdateien sind einfache Textfiles, die allerdings eine vorgegebene Textstruktur aufweisen. Lesbar und bearbeitbar sind die Dateien mit jedem herkömmlichen Texteditor. Ähnliches gilt für Protokolldateien. Einige Protokolldateien sind im *Rich Text* Format gespeichert und brauchen zum Lesen einen Texteditor, der mit Formatierungsbefehlen umgehen kann. SPSS-Syntaxdateien sind ebenfalls Textdateien, erfordern zu ihrer Verwendung aber die Lizenz und Installation der SPSS-Software. Bilddateien werden im Bitmap-Format ausgegeben, das zu den meisten gängigen Gra-

Tabelle C.1: Liste der Dateitypen in GIER[1.0]

Datei-Typ	Dateiendung	Dateiart
Downhill	*.gdh	Prototyp
GA	*.gga	Prototyp
GASet	*.ggs	Prototyp
MetaIsland	*.gmi	Prototyp
R2-F2	*.gr2	Prototyp
Swarm	*.gsw	Prototyp
Swarmset	*.gss	Prototyp
TSP-GA	*.gts	Prototyp
TYPAG	*.gty	Prototyp
Script	*.gcs	Script-Datei
Meta-Script	*.gms	Meta-Script-Datei
Script-Speicherung	*.gin	Fortschrittsdatei
Daten	*.dat	Datendatei
Diagrammeinstellungen	*gdg	Parameterdatei
Programmeinstellungen	*.ini	Parameterdatei
Diagramm	*.bmp	Bilddatei
Protokoll	*.txt	Textdatei
Formatiertes Protokoll	*.rtf	Formatierte Textdatei
SPSS-Syntax	*.sps	SPSS-Syntax-Datei
Programmdatei	*.exe	Ausführbare Datei

fikprogrammen kompatibel ist.

Es können in die Software alle ASCII-Textdateien eingelesen werden, die die Fälle als einzelne Zeilen enthalten und tabulatorgetrennt die Werte für alle Variablen in jeder Zeile. Wahlweise können in der ersten Zeile die Variablennamen enthalten sein. Durch die spätere Möglichkeit, Kriterium und Prädiktoren aus der Menge der Variablen auszuwählen, sind keine weiteren Strukturanforderungen zu erfüllen. Wichtig ist, dass zulässige Werte für das Kriterium ausschließlich die Werte 0 und 1 sein dürfen, zulässige Werte für Prädiktoren die ganzzahligen Werte 0 bis 100[2]. Fehlende Werte dürfen nicht auftreten, hier ist unter Umständen eine Vorbearbeitung der Datensätze oder eine Bearbeitung in der Datentabelle notwendig.

Das Programm selbst liegt als *Executable* vor, als ausführbare Datei mit einer „*.exe-Endung“ und eigenem Icon. Die erforderlichen Windows-*Libraries* sind bei einem Windows-Rechner mit Standardinstallation entweder bereits installiert oder in die ausführbare Datei eingebunden.

C.1.4 Beschreibung der Oberfläche und Steuerung

Verzicht auf Standardeinstellungen und Direktauswahlen

Es ist eine Konvention in vielen Statistikpaketen, Standardwerte für Parameter und Prozeduren zu implementieren. Diese Werte fehlen in der entwickelten Software aus folgendem Grund: Standardwerte implizieren aus der Perspektive vieler Anwender das Vorliegen einer Original- oder Ursprungsmethode. Es entsteht der Eindruck, dass man sich bei Abweichung von den Empfehlungen für eine Alternativmethode entscheidet, sich bei Übernahme der Standardeinstellungen aber nicht entscheiden muss. Denn „... naïve users tend to use defaults of available procedures [...] and reproduce output indiscriminately. No wonder, many are the instances when output that has no bearing on the problem being investigated is reported and „interpreted““(Pedhazur & Schmelkin, 1991, S. 345). Jede Setzung von Parametern stellt aber eine Entscheidung dar, die vom Anwender zu verantworten ist. Die Abgabe dieser Verantwortung an die Software aber auch das Aufgreifen der häufig fälschlich getroffene Annahme der Überlegenheit bestimmter Einstellungen bei allen Problemen soll durch den Verzicht auf Standardwerte verhindert werden. Ein Vergleich der Ergebnisse von „Standardverfahren“ in verschiedenen Softwarepaketen zeigt oft große und relevante Diskrepanzen aufgrund anderer Parametersetzungen, ohne dass *Bugs* in einem der Programme vorhanden sein müssen. Dementsprechend erfordert die Konstruktion von Algorithmen die

[2] Für GIER[1.0].simple_heuristics sind auch andere Prädiktorvariablenwerte zugelassen.

Kenntnis über die Implikationen der Parametersetzungen. Die damit im Design verankerte Notwendigkeit der Aufmerksamkeit ist in vollem Maße gewünscht. Die mit der Verwendung einer exorbitant hohen Zahl verschiedener Algorithmen in der „Explorationsphase“ einer Studie kann nicht im Interesse des Forschenden liegen, wenn die eingesetzten Methoden nicht zumindest voll verantwortet werden können:

> It has almost become the norm to analyze data every which way, mining it, bombarding it with every available analysis, regardless of whether or not it is appropriate and/or relevant. (Pedhazur & Schmelkin, 1991, S. 345)

Grobe Fenster- und Menüstruktur

Eine Darstellung von Details der programminternen Verknüpfungen oder der Feinheiten der Interfacegestaltung würde an dieser Stelle zu viel Raum einnehmen. Statt dessen soll hier für eine feinere Analyse der Umgang mit der Software selbst empfohlen werden. Hier soll lediglich die Grobstruktur des Programms beschrieben werden. Die Anwendung besteht aus einem Hauptfenster und mehreren Unterfenstern, in denen Einstellungen vorgenommen werden, oder Ergebnisse und Diagramme ausgegeben werden, einigen Modalfenstern zur Eingabe von Parametern und mehreren Dialogen zum Laden und Speichern.

Das Hauptfenster von GIER$_{[1.0]}$enthält die Datentabelle sowie Menüs und *Buttons* zur Steuerung des Programms. Die Datentabelle ist als *Stringgrid*-Komponente realisiert, enthält Scrollleisten und ermöglicht das Editieren von Werten. Die erste Zeile ist für Variablennamen reserviert, die erste Spalte für Fallbezeichnungen. Die Buttonleiste enthält Buttons zum Laden und Speichern von Datensätzen, zur Auswahl von Variablen und Fällen im Datensatz, zur Erstellung und zum Ausführen von Script- und Meta-Scriptdateien, sowie zum Aufrufen von Zusatzmodulen. Im Hauptmenü sind alle Optionen anwählbar, die auch als *Button* repräsentiert sind, zudem können einige Diagrammtypen direkt angewählt werden, zudem sind Menüeinträge für Programmparameteränderungen und die Festlegung von Programmverzeichnissen vorhanden. Auch ein Fenster zur Erstellung von Grafikparameterdateien wird über das Menü gesteuert.

Zur Erstellung von Prototypen aller Algorithmenformen wird das Prototypenfenster eingeblendet, in dem über *Buttons* der gewünschte Algorithmentyp ausgewählt wird. Die Eingaben für Parameteroptionen werden über eine *TPageControl*-Komponente auf mehrere Seiten entsprechend den Komponenten des Algorithmus aufgeteilt und werden durch *Radio-Buttons*, *Edit*-Felder, *Buttons*, *Check-Boxes* und *Stringgrids* vorgenommen. Aus diesem Fenster heraus können die Einstellungen dann in Prototypen-Dateien gespeichert werden. Scripte und Meta-Scripte werden in ähnlicher Form in einem eigenen Fenster zusammengestellt und können als Script-Dateien bzw. Meta-Script-Dateien gespeichert werden. Nach dem Starten von Scripten und Metascripten wird ein Fortschrittsfenster eingeblendet, dass Angaben zum Fortschreiten der Berechnungen bietet und je nach Algorithmus auch Visualisierungen von Zwischenständen oder Entwicklungen. Auf ein eigenes Ausgabefenster wurde verzichtet, da alle Ergebnisse als Dateien in Auswertungsverzeichnis gespeichert werden, was zum Punkt der Verzeichnisstruktur überleitet.

Verzeichnisstruktur

Die Anwendung besteht aus einer ausführbaren Datei, während Parameterwerte für das Gesamtprogramm in einer *.ini*-Datei abgespeichert werden. Neben dem Anwendungsverzeichnis gibt es ein Datenverzeichnis, ein Prototypenverzeichnis, ein Scriptverzeichnis und ein Auswertungsverzeichnis (siehe Abbildung C.1).

Der Inhalt der Verzeichnisse entspricht der Namensgebung, wobei das Auswertungsverzeichnis differenzierter betrachtet werden muss: Für jedes Script wird ein Unterverzeichnis angelegt, das mit dem Namen des Scripts benannt wird, gegebenenfalls um einen Zähler ergänzt. Jedes Unterverzeichnis enthält neben einer Kopie der verwendeten Script-Datei und einer Datei, die den Fortschritt bei der Abarbeitung des Scripts speichert, seinerseits eine Reihe von Unterverzeichnissen(s. Abbildung C.2):). In einem ersten Unterverzeichnis werden Kopien der verwendeten Prototypen gespeichert, in einem zweiten Kopien der verwendeten Originaldatensätze. Ein drittes Unterverzeichnis enthält Diagrammausgaben auf der Ebene des Scripts, etwa Variablendiagramme oder Scatterplots. Ein viertes Verzeichnis enthält Datensätze mit Klassifikationsergebnissen ein fünftes Verzeichnis die Protokolldateien für das Script. In einem weiteren Verzeichnis sind dann schließlich Angaben über die zum Script gehörenden einzelnen Tasks angelegt. Diese Verzeichnisse enthalten Taskdateien und besitzen ebenfalls eine Reihe von Unterverzeichnissen (vgl.Abbildung C.3):

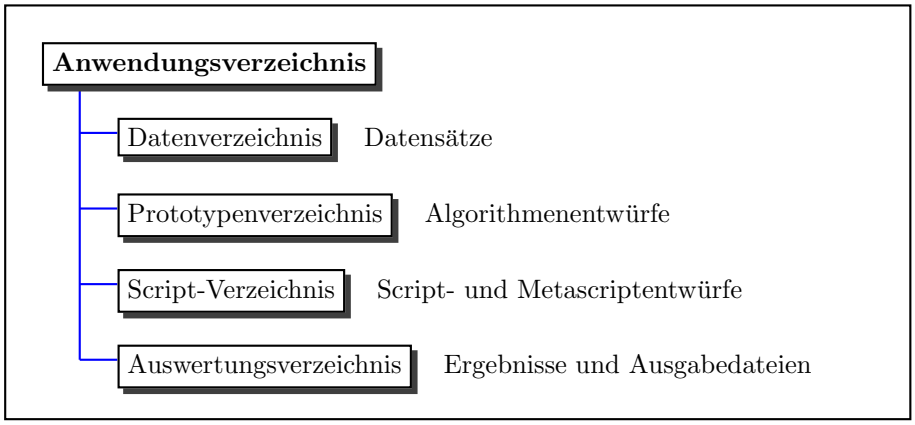


Abbildung C.1: Verzeichnisstruktur von GIER$_{[1.0]}$

Ein Diagrammordner enthält Diagramme auf Taskebene, etwa Regelvisualisierungen, Entscheidungsbäume oder Verlaufsdiagramme für Algorithmen, ein Klassifikationsordner enthält klassifizierte Datensätze auf Taskebene, der Kreuzvalidierungsordner dementsprechend kreuzvalidierte Datensätze auf Taskebene. Schließlich werden im Protokollordner die Protokolldateien für die einzelnen Subtasks gespeichert, sowie Syntax-Dateien für die Klassifikation mit SPSS, Regelprotokolle oder TYPAG-Ergebnisse. Diese doch recht verzweigte Dateistruktur hat die Aufgabe, die entstehenden Ausgabedokumente übersichtlich zu sortieren, da der Datei-*Explorer* ein gesondertes Ausgabemodul innerhalb der Software ersetzen muss. Da die Dateien in der gewählten Form aber auch leichter zu exportieren und in anderen Programmen verwendbar bleiben, wurde die Entscheidung zugunsten der Ausgabeverteilung auf Ordner getroffen. In der Verzeichnisstruktur werden Prototypendateien, Scripte, Datensätze und Ergebnisse getrennt voneinander in eigenen Ordnern gespeichert, die für eine Auswertung relevanten Elemente werden aber alle in den Ausgabeordner kopiert, so dass eine saubere Dokumentation der Berechnungsdetails erfolgen kann.

C.2 Objektorientierte Programmierung von Algorithmen

C.2.1 Relevante Konzepte Objektorientierter Programmierung

Es stellt ein Problem dar, dass Computerprogramme in ihrer Funktionsweise von außen oft als kryptisch betrachtet werden:

> Often likened to black boxes, the inner workings of which are obscured from vision, hence, from comprehension, computers and their programs are invested with magical qualities, and infallibility. Faced with neatly printed output, it appears almost inconceivable to entertain the notion that it may contain errors. Complicating matters is the increasing use of computers to solve problems the answers to which cannot possibly be checked by humans. (Pedhazur & Schmelkin, 1991, S. 344)

Um die profane Natur von Programmen, aber auch ihren Nutzen für eine psychologische Methodik zu veranschaulichen, soll deshalb in diesem Abschnitt ein Überblick über relevante Aspekte der objektorientierten Programmierung und Modellierung gegeben werden. Dabei soll die Nützlichkeit für die Umsetzung von statistischen Methoden verdeutlicht werden, auf eine umfassende Darstellung der ontologischen und erkenntnistheoretischen Aspekte wird hier verzichtet (siehe dazu aber Thanh Hai, 1999). Die Programmiersprache Delphi 6.0, die zur Programmierung von GIER$_{[1.0]}$zum Einsatz kam, zwingt nahezu zum Umgang mit Objekten (Cantù, 2001), da bereits die Standardkomponenten der Fenstergestaltung als Objekte verwaltet werden.

Klassen und Objekte

Die erste objektorientierte Programmiersprache wurde war Smalltalk-80, sie wurde zwischen 1970 und 1980 in Kalifornien entwickelt und übernahm das Klassenkonzept von der Sprache SIMULA 67 (Balzert, 2001). Ein Objekt ist in diesem Kontext am besten zu verstehen als ein „Gegenstand des Interesses,

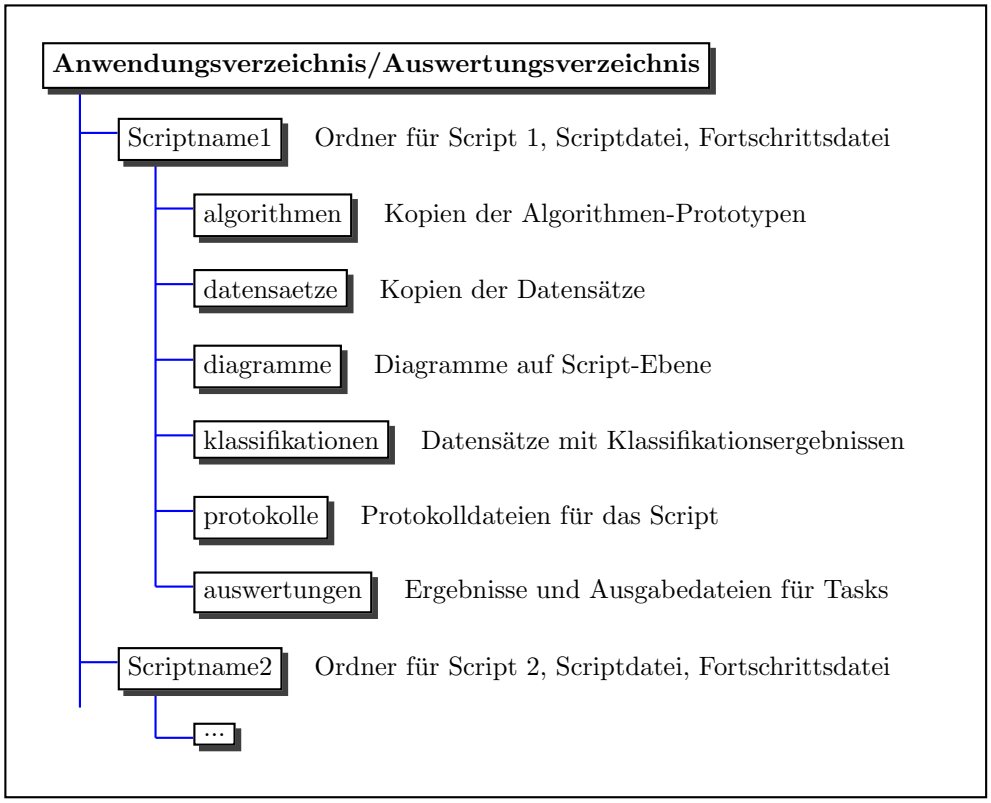


Abbildung C.2: Unterverzeichnisstruktur für Auswertungen

insbesondere einer Beobachtung, Untersuchung oder Messung“ (Balzert, 2001, S.156). In der Programmierung wird mit Objekten zumeist auf eine konkrete Struktur im Speicher verwiesen (vgl. Cantù, 2001). Objekte haben Eigenschaften (oder Attribute) und die Summe ihrer aktuellen Eigenschaften kann als Zustand oder auch Repräsentation beschrieben werden. Ebenso entspricht die Menge der dem Objekt möglichen Operationen (oder Methoden) seinem Verhalten. Gleichzeitig ist jedes Objekt eine Instanz einer Objektklasse. Beispielsweise ist auch der vorliegende Text eine Instanz der Klasse „Dissertationen“. Klassen sind notwendigerweise Abstraktionen und natürlich kann es auch Klassen von Klassen geben, allerdings nicht Klassen beliebiger Klassen (z.B. nicht die Klasse aller Klassen, die sich nicht selbst enthalten), als Objekt wird aber etwas bezeichnet, dass eine konkrete individualisierte Entität darstellt. Das Objekt kann mit der Beschreibungsstruktur der Klasse beschrieben werden, denn für jedes Attribut, das Bestandteil der Klasse ist, hat das Objekt einen korrespondierenden Attributwert. Enthält die Klasse der Dissertationen bspw. das Attribut „Sprache“, so hat die vorliegende Instanz den Attributwert „Deutsch“ bzw. „Sprache=Deutsch“. Damit sind die Begriffe „Objekt“, „Instanz“, „Klasseninstanz“ und „Exemplar“ im Kontext der objektorientierten Programmierung (OOP) Synonyme. Eine Klasse kann definiert werden als „eine Struktur von Feldern, Methoden und Eigenschaften“ (Borland GmbH, 2001, S.7-1). Diese Bestandteile heißen Komponenten oder Elemente einer Klasse. Dabei sind Felder die zu einem Objekt gehörenden Variablen, Methoden die zu einem Objekt gehörenden Prozeduren oder Funktionen. Man könnte Klassen auch allgemeiner fassen als „Menge von Objekten mit denselben Eigenschaften (Attributen), demselben Verhalten (Operationen) und denselben Beziehungen“ (Balzert, 2001, S. 161). Dabei ist es allerdings wichtig, dass mit „denselben Eigenschaften“ nicht „dieselben Eigenschaftswerte“ gemeint sind. Klassen haben Mechanismen, um Objekte als Instanz von sich selbst zu erzeugen.

Properties sind eine spezielle Art von Variablen, die über Zugriffsangaben bestimmen, wie enthaltene Daten modifiziert oder ausgelesen werden können. Der Klassentyp bestimmt damit die Struktur, die einem Objekt im Speicher zugewiesen werden muss, deshalb lässt sich eine Klassentypvariabe als Zeiger bzw. *Pointer* auf das Objekt verstehen. Methoden sind allen Instanzen einer Klasse gemeinsam, aber jedes Objekt verfügt über eine eigene Variante der Variablenstruktur. Objekte müssen anders als Standardvariablen kreiert und zerstört werden, um zum einen eingesetzt werden zu können und zum anderen ihren Speicherbedarf wieder zur Verfügung stellen zu können. Eine wesentliche Eigenschaft von Objekten in der OOP ist die Einkapselung (*Encapsulation*) oder Geheimhaltung von Daten. Daten können als

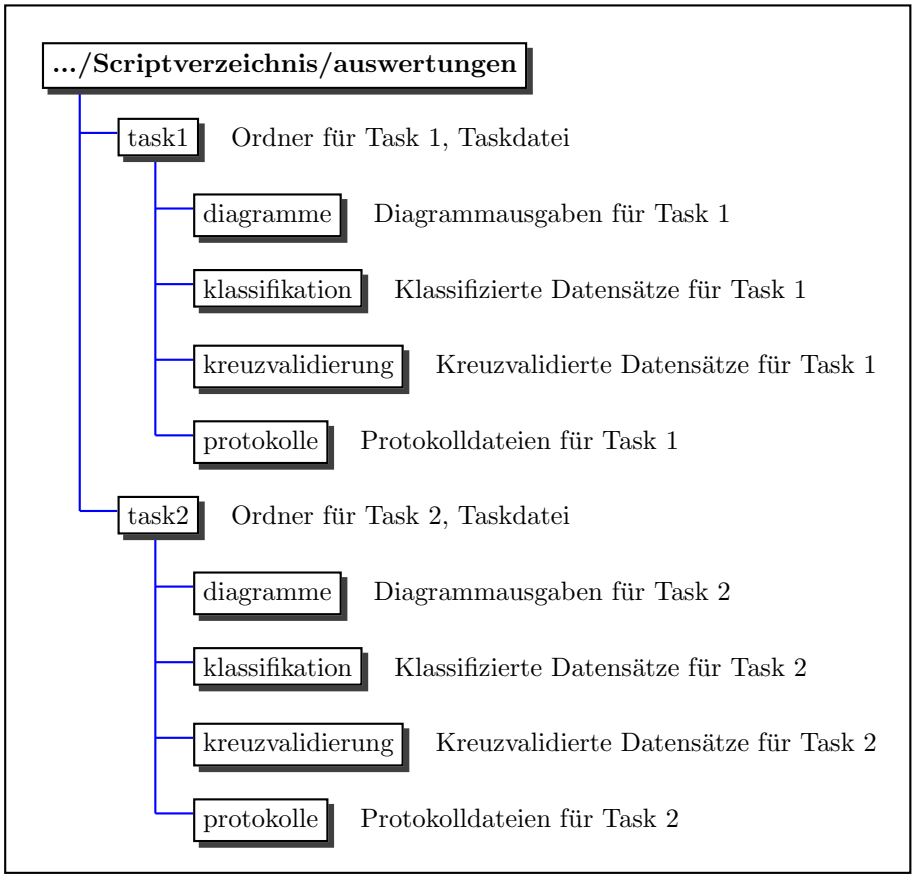


Abbildung C.3: Unterverzeichnisstruktur für Scripte

private, *protected* oder *public* deklariert werden, wobei die ersten beiden Zugriffsvarianten den Zugriff auf bestimmte Programmteile einschränken: Als *private* deklarierte Daten können nur in der *Unit* verwendet werden, in der sie deklariert wurden, als *protected* deklarierte Daten nur durch die Klasse und ihre Unterklassen (s.u.). Die angesprochenen *Properties* stellen die schärfste Einschränkung der Sichtbarkeit dar, da der Datenzugriff nur noch über klasseneigene Methoden erfolgen kann.

Zur Visualisierung von Objekten stellt die *Unified Modeling Language* (vgl. Amber, 2005; Roussev, 2006) Diagrammformen zur Verfügung. Abbildung C.4 zeigt eine Visualisierung für eine Beispielklasse und ein Beispielobjekt, das eine Instanz dieser Klasse darstellt. Die Visualisierung besteht aus einem dreigeteilten bzw. zweigeteilten Rechteck. In der ersten Zeile steht bei Klassen der Name, bei Objekten der Name gefolgt von einem Doppelpunkt und der Klasse, deren Instanz das Objekt ist. Im zweiten Feld werden die Attribute und *Properties* der Klasse aufgeführt, bei Objekten werden zusätzlich die Wertbelegungen der Attribute und *Properties* angegeben. Zwei unabhängige Objekte können dabei im Rahmen der zulässigen Wertzuschreibungen vollkommen unterschiedliche Wertbelegungen aufweisen, können in ihren Belegungen aber auch gleich sein. Im dritten Feld sind schließlich die Operationen der Klasse auf-

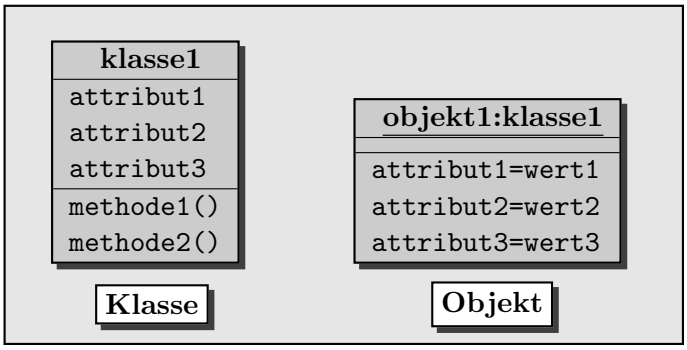


Abbildung C.4: UML-Darstellung von Klassen und Objekten bzw. Klasseninstanzen

geführt. In Klammern hinter dem Namen werden die Parameter aufgeführt, die zum Aufruf der Methode notwendig sind. Da Operationen auf der Ebene der Klasse festgelegt sind, werden sich Objekte nicht hinsichtlich ihrer verfügbaren Operationen oder Methoden unterscheiden können. Deshalb ist es üblich, diese Information bei Objekten nicht aufzuführen. Zusätzlich kann auch die Sichtbarkeit der Attribute und Methoden angegeben werden, dabei wird auf der Sichtbarkeitsstufe *protected* ein „#“, auf der Stufe *private* ein „-“ und auf der Stufe *public* ein „+“ vor den Namen des Attributs oder der Methode gestellt. Der Übersichtlichkeit halber wird hier auf diese Ergänzung verzichtet.

Aggregation und Vererbung

Die Modellierungssprache ermöglicht neben der Darstellung individueller Klassen und Objekte auch eine Veranschaulichung der Beziehungen zwischen Klassen. Von besonderer Bedeutung sind im vorliegenden Kontext die Konzepte der *Aggregation* und der *Vererbung*. Als Aggregation wird eine Beziehung bezeich-

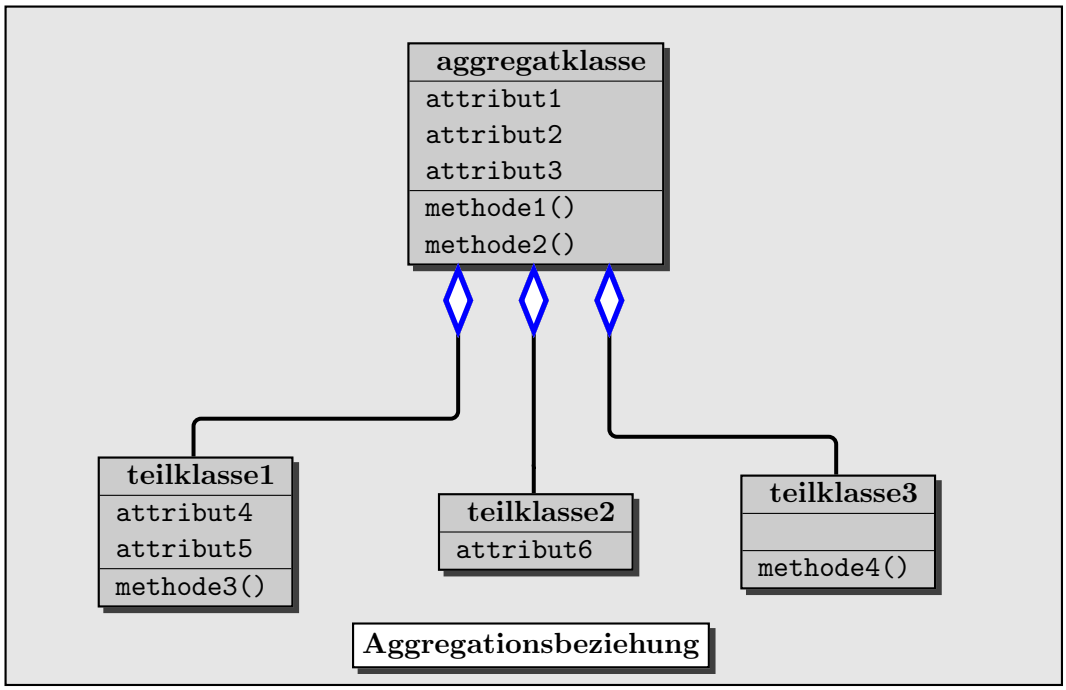


Abbildung C.5: UML-Darstellung einer Aggregationsbeziehung: Die Klasse „aggregatklasse“ ist im Beispiel die Oberklasse, die Klassen „teilklasse1“, „teilklasse2“ und „teilklasse3“ sind Unterklassen dieser Klasse

net, wenn eine „ist Teil von“ Beziehung zwischen Objekten besteht. Diese Beziehung ist asymmetrisch und wird durch einen gerichteten azyklischen Graphen visualisiert (Balzert, 2001). Dabei geht die Verbindung von der Teilklasse aus und und endet in einem Parallelogramm an der Aggregatklasse. Eine stärkere Form der Aggregation, die Komposition, wird von (Balzert, 2001) durch folgende zusätzliche Merkmale definiert:

- Ein Objekt einer Teilklasse darf nur zu einem Objekt der Aggregatklasse gehören, d.h. Objekte werden nicht geteilt.
- Die dynamische Semantik des Ganzen gilt ebenso für die Teile. Es kommt nicht vor, dass selektiv Teile kopiert werden.
- Eine Löschung des Ganzen führt automatisch zur Löschung seiner Teile.

Abbildung C.5 zeigt ein Beispiel für eine Aggregationsbeziehung.

Eine Vererbungsbeziehung besteht zwischen Klassen dann, wenn eine spezialisierte Klasse oder Unterklasse über die vollständige Repräsentation und das vollständige Verhalten einer allgemeinen Klasse oder Oberklasse verfügt. Die Vererbungsbeziehung ist asymmetrisch und transitiv, sie kann zudem nicht graduell festgestellt werden: sie besteht oder sie besteht nicht. Für eine Oberklasse sind die spezifischen

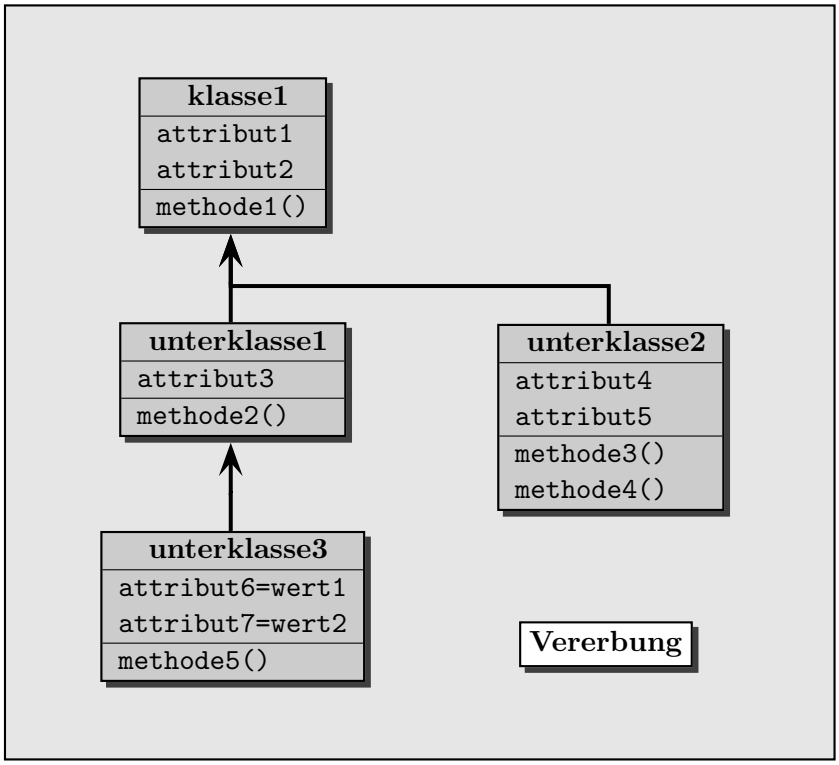


Abbildung C.6: UML-Darstellung einer Vererbungsbeziehung: Im Beispiel sind die Klassen „unterklasse1“ und „unterklasse2“ Unterklassen der Oberklasse „klasse1“, gleichzeitig ist „unterklasse1“ auch Oberklasse für die Unterklasse „unterklasse3“. Letztere erbt ebenfalls alle Eigenschaften und Methoden von „klasse1“

Attribute und Methoden ihrer Unterklassen nicht sichtbar, während alle nicht speziell geschützten Attribute und Methoden der Oberklasse für die Unterklassen sichtbar sind. Eine Vererbungsbeziehung wird in UML-Schreibweise als Pfeil repräsentiert, wobei der Pfeil von der erbenden Klasse auf die vererbende Klasse zeigt. Auf diese Art und Weise bildet das Vererbungsgefüge eine graphische Baumstruktur, wenn es eine allgemeine Klasse gibt („TObject“ in Delphi) oder eine graphische Waldstruktur, wenn es keine allgemeine Klasse gibt. In Abbildung C.6 ist ein zweistufiger Vererbungsbaum abgebildet. Es wird deutlich, dass „unterklasse1“ im Beispiel zugleich Unterklasse und Oberklasse ist, natürlich in Relation zu zwei unterschiedlichen Klassen. Die Konsequenzen dieser Struktur für aus den Klassen abgeleitete Objekte ist in Abbildung C.7 dargestellt. Entgegen der Konvention sind auch die für die Objekte anwendbaren Methoden aufgeführt. In der Grafik wird deutlich, dass die Vererbung von Eigenschaften nicht die Vererbung von Eigenschaftswerten instantiierter Objekte impliziert, sondern lediglich das Vorhandensein irgend eines Eigenschaftswertes abgeleitet werden kann. Zwischen den Objekten besteht ausdrücklich keine Vererbungsbeziehung, da diese Relation nur zwischen Klassen definiert ist.

Abstrakte Klassen, Polymorphismus und späte Bindung

Eine für die Programmierung ungemein nützliche Eigenschaft von OOP ist die Möglichkeit *später Bindungen* oder von *late coupling*, das mit dem Konzept des *Polymorphismus* verbunden ist. Ein Polymorphismus ermöglicht es, „den gleichen Namen für gleichartige Operationen zu verwenden, die auf Objekten verschiedener Klassen auszuführen sind“ (Balzert, 2001, S. 829). Der Nutzen einer solchen Möglichkeit ergibt sich aus der Konsequenz, dass ein Aufruf der gleichen Methode unterschiedliche Konsequenzen hat, je nachdem welcher Klasse ein Objekt angehört, ohne dass die Klassenzugehörigkeit des Objekts geprüft werden muss. Damit die Methode aber in geeigneter Weise aufgerufen werden kann, sind *abstrakte Klassen* und *abstrakte Methoden* von entscheidender Bedeutung. Eine abstrakte Klasse bezeichnet eine Klasse, die keine Objekte instantiiert im Gegensatz zu Klassen, die von der abstrakten Klasse abgeleitet wurden. Abstrakte Klassen besitzen im Allgemeinen eine oder mehrere abstrakte Methoden, d.h. Methoden die nur aus einer Deklaration ohne Implementierung bestehen. Abgeleitete Klassen hingegen können diese Deklaration „mit Inhalt füllen“, also die abstrakte Methode durch eine nicht-abstrakte Methode

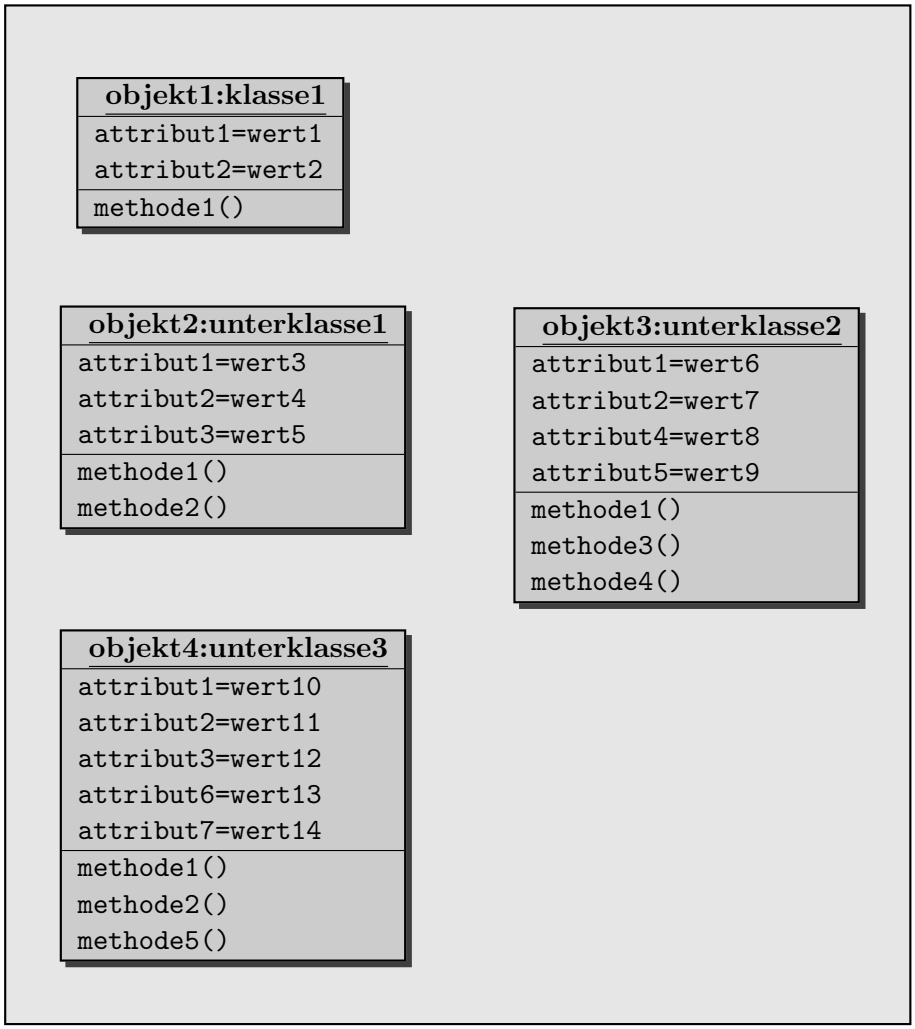


Abbildung C.7: Modifizierte UML-Darstellung von Objekten mit vererbten Attributen: Die Klassen stehen in der Vererbungsrelation aus Abbildung C.6

ersetzen. Da es möglich ist, jedes Objekt beim Zugriff wie ein allgemeineres Objekt zu behandeln, kann durch einen Aufruf der Methode von der Ebene der abstrakten Klasse aus, die Implementierung aller entsprechenden abgeleiteten Klassen aufgerufen werden, je nachdem welches Objekt den Platz besetzt, der mit dem Namen der abstrakten Klasse freigehalten wurde. Ist eine Operation in diesem Sinne nicht an eine Klasse gebunden, spricht man von einer polymorphen Operation oder von später Bindung (Balzert, 2001). Neben der Verwendung abstrakter Klassen ist es möglich, polymorphe Operationen durch das Überschreiben von Operationen der Oberklasse zu erreichen. In jedem Fall müssen Name und Parameter der abstrakten bzw. ursprünglichen Operation und der Operation der abgeleiteten Klasse identisch sein.

Ein Beispiel mit beiden Vorgehensweisen ist Abbildung C.8 zu entnehmen: Der Aufruf von „methode2“ der Oberklasse führt je nach Klassenzugehörigkeit des Objekts zum Aufruf einer unterschiedlich implementierten Methode. Gleichzeitig führt der Aufruf von „methode1“ bei Objekten der zweiten Unterklasse zu einer Ausführung einer anderen Implementierung als bei Objekten, die aus den anderen Unterklassen abgeleitet wurden. Die Nützlichkeit der vorgestellten OOP-Konzepte für die Entwicklung und Weiterentwicklung statistischer Methoden soll im folgenden Abschnitt erläutert werden.

C.2.2 Objektorientierte Programmierung und visuelle Methodenbaukästen

Zunächst einmal soll gezeigt werden, welche Vorteile eine konsequente Objektorientierung für die Entwicklung statistischer Algorithmen birgt, um an einem einfachen Beispiel aus der Software GIER$_{[1.0]}$das Wirken dieser Prinzipien zu veranschaulichen. Zuletzt soll eine visuelle Sprache vorgestellt werden, mit der die Struktur der Software übersichtlicher beschrieben werden kann.

Zunächst einmal beinhaltet OOP eine saubere Differenzierung der Aufgabenstruktur und eine Ordnung der Komponenten. Abbildung C.9 zeigt ein fiktives Beispiel für Hierarchiebeziehungen zwischen

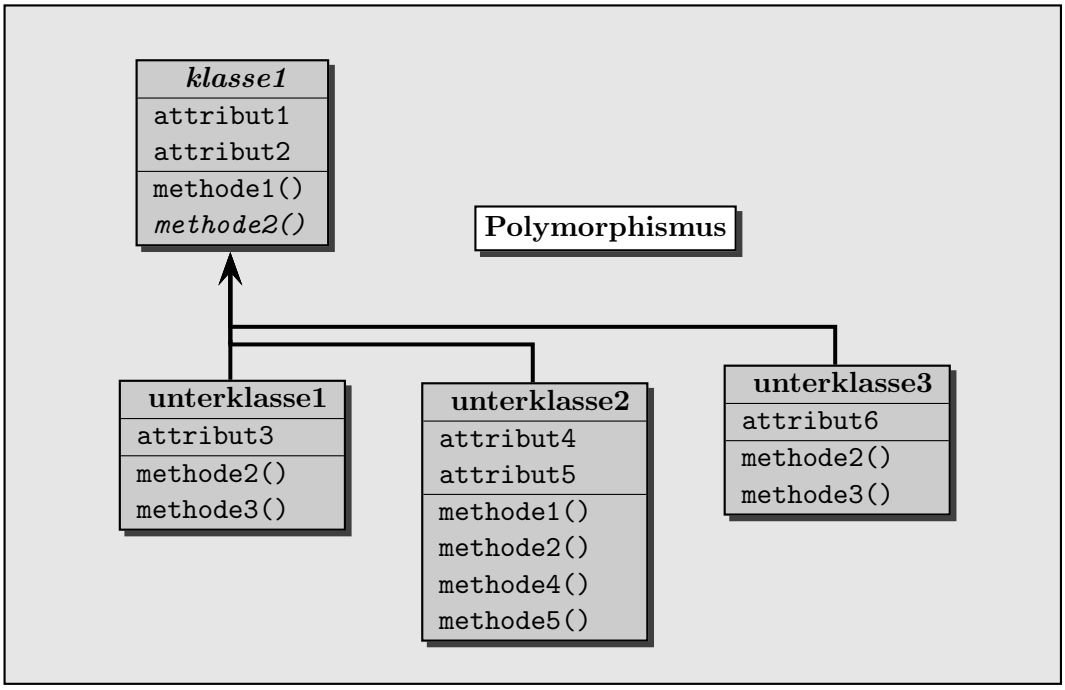


Abbildung C.8: Beispiel für polymorphe Methoden in UML-Darstellung: Die Unterklassen „unterklasse1“ und „unterklasse2“ implementieren die abstrakte Methode „methode2“ der Oberklasse. Die Unterklasse „unterklasse2“ überschreibt zusätzlich die Operation „methode1“ der Oberklasse.

Objekten, die verschiedene statistische Klassifikationsverfahren umsetzen sollen. In der abstrakten Klasse „t_klassifikationsverfahren“ wird die Grundstruktur der Operationen eines Klassifikationsverfahren festgelegt. Eine nicht gezeigte Klasse könnte die drei Routinen der Klasse in angegebener Reihenfolge aufrufen, unabhängig davon, welcher Unterklasse das Objekt angehört, das diese Routinen implementiert. Aufgrund von Benutzereingaben könnte ein Objekt der gewünschten Klasse instantiiert werden, so dass die Routine je nach Wunsch in diesem Fall eine Logistische Regression, einen CHAID-Baum oder einen CART-Baum als Modell zur Klassifikation verwenden würde. Auch wenn diese Konstruktion zunächst nach zusätzlichem Aufwand aussehen mag, so hat die Vorgehensweise einige Vorzüge:

- Das Ausmaß an Redundanz wird gesenkt. Wenn die Hauptsteuerungsroutine einmal geschrieben ist, kann sie für jedes Verfahren verwendet werden und muss nicht jedes Mal neu geschrieben werden. Redundanz ist eine der Hauptfehlerquellen bei einer Anpassung von Programmen an neue Funktionen oder Umgebungen.
- Jeder Verfahrensklasse stehen nur die Routinen zur Verfügung, die für das Verfahren relevant sind. Beim Programmieren wird die Möglichkeit der Anwendung unpassender Methoden eingeschränkt.
- Gibt es Methoden, die eine Gruppe von Verfahren teilt, so kann wie im Beispiel mit abstrakten Zwischenklassen gearbeitet werden, die Methoden implementieren, die für jede Unterklasse in gleicher Weise funktionieren (im Beispiel die Methode „zeige_baumstruktur()“), und wieder andere Methoden abstrakt deklarieren, um durch späte Bindung den gewünschten Code starten zu können (im Beispiel die Methode „berechne_baum()“)
- Bei einer Erweiterung um zusätzliche Verfahren steht bereits ein Gerüst zur Verfügung, in das die neuen Verfahren zeitsparend eingepasst werden können.

Der letztgenannte Punkt gilt natürlich nur, wenn die neuen Verfahren keine strukturell tiefgreifenden Veränderungen notwendig machen, die auch andere Verfahren betreffen. Beispielsweise kann eine grundlegende Veränderung der Fragestellung oder Funktion der Software, oder die Verwendung einer neuer Repräsentation der Daten ein bestehendes System aus den Angeln heben, weshalb in der Planungsphase die gewünschte Erweiterbarkeit kategorial miteinbezogen werden sollte. Ermöglichen Polymorphismen und Vererbung die beschriebenen gewünschten Eigenschaften auf der Ebene zwischen Verfahren, so kommt es

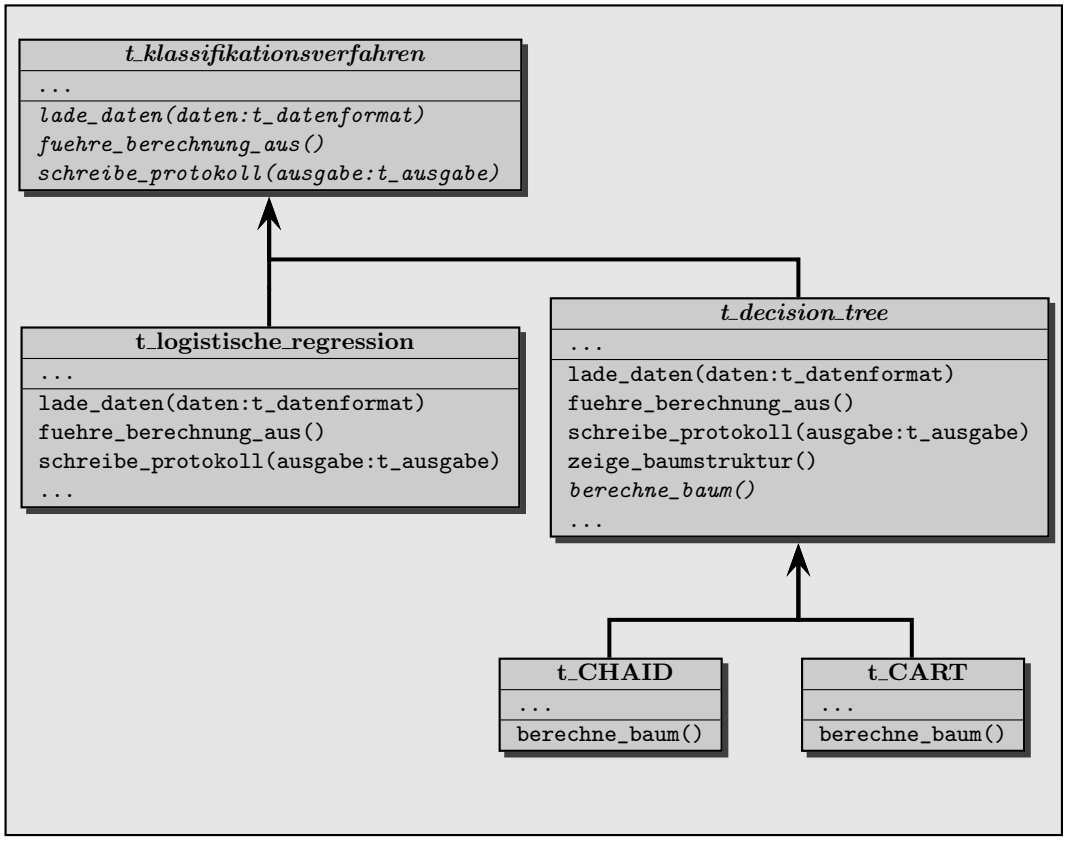


Abbildung C.9: UML-Darstellung eines fiktiven Softwareprojektes zur Klassifikation: Von der abstrakten Klasse „t_klassifikationsverfahren“ sind alle übrigen dargestellten Klassen abgeleitet. Die Klasse „t_decision_tree“ ist eine weitere abstrakte Klasse, deren Methoden von „t_CHAID“ und „t_CART“ implementiert werden.

zu einer noch deutlicheren Verbesserung de Programmierungsmöglichkeiten durch eine Verbindung mit der Aggregationsidee:

Für eine Vielzahl von Algorithmen ist es notwendig, Entscheidungen zwischen verschiedenen Optionen zu treffen. Dabei gibt es solche, die das Setzen von Parametern betreffen und andere, die prinzipiell eine Vorgehensweise verändern können. Wenn es bspw. nur darum ginge, den Exponenten in einer Formel zu verändern, so würde es ausreichen, den Attributwert eines Objekts zu verändern, bevor eine Berechnungsroutine aufgerufen wird. Ginge es hingegen darum, dass ein ganzer Term in der Gleichung verändert werden soll, wäre eine Lösung über Klassen denkbar. Selten ist es eine einzelne Entscheidung, die zu treffen ist. Die Idee, für jede Kombination aus Einstellungen eine eigene Klasse zu definieren stößt schnell an ihre Grenzen, wenn die Anzahl der notwendigen Klassen überdacht wird. Vier Entscheidungen mit jeweils fünf Alternativen würden schon eine Definition von 625 Klassen notwendig machen. Der Algorithmus $\text{GIER}_{[1.0]}$.ga käme bei diesem Vorgehen und einer Entscheidung für lediglich zwei mögliche Operatoren auf 29 568 000 notwendige Klassendefinitionen. Aus diesem Grund bietet es sich an, den Algorithmus als Komposition oder Aggregat zu verstehen, wie es in Abbildung C.10 demonstriert wird:

Die Algorithmenklasse „t_algorithmus1“ wird hier zusammengesetzt aus vier Bestandteilen, auf die bspw. von der Hauptroutine des Algorithmus aus Bezug genommen werden kann. Zusätzlich sind diese Hauptbestandteile jetzt aber abstrakte Klassen mit jeweils einer Reihe von abgeleiteten Klassen, die die variablen Bestandteile der Entscheidungsalternativen umsetzen. Dabei kann es auch die Variante geben, dass in der Zwischenklasse einige Operationen implementiert werden, die nur ergänzt werden durch die spezifischen Beiträge der Unterklassen. Die Zwischenklasse bliebe in dem Sinne eine abstrakte Klasse, als dass keine Objekte von ihr instantiiert würden, obwohl sie implementierte Methoden aufweist. Im Beispiel sind aus Gründen der Übersichtlichkeit nur für einen Teilschritt die alternativen Unterklassen aufgeführt worden.Als zu den Vorteilen dieser Vorgehensweise gehörend lassen sich benennen:

- Die formale Struktur macht eine saubere konzeptuelle Gliederung des Algorithmus notwendig.
- Eingabe- und Ausgabeformate des Algorithmus können in der Hauptklasse festgelegt werden, Einga-

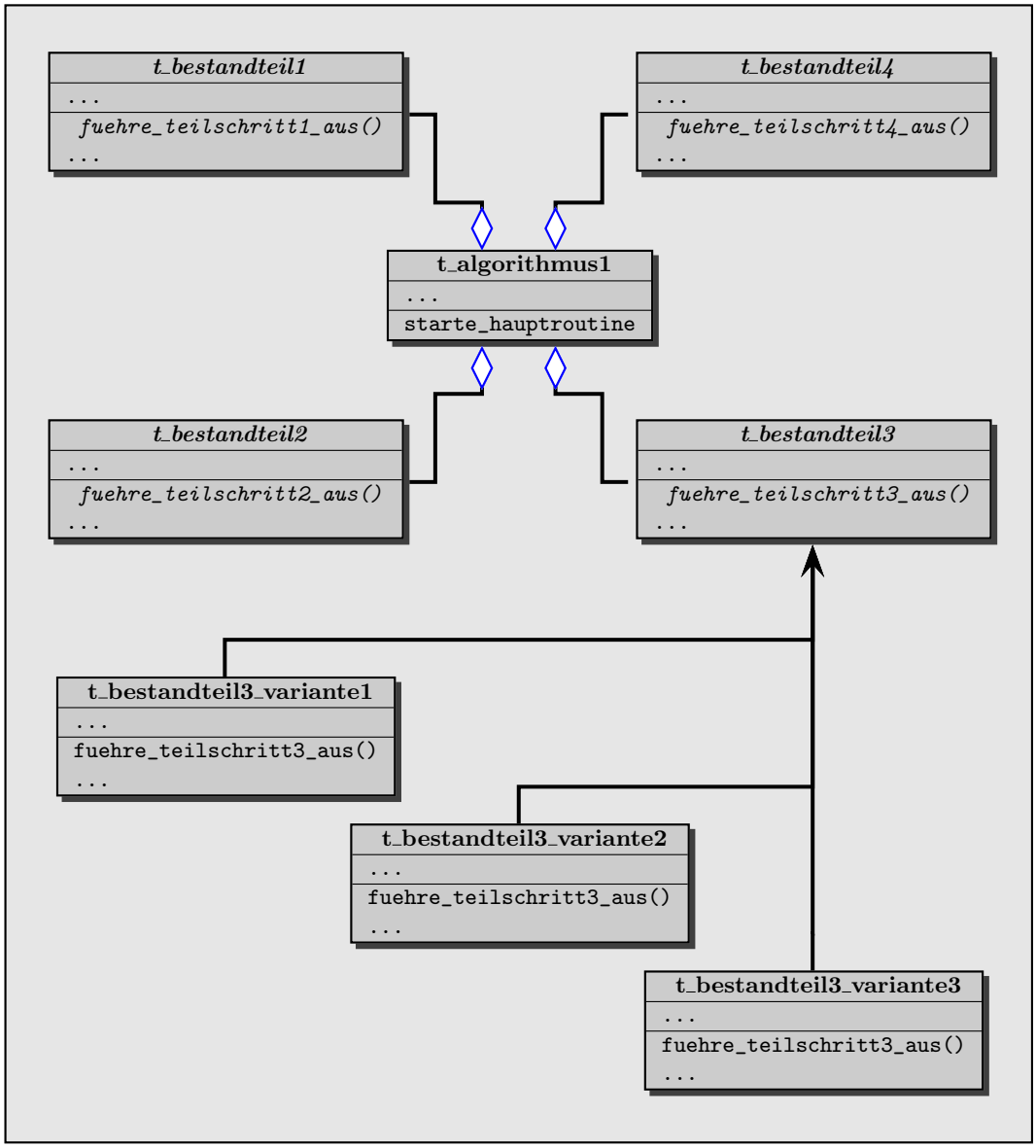


Abbildung C.10: Beispiel für Algorithmenklassen als Aggregatklassen: Die Klasse „t_algorithmus1“ setzt sich aus vier Bestandteilen zusammen, die wiederum aus absrakten Klassen bestehen. Für die Teilklasse „t_bestandteil 3“ sind drei abgeleitete Klassen mit Implementierung der abstrakten Methoden eingezeichnet.

be und Ausgabe der Teilschritte in den Zwischenklassen. Routinen der Datenübermittlung müssen aus diesem Grund oft nur einmal programmiert werden.

- Die Ergänzung um eine neue Alternative ist bei sauberer Schnittstellendefinition äußerst einfach, da sich die Veränderungen nur auf den Teilschritt beziehen müssen. In diesem Fall ist die Vertrautheit des Programmierers mit dem restlichen Code nicht einmal notwendig.
- Eine Erweiterung um neuen Code lässt den ursprünglichen Teil des Codes intakt. Fehler in anderen Teilen können dem Projekt lediglich durch unerwartete Konsequenzen der neuen Teile hinzugefügt werden.
- Der gleiche Code kann eine Vielzahl möglicher Varianten des Algorithmus realisieren, was eine notwendige Voraussetzung etwa für Methodenvergleiche oder Überprüfungen von Entscheidungskonsequenzen darstellt.
- Die formale Struktur bleibt auch über längere Zeiträume hinweg eine Hilfe zum Verständnis von Abläufen und Bezügen.

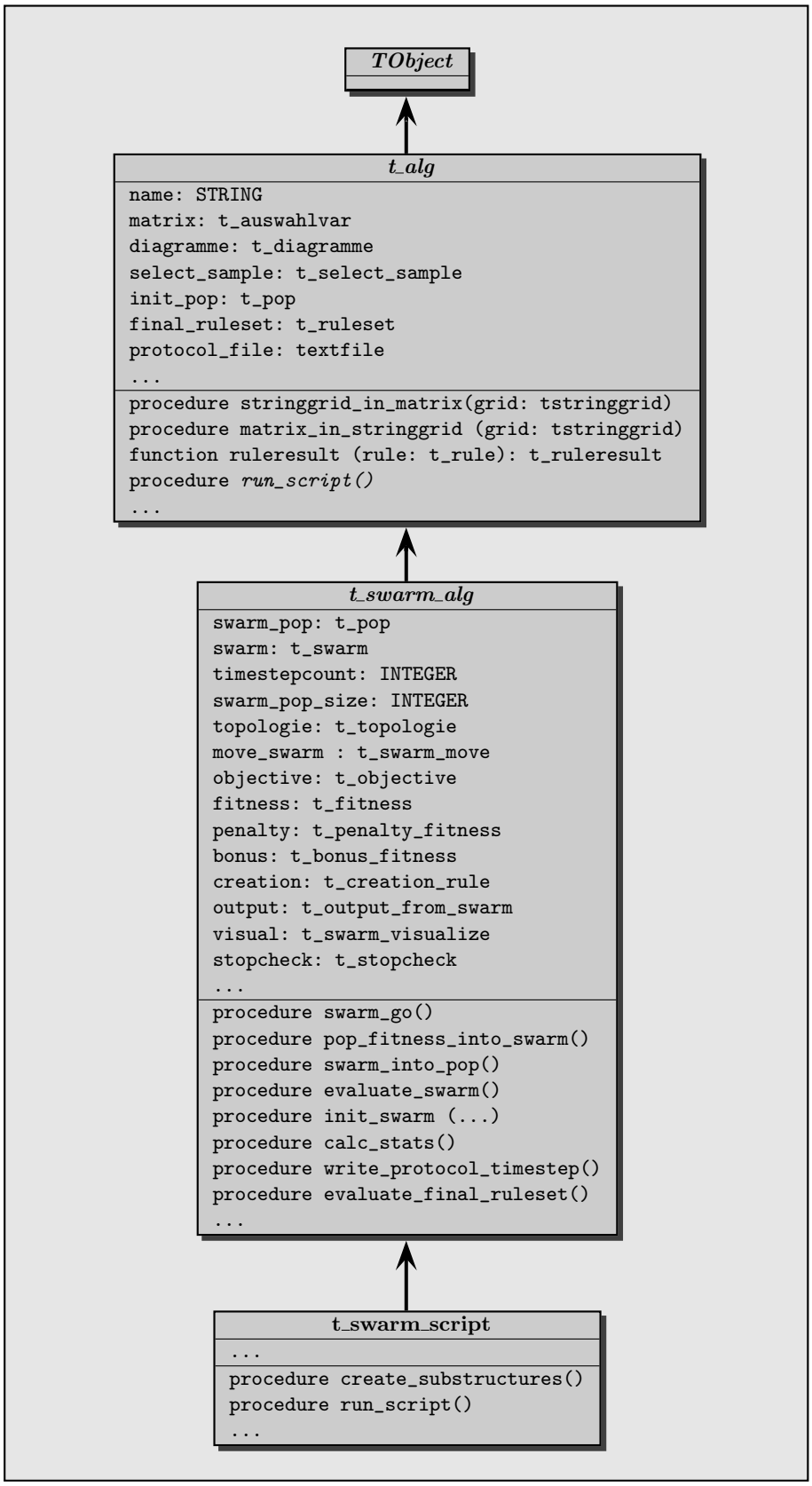


Abbildung C.11: UML-Darstellung von ausgewählten Aspekten einer Klasse aus GIER$_{[1.0]}$: In der Klasse „t_swarm_script“ wird die abstrakte Methode in „*run_script()*“ überschrieben. Einige Elemente wurden für die Darstellung umbenannt.

Der direkte Bezug zur vorliegenden Arbeit soll an einem Beispiel aus der entwickelten Software verdeutlicht werden. Ein Auszug aus den Eigenschaften und Relationen von drei der zahlreichen Klassen in GIER$_{[1.0]}$ist in Abbildung C.11 wiedergegeben. Die Klasse „TObject“ ist die allgemeine Objektklasse in Delphi. Von der Klasse „T_alg“ sind alle verwendeten Algorithmen abgeleitet worden. Sie enthält bspw. Methoden zum Einlesen von Daten und zur Berechnung von Statistiken über Regelmengen. Zum anderen wird die Methode „run_script“ von allen scriptfähigen Unterklassen überschrieben. In „t_swarm_alg“ gibt es zunächst Felder für Schwarm und Schwarmpopulation, die Nummer des Zeitschritts oder die Größe des Schwarms. Es folgt eine Liste mit Komponenten, die aus Objekten bestehen, die die Teilschritte des Algorithmus definieren. Der Schwarmalgorithmus setzt sich also aus mehreren Teilmodulen zusammen, die in der Routine „swarm_go()“ angesteuert werden können. Die Klasse „t_swarm_script“ dient der Steuerung durch Scriptdateien.

Auf die Bedeutung der Aggregation für „t_swarm“ wurde bereits hingewiesen. Polymorphismen werden im Diagramm an zwei Stellen angesprochen: Die Klasse „t_alg“ ist Oberklasse jedes Algorithmus, so dass die Prozedur „run_script()“ jeden dieser Algorithmen aufrufen kann. Nicht sichtbar im Diagramm, bestehen die benannten Modul-Komponenten von „t_swarm“ aus abstrakten Klassen, so dass eine späte Klassenbindung ermöglicht wird und verschiedene Varianten der Module durch „swarm_go“ aufgerufen werden können.

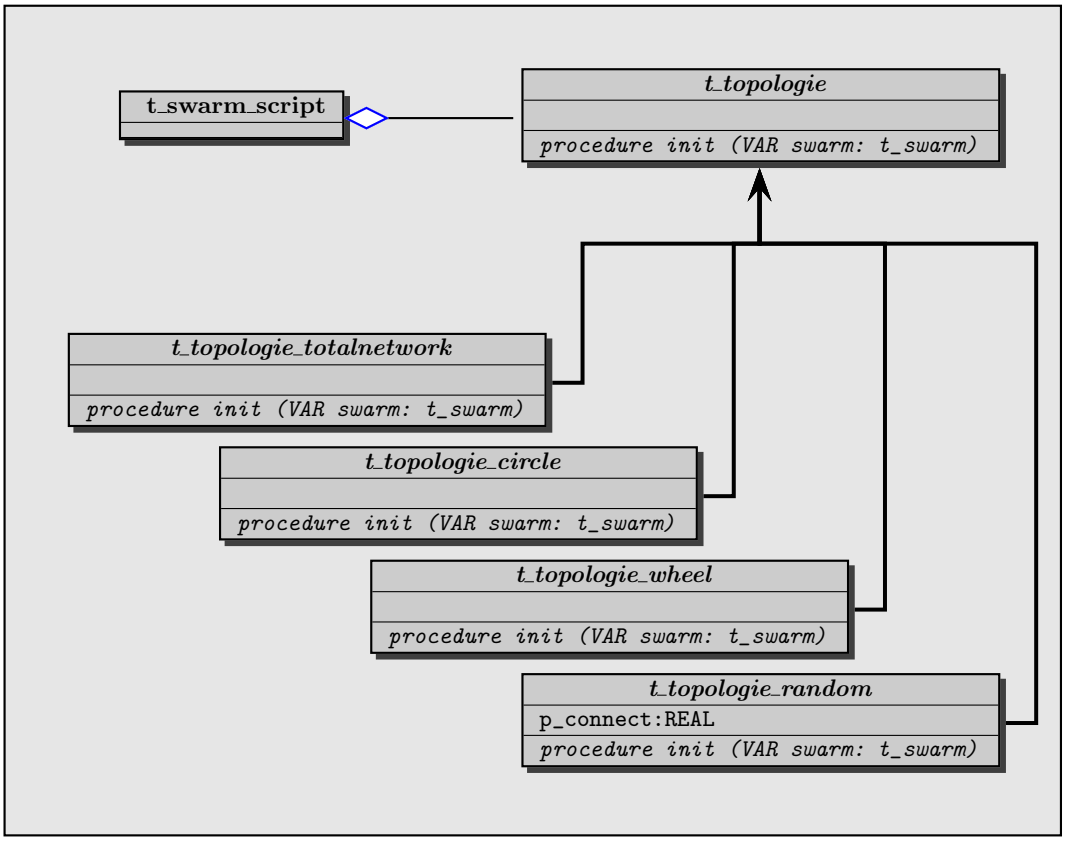


Abbildung C.12: UML-Darstellung eines Moduls von t_swarm: für die abstrakte Klasse „t_topologie“ sind alle konkreten abgeleiteten Klassen aufgeführt.

An dieser Stelle soll die Beziehung zwischen Objektstruktur und einer Bildsprache verdeutlicht werden, die zur Darstellung der Algorithmenklassen im Folgenden zum Einsatz kommen soll. Aus diesem Grund wird hier zunächst ein Teil des UML-Klassendiagramms abgebildet, das die Beziehung zwischen der Klasse „t_swarm“, der abstrakten Klasse „t_topologie“ und ihren konkreten Unterklassen veranschaulicht (s. Abbildung C.12. Es wird gezeigt, wie aus der abstrakten Komponentenklasse „t_topologie“ vier konkrete Unterklassen abgeleitet werden, die den Teilschritt unterschiedlich realisieren. Drei der Unterklassen haben keine Attribute, die vierte hat ein Attribut für die Verbindungswahrscheinlichkeit, das im Sinne eines Parameters fungiert. Dieses einfache Beispiel soll stellvertretend für andere Beispiele genügen.

Zur Verbesserung der Übersichtlichkeit wurde für GIER$_{[1.0]}$eine visuelle Sprache entworfen, die helfen soll, die Struktur der Klassenhierarchie und der Algorithmen auch ohne UML-Diagramme nachzeichnen zu können. Diese Sprache soll im Folgenden als *Visueller Baukasten* (*VBk*) bezeichnet werden. Ein Beispiel

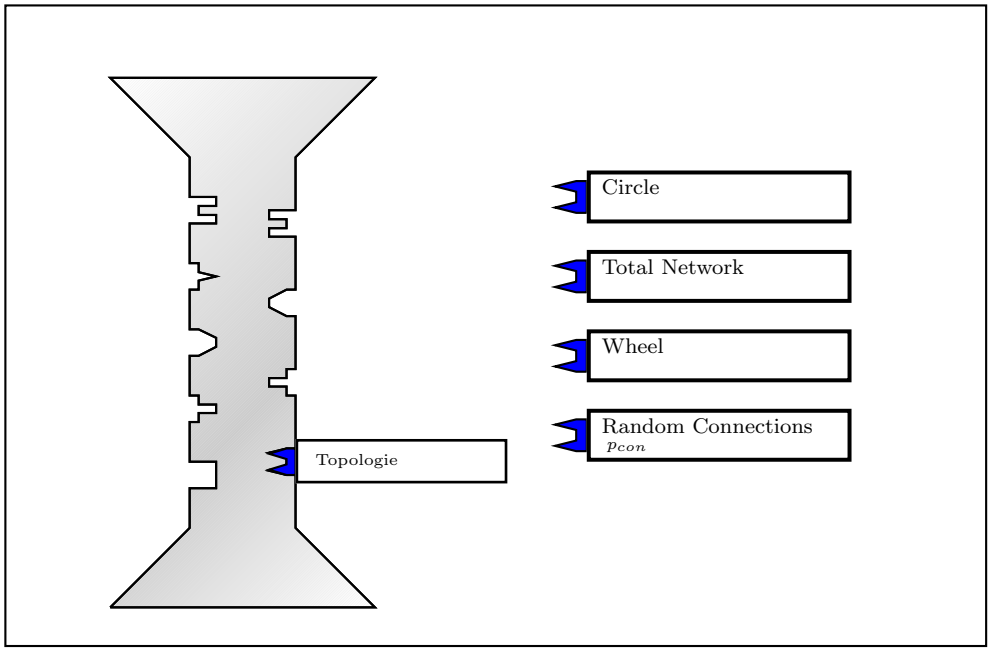


Abbildung C.13: Auszug der VBk-Darstellung von Schwarmalgorithmen: In die Form, die die Schwarmstruktur repräsentiert, ist eine abstrakte Steckkarte für das Modul „Topologie“ eingesetzt worden. Die möglichen Varianten sind ebenfalls als Steckkarten repräsentiert. Die Darstellung ist in ihrer Bedeutung vergleichbar mit Abbildung C.12

ist in Abbildung C.13 dargestellt. Dieser Baukasten besteht aus drei verschiedenen visuellen Objekttypen:

1. Algorithmenklassen werden als geometrische Figuren, mit mehreren Andockstellen bzw. Steckplätzen visualisiert.

2. Abstrakte Komponentenklassen werden als Steckkarten mit dem Namen der Komponente in die Figur eingefügt .

3. Konkrete Klassen werden als Steckkarten repräsentiert, deren Kontaktstelle der entsprechenden Andockstelle an der Figur entspricht. Dabei ist die Richtung der Kontaktstelle (links oder rechts) von Bedeutung.

Durch individuelle Formen der Kontaktstelle wird eine eindeutige Zuordnung von Komponenten zu Algorithmen ermöglicht. Dabei ist zum einen die Mehrfachverwendung von Klassen für mehrere Algorithmen visuell darstellbar über die Identität der Kontaktstelle, zum anderen kann Polymorphie repräsentiert werden durch das Vorhandensein mehrerer Komponenten mit identischer Kontaktstelle. Abgebildet ist in der VBk-Darstellung damit eine der neun Komponenten des Schwarmalgorithmus zusammen mit den vier Entscheidungsalternativen für die Form dieses Moduls und den Parametern, die in den Alternativen benötigt werden.

C.3 Darstellung der Algorithmensammlung in Gier$_{[1.0]}$

Im Folgenden sollen die in Gier$_{[1.0]}$implementierten Algorithmenklassen mit Hilfe der VBk-Sprache vorgestellt werden. Dabei werden die Algorithmen jeweils kurz umrissen, bevor im Detail auf ihre Komponentenklassen eingegangen wird. Es werden dann die abstrakten Komponentensteckkarten in allen Variationsmöglichkeiten aufgeführt.

C.3.1 Algorithmenübergreifende Eigenschaften und Parameter

Die Kodierung von Regelmengen

Im Ergebnis müssen für eine Regelmenge folgende Angaben verfügbar sein:

1. Die Anzahl der Regeln in der Menge.
2. Die Prädiktorintervalle für jede Regel.
3. Die Entscheidungen für jede Regel.

Für die Kodierung der Entscheidung wurde ein einzelnes Zeichen verwendet, das aber nicht binäre Werte annehmen kann, sondern die Werte $\{0, 1, 2\}$. Die ersten beiden Zeichen stehen dabei für eine Entscheidung für die Kategorien mit gleicher Ziffer. Die Ziffer 2 kodiert die Möglichkeit, sich entsprechend den empirischen Verhältnissen für die günstigere der beiden Kategorien zu entscheiden, damit wird ein Individuum lernfähig.

Die Anzahl der Regeln ergibt sich aus der Stringlänge und der Variablenanzahl, bzw. die Variablenanzahl ergibt sich aus Stringlänge und Regelzahl. Für die Intervalle selbst wurde jeweils eine Zeichenkette mit vierzehn Bits verwendet, von denen die ersten sieben die untere Grenze und die zweiten sieben die obere Grenze kodieren. Alternativ hätte das zweite Intervall zur Kodierung der (maximalen) Intervallgröße genutzt werden können, um das Problem ungültiger Wertkombinationen zu vermeiden, damit wäre es aber nicht möglich gewesen, die zweite Grenze konstant zu halten, wenn sich die erste durch eine Mutation verändert. Sieben Bits kodieren $2^7 = 128$ verschiedene Zustände. Da nur die Zahlen 0 bis 100 als Werte vorkommen können, werden die Werte, die einer binären Kodierung von 0 bis 13 entsprechen als Kodierungen für 0 interpretiert und Werte, die eine binären Kodierung von 115 bis 127 entsprechen als Kodierungen für 100 interpretiert, da alle ganzzahligen Intervalle kodiert werden sollen. Damit gibt es einen Bias bei einer Zufallsauswahl von Werten für die möglichen Extremwerte, der als gewünscht betrachtet wird. Damit besteht der Zeichenstring bzw. das Genom für die bereits verwendete Regelmenge mit den vier Regeln

linebreak

$$
\begin{aligned}
R_1 :&\quad (10 \leq X_{1,j} \leq 90) \wedge (40 \leq X_{2,j} \leq 70) \rightarrow P_j = 1\\
R_2 :&\quad (43 \leq X_{1,j} \leq 59) \wedge (67 \leq X_{2,j} \leq 89) \rightarrow P_j = 0\\
R_3 :&\quad (22 \leq X_{1,j} \leq 49) \wedge (16 \leq X_{2,j} \leq 52) \rightarrow P_j = 0\\
R_4 :&\quad (3 \leq X_{1,j} \leq 85) \wedge (4 \leq X_{2,j} \leq 98) \rightarrow P_j = 1.
\end{aligned}
$$

aus den vier Einzelstrings bzw. Genen:
„**100011001101000011011010100**“ („1“ für die Vorhersage, „**0001100**“ und „**1101000**“ für die ersten beiden Intervallgrenzen, „**0110110**“ und „**1010100**“ für das zweite Grenzenpaar), dementspechend auch:
„**001110011001001101000111001111**“, zudem
„**001001000111111001111010000010**“ und schließlich
„**100100011100011001001011110000**“ zusammen und ergibt sich damit als

$$
\begin{aligned}
&10001100110100001101101010100001110011001001101000111001110010010001111110011110100001\\
&\qquad 010010001110001100100101110000
\end{aligned}
\tag{C.1}
$$

Ebenso kann aus der Zeichenkette bei Kenntnis über Variablen- oder Regelzahl die Regelmenge rekonstuiert werden. Je nach Algorithmus werden Kodierungen von Einzelregeln oder Kodierungen von Regelmengen eingesetzt.

Variablenaktivierungs-Kodierung

Eine Variante der Kodierung benutzt einen zusätzlichen Genomteil, der als Aktivierungsgenom fungiert. Dieser Teil hat genauso viele Bits wie Variablen, wobei jedes Bit in der Reihenfolge der Prädiktoren eine 1 für eine berücksichtigte und eine 0 für eine nicht berücksichtigte Variable annimmt . Bei einem berücksichtigten Prädiktor wird der Grenzenteil des Genoms unverändert interpretiert, bei einem nicht berücksichtigten Prädiktor werden die Grenzen auf 0 und 100 gesetzt, so dass die entsprechenden Prädiktorenwerte kein Gewicht haben, da durch diese Prädiktoren keine Antezedenzbedingung falsifiziert werden kann.

Parameter für Regelmengen

Generell können für Regelmengen verschiedene Parameter festgelegt werden. Zu diesen gehören im Einzelnen:

- Regelzahl: Die Anzahl der Regeln kann bei einigen Algorithmen entweder direkt oder als Höchstgrenze festgelegt werden.
- Mindestfallzahl pro Regel: Es wird die Mindestzahl von Objekten festgelegt, für die die Antezedenzbedingung einer Regel erfüllt sein muss, damit die Regel als gültig betrachtet werden kann. Damit wird bei relativ zu dieser Zahl kleinen Trainingsstichproben auch indirekt die Höchstregelzahl festgelegt.
- Restregel: Es wird eine Regel mit immer erfüllter Antezedenzbedingung als letzte Regel aufgenommen, die als Vorhersage die Majoritätsklasse für in der Trainingsstichprobe bis zu dieser Regel nicht klassifizierte Objekte trifft.

Die Einstellungen für Regelzahl und Mindestfallzahl sollten die besonderen Eigenschaften des Problems berücksichtigen sowie die erwartete Stabilität von Generalisierungen. Werden Regeln auf der Grundlage weniger Beobachtungen gebildet, so wäre eine hohe Stabilität erforderlich, damit die Regeln sich bei der Vorhersage als sinnvoll erweisen können. Was bei der gewählten Art der Aufschlüsselung verloren geht, ist die Möglichkeit, bereits klassifizierte Fälle zur Bewertung der Stabilität folgender Regeln nutzen zu können. In diesem Sinne ist die Analyse in dieser Hinsicht strikt konditional.

Random Repair, Universal Repair und One-Scale Repair

Für den Umgang mit ungültigen Regeln werden im Algorithmenbaukasten zwei Hauptprinzipien für die Reparatur verwendet: Regeln werden ungültig durch die Unterschreitung der Mindestzahl von Objekten, die die Antezedenzbedingungen der Regel satisfizieren. Das ist gleichbedeutend mit der Feststellung, dass die Intervallgrenzen für die Stichprobe zu eng gesteckt sind. Bei dem Prinzip des *Universal Repair* werden die Intervallgrenzen so erweitert, dass solange alle Grenzen gleichzeitig um einen Punkt verschoben werden (also die Mindestgrenzen nach unten, die Höchstgrenzen nach oben), wenn dies möglich ist, bis genügend Objekte die so generalisierte Regelantezedenzbedingung erfüllen. Da der Algorithmus nur gestartet werden sollte, wenn noch genügend unklassifizierte Objekte in der Stichprobe verbleiben. Beim *Random Repair*-Mechanismus, dessen Zeitbedarf je nach Problemstruktur höher sein kann, werden nicht alle Grenzen verschoben, sondern lediglich eine zufällig ausgewählte Grenze in Richtung einer Intervallvergrößerung verschoben. Dieser Schritt wird ebenfalls so lange ausgeführt, bis genügend Objekte durch die Regel klassifiziert werden können. Die One-Scale-Repair-Option ist konzipiert für Fälle, in denen eine ungünstige Veränderung die Grenzen nur eines Prädiktors betroffen hat. Der Reihe nach wird ausgetestet, ob die Erweiterung des Intervalls für einzelne Variablen ausreicht, um das Ziel der Mindestabdeckung zu erfüllen. Nur wenn auf diesem Weg keine Lösung gefunden werden kann, wird eine der beiden anderen Reparaturmethoden eingesetzt.

Penalty und Bonus

Der Unterschied zwischen einer gültigen und einer ungültigen Lösung ist quantifizierbar, wenn als Maßstab die Abweichung von der geforderten Mindestklassifikationszahl herangezogen wird. Dementsprechend kann diese Abweichung als Bemessungsgrundlage für Belohnungen und Bestrafungen genutzt werden. Als mögliche Varianten kommen auch hier die Etablierung von Cut-Off-Werten in Frage, sowie Funktionen der Abweichungswerte. Dabei werden bei der Bemessung von Penalty-Werten Abweichungen nach unten berücksichtigt, bei der Bestimmung von Bonus-Werten Abweichungen nach oben über einen festgesetzten Wert hinaus. In der Folge sollen alle Algorithmen der $\text{GIER}_{[1.0]}$toolbox vorgestellt werden.

Die Snowboard-Erweiterung des $\text{Gier}_{[1.0]}$.downhill-Algorithmus

Eine Erweiterung des Downhill-Algorithmus widmet sich dem Problem lokaler Minima, in denen der Algorithmus zum Stillstand kommen kann. Jeder gierige Algorithmus trägt aufgrund seiner lokalen Optimierungsstrategie das Risiko, auf der Suche nach der optimalen Lösung abgelenkt zu werden, was auch

durch die Erweiterung nicht als Problem ausgeräumt werden kann. Was hier aber verändert wird, ist die Flexibilität des Algorithmus, weshalb die Metapher des Snowboardfahrers im Gegensatz zum Abfahrtsläufer gewählt wurde: Nachdem der GIER$_{[1.0]}$.downhill-Algorithmus eine Regel erstellt hat, wird diese Regel der Snowboard-Optimierung als Input gegeben. Diese besteht aus den gleichen Schritten wie der Downhill-Algorithmus, ermöglicht aber auch die Vergrößerung von Intervallen unter Angabe einer Mindestschrittgröße. Auch dieser Algorithmus nimmt Veränderungen nur bei Verbesserungen der Klassifikationsleistung vor, so dass die Schrittzahl bei endlichen Stichproben endlich bleibt, eine ringförmige Veränderungsabfolge ist nicht möglich, kein Zustand wird mehr als einmal erreicht. Durch die Erweiterung gibt es pro Variable bis zu vier mögliche Intervallveränderungen. Interaktionen zwischen Variablenausprägungen können durch diese Erweiterung auf Kosten einer längeren Laufzeit mit höherer Wahrscheinlichkeit aufgedeckt werden als durch den Standardalgorithmus.

C.3.2 Charakterisierung des gier.ga-Algorithmus

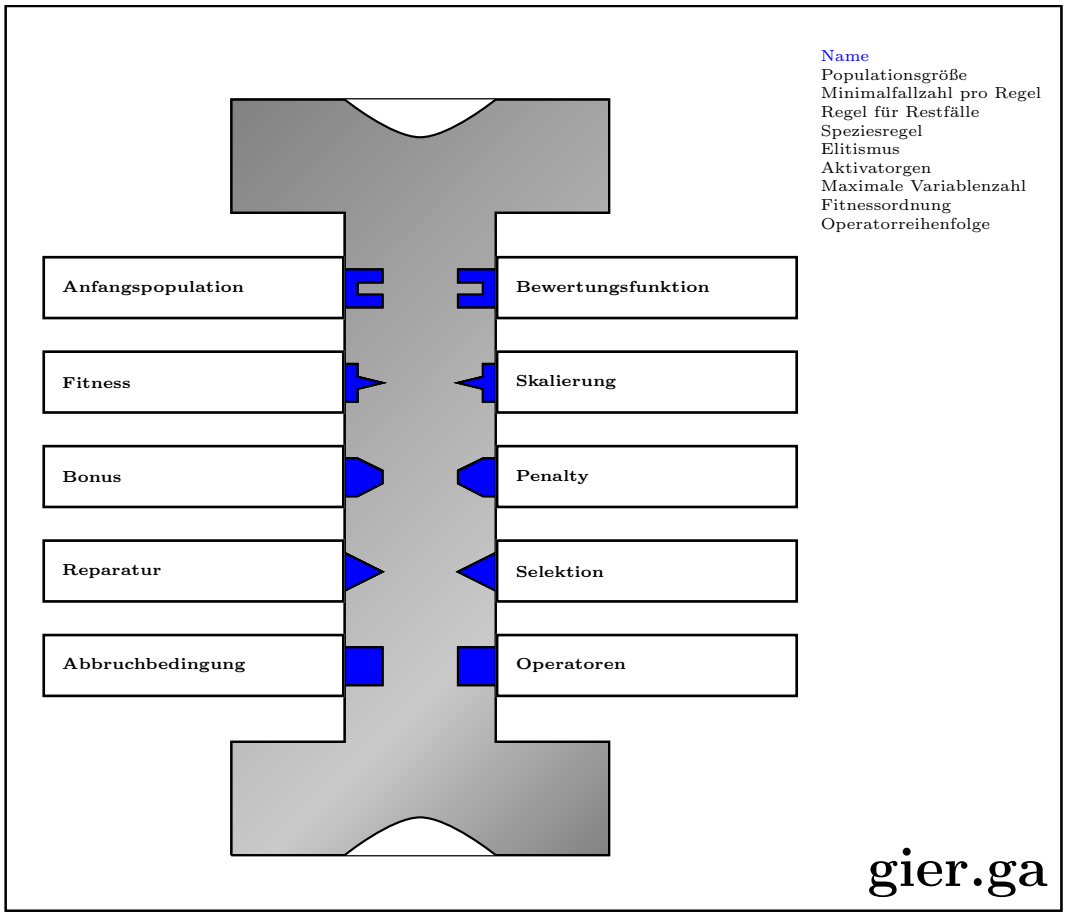


Abbildung C.14: VBk-Darstellung des Algorithmus GIER$_{[1.0]}$.ga: Die Abbildung zeigt die abstrakten Platzhalter für variable Komponentenklassen und Parameter für den Gesamtalgorithmus.

Der GIER$_{[1.0]}$.ga Algorithmus ist wie die meisten der übrigen Algorithmen eigentlich eine Klasse zahlreicher unterschiedlicher Algorithmen. Der Algorithmus dient zu Erstellung hierarchischer Regelmengen. Populationen von Individuen, die Regeln kodieren, werden evolutionär entwickelt, wobei für jede Regel ein eigener Durchlauf des Algorithmus vollzogen wird bis die Anzahl der verbleibenden Fälle die Mindestanzahl pro Regel unterschreitet.

Generelle Parameter des Algorithmus

Abbildung C.14 zeigt die Grundstruktur des Algorithmus. Dabei sind die zehn Komponenten, für die jeweils mehrere Varianten existieren, in der abstrakten Form als Steckkarten eingezeichnet und die Parameter, die unabhängig von den Komponenten gesetzt werden müssen, sind im Diagramm rechts oben

aufgeführt. Die Liste der Parameter besteht neben dem erwähnten Parameter für die Mindestfallzahl pro Regel zunächst aus den folgenden Einstellungen:

- **Name**: Der Name des Prototypen.
- **Populationsgröße**: Die Anzahl der Individuen in der Population.
- **Speziesregel**: Für die Rekombination werden bei Aktivierung der Speziesregel nur Genome mit gleichem Entscheidungswert berücksichtigt. Durch die Speziesregel wird die unter Umständen unsinnige Kombination von Lösungen verhindert, die auf heterogene Klassifikationen abzielen.
- **Elitismus**: Die Regel mit dem höchsten Fitnesswert einer Generation wird mit mindestens einer Kopie in die Folgegeneration aufgenommen, so dass diese Lösung gesichert wird.
- **Aktivatorgen, maximale Variablenzahl und Fitnessordnung**: Bei Verwendung des Aktivatorgens gibt die maximale Variablenanzahl an, wie viele Variablen maximal aktiviert werden dürfen, so dass effektiv eine Beschränkung der berücksichtigten Information pro Regel erreicht wird. Die Fitnessordnung entscheidet darüber, ob bei gleicher Fitness die Anzahl der Variablen (negativ gewichtet) oder die durch die Regel abgedeckte Stichprobe (positive gewichtet) über die Rangfolge der Lösungen entscheidet.
- **Operatorreihenfolge**: In der Operatorreihenfolge wird die Reihenfolge der Anwendung der ausgesuchten Operatoren spezifiziert. Sie ist deshalb als genereller Parameter aufgeführt, weil die Reihenfolge formal rein über Identifikationszahlen der Operatoren ohne die Notwendigkeit von Informationen über die Operatorenauswahl definiert wird.

Die Erstellung der Anfangspopulation

Abbildung C.15: VBk-Darstellung der Komponente „creation“ des Algorithmus GIER$_{[1.0]}$.ga: Eine legitime Auswahl besteht aus der Kombination eines Elements der linken Seite mit einem Element der rechten Seite. Der erste Bestandteil beschreibt das Festsetzen der Prädiktionsklassen, das zweite die Art der Algorithmenerstellung

Die möglichen Komponenten für den Schritt der Erstellung der ersten Population sind in Abbildung C.15 aufgeführt. Die Komponente stellt einen Sonderfall dar, da zwei Teilelemente für eine vollständige Komponenten kombiniert werden müssen. Der eine Teil betrifft die Bestimmung der Prädiktionswerte von Regeln, das zweite die Festlegung der Intervalle. Für die Kreierung eines neuen Individuums können die Intervallgrenzen auf folgende Arten bestimmt werden:

- **Zufallserstellung mit Reparatur (*Uniform*)**: Die Parametergrenzen werden zufällig bestimmt, dabei hat jeder der Werte von 0 bis 100 die gleiche Wahrscheinlichkeit für eine Grenze ausgewählt zu werden. Als untere Grenze wird der kleinere der beiden Werten, als obere Grenze der größere Wert bestimmt, bei gleichen Werten fallen obere und untere Grenze zusammen. Verletzt die so kreierte Regel das Mindestfällekriterium, wird der *Uniform-Repair*-Operator appliziert.
- **Zufallserstellung mit Reparatur (*Random*)**: Die Parametergrenzen werden in gleicher Weise bestimmt, als Reparaturmechanismus wird das *Random-Repair*-Verfahren genutzt.
- **Zufallserstellung**: Die Parametergrenzen werden zufällig bestimmt; es gibt keine Reparatur.

Dabei gibt es für alle drei Verfahren die Parameter „Snowboard-Optimierung“ und“Snowboard-Schrittzahl“: Wahlweise können die kreierten Anfangslösungen durch den Snowboard-Operator (s.o.) lokal optimiert werden bei Angabe der Mindestfallzahl pro Schritt.

Für die Bestimmung des Regelprädiktionswerts gibt es die folgenden Optionen:

- **Zufällig 0, 1 oder 2**: Es werden zusätzlich die Parameter $p(1)$ und $p(2)$ mit der Auflage dass $p(1) \geq 0$, $p(2) \geq 0$ und $p(1)+p(2) \leq 1$ bestimmt. Der Entscheidungswert wird zufällig bestimmt mit $p(0) = 1 - (p(1) + p(2))$, $p(1)$ und $p(2)$ als Wahrscheinlichkeiten für die Auswahl der Entscheidungscodes 0, 1 und 2.
- **Zufällig 0 oder 1**: Zusätzlich wird der Parameter $p(1)$ angegeben mit $0 \leq p(1) \leq 1$. Der Entscheidungswert wird zufällig auf 0 oder 1 gesetzt mit $p(0) = 1 - p(1)$.
- **Wechseln zwischen 0, 1 und 2**: Zusätzlich wird ein Parameter für die Startposition angegeben. Dann werden die Werte beginnend mit der Startposition der Reihe nach an neue Individuen der Population verteilt. Nach einer Vergabe des Werts 2 wird an das nächste Individuum wieder der Wert 0 verteilt.
- **Wechseln zwischen 0 und 1**: Zusätzlich wird ein Parameter für die Startposition angegeben. Dann werden die Werte 0 und 1 beginnend mit der Startposition abwechseln an neue Individuen der Population verteilt.
- **Alle Werte auf** k: Der Entscheidungswert wird für alle Genome auf den Wert k gesetzt. Zulässige Parameterwerte für k sind dementsprechend 0, 1 und 2.

Die Kombination beider Elemente zusammen mit der entsprechenden Parametersetzung legt die Komponente für den Algorithmus erschöpfend fest.

Komponenten zur Fitnessbestimmung

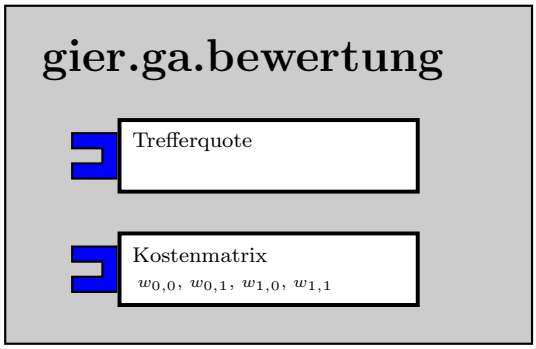


Abbildung C.16: VBk-Darstellung der Komponente „bewertung“ des Algorithmus GIER$_{[1.0]}$.ga

Zur Festlegung der Bewertungsfunktion stehen die beidendiskutierten Optionen der Klassifikationsgütebestimmung zur Verfügung (s. Abbildung C.16). Dies sind:

- **Die Klassifikationsgenauigkeit** A: Für die Bestimmung von A sind keine Parameter notwendig.

- **Der Klassifikationsnutzen**: Zur Verwendung des KNMK ist die Angabe der Werte o_{00}, o_{01}, o_{10} und o_{11} notwendig.

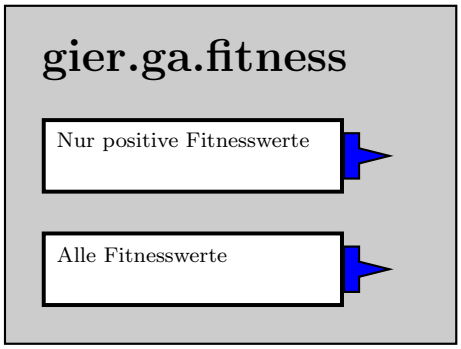


Abbildung C.17: VBk-Darstellung der Komponente „fitness“ des Algorithmus GIER$_{[1.0]}$.ga

Zur Bestimmung des Fitnesswertes stehen ebenfalls zwei Komponenten zur Auswahl (vgl. Abbildung C.17:

- **Nur positive Fitnesswerte**: Negative Bewertungswerte werden auf 0 gesetzt.
- **Alle Fitnesswerte**: Es wird keine Transformation der Bewertungen vorgenommen.

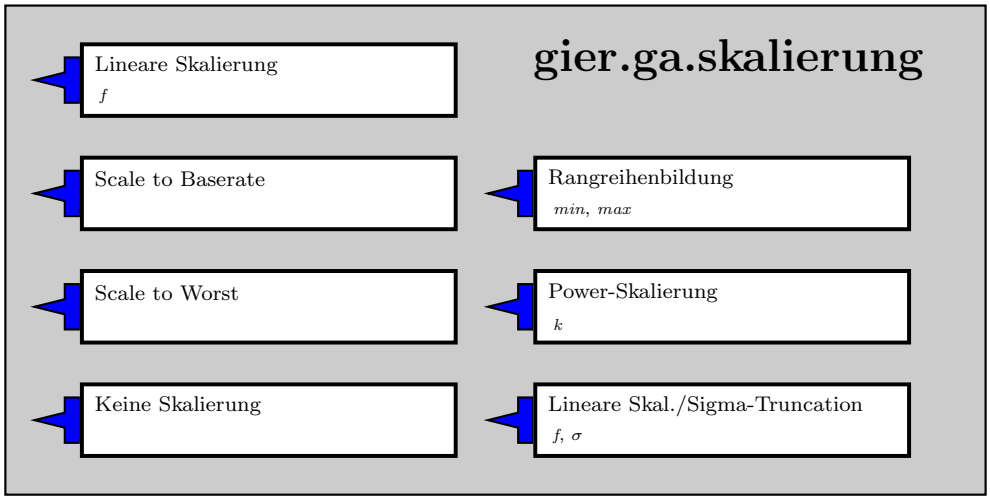


Abbildung C.18: VBk-Darstellung der Komponente „skalierung“ des Algorithmus GIER$_{[1.0]}$.ga

Die erste Option ist nur bei Verwendung des KNMK relevant, da das A-Kriterium keine negativen Werte annehmen kann.

Für die Selektion werden sieben Komponenten angeboten (s. Abbildung C.18):

- **Keine Skalierung**: Es wird keine Transformation der Fitnesswerte vorgenommen.
- ***Scale to Baserate***: Von allen Werten wird der Anteil der Majoritätsklasse an der Trainingsstichprobe subtrahiert.
- ***Scale to Worst***: Von allen Fitnesswerten wird der kleinste Wert subtrahiert. Die Ergebnisse der Transformation sind alle nichtnegativ.
- **Lineare Skalierung**: Nach Angabe des Parameters f wird eine lineare Skalierung durchgeführt.
- **Lineare Skal./*Sigma-Truncation***: Nach Angabe der Parameter f und σ wird eine lineare Skalierung mit *Sigma-Truncation* durchgeführt.

- **Power-Skalierung**: Unter Verwendung des anzugebenden Parameters k wird eine Power-Skalierung durchgeführt.
- **Rangreihenbildung**: Es werden nach Bestimmung der Parameterwerte min und max Rangplätze an alle Fitnesswerte vergeben, so dass der kleinste Wert den Rang min und der größte Wert den Rang max erhält. Alle anderen Werte werden durch das Inkrement $\frac{max-min}{n-1}$ bestimmt. Die Reihenfolge der Fitnesswerte bleibt bei der Transformation unverändert.

Die Komponenten „bewertung“, „fitness“ und „skalierung“ werden in angegebener Reihenfolge durch die Hauptroutine des GIER$_{[1.0]}$.ga-Verfahrens appliziert.

Bonus und Penalty

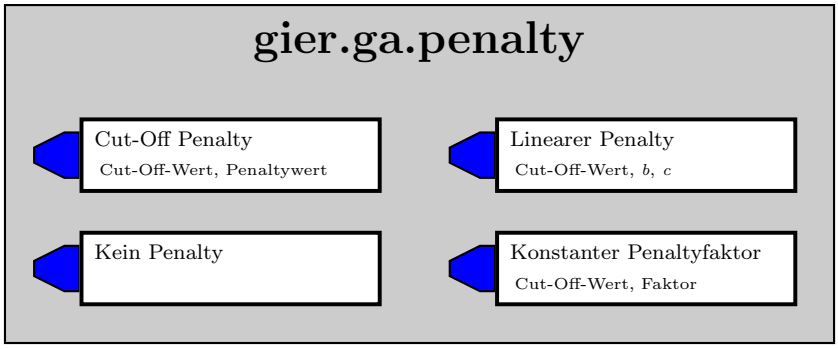


Abbildung C.19: VBk-Darstellung der Komponente „penalty“ des Algorithmus GIER$_{[1.0]}$.ga

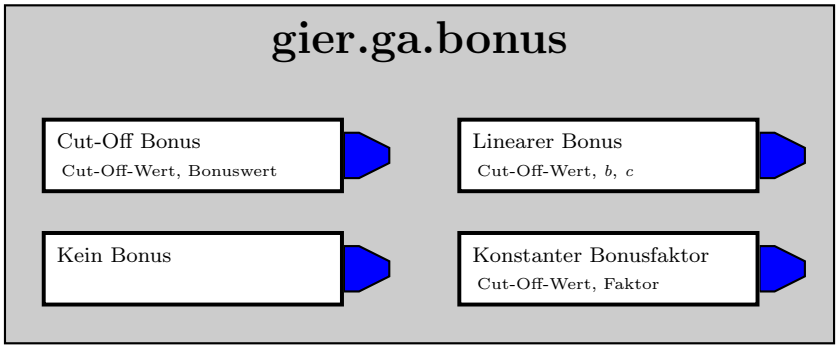


Abbildung C.20: VBk-Darstellung der Komponente „bonus“ des Algorithmus GIER$_{[1.0]}$.ga

Es gibt vier *Penalty*-Varianten (vgl. Abbildung C.19):

- **Kein *Penalty***: Es wird keine Transformation der Fitnesswerte vorgenommen.
- **Cut-Off *Penalty***: Bei Unterschreitung eines angegebenen *Cut-Off-Werts* wird der Fitnesswert durch den angegebenen *Penalty-Wert* ersetzt.
- **Konstanter Penaltyfaktor**: Bei Unterschreitung eines angegebenen *Cut-Off-Werts* wird der Fitnesswert mit dem angegebenem *Penalty-Faktor* multipliziert.
- **Linearer Penalty**: Bei Unterschreitung eines angegebenen *Cut-Off-Werts* wird der Fitnesswert mit dem Parameterwert von b multipliziert und der angegebene Paramterwert von c wird subtrahiert.

Die Bonus-Varianten sind dementsprechend vier analoge Komponenten (vgl. Abbildung C.20):

- **Kein Bonus**: Es wird keine Transformation der Fitnesswerte vorgenommen.
- **Cut-Off Bonus**: Bei Überschreitung eines angegebenen *Cut-Off-Werts* wird der Fitnesswert durch den angegebenen *Bonuswert* ersetzt.

- **Konstanter Bonusfaktor**: Bei Überschreitung eines angegebenen *Cut-Off-Werts* wird der Fitnesswert mit dem angegebenem *Bonusfaktor* multipliziert.
- **Linearer Bonus**: Bei Überschreitung eines angegebenen *Cut-Off-Werts* wird der Fitnesswert mit dem Parameterwert von b multipliziert und der angegebene Paramterwert von c wird addiert.

Selektion

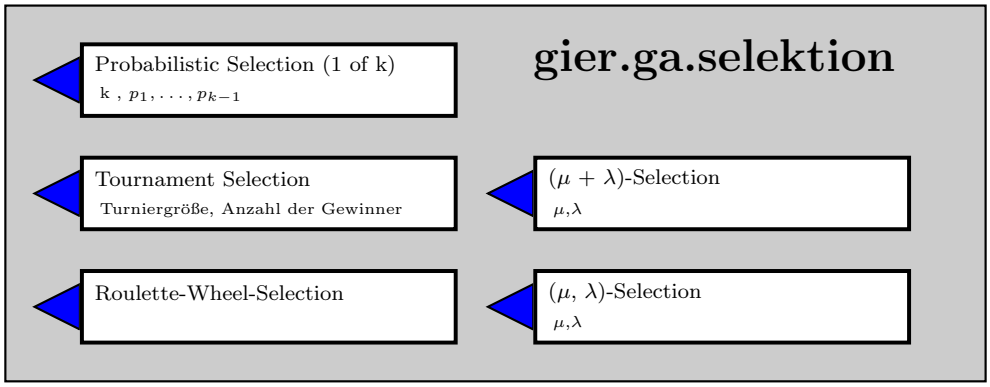


Abbildung C.21: VBk-Darstellung der Komponente „selektion“ des Algorithmus GIER$_{[1.0]}$.ga

Die Selektionskomponenten implementieren die vorgestellten Selektions-Mechanismen. Insgesamt stehen fünf Varianten zur Auswahl (s. Abbildung C.21):

- ***Roulette-Wheel-Selection***: Ein proportionales Selektionsverfahren wird eingesetzt.
- ***Tournament Selection***: Es wird eine Tournament Selection mit einer durch einen Parameter definierten *Turniergröße* und einer durch einen zweiten Parameter gegebenen *Anzahl von Gewinnern* durchgeführt.
- ***Probabilistic Selection (1 of k)***: Es wird eines von k zufällig ausgewählten Individuen ausgewählt. Zusätzlich zum Parameter k werden $k-1$ Wahrscheinlichkeiten angegeben, die bestimmen, wie wahrscheinlich die Auswahl welches Individuums in Reihenfolge der Fitnesswerte ist.
- **(μ + λ)-*Selection***: Unter Verwendung der Parameter μ und λ wird eine (μ + λ)-*Selection* durchgeführt.
- **(μ, λ)-Selection**: Unter Verwendung der Parameter μ und λ wird eine (μ, λ)-*Selection* durchgeführt.

Genetische Operatoren

Von den genetischen Operatoren können prinzipiell beliebig viele ausgewählt werden, einzelne Operatoren auch mehrfach. Die Reihenfolgeliste bestimmt die Reihenfolge der Anwendung der Operatoren. Einige Operatoren stehen nur zur Auswahl, wenn die Aktivierungsgen-Option des Algorithmus genutzt wird. Mögliche Operatoren sind (s. Abbildung C.22):

- **Standard-Mutation**: Der Parameter p bestimmt die Wahrscheinlichkeit einer Mutation pro Bit pro Individuum.
- **Generalisierungs-Mutation**: Der Parameter p bestimmt die Wahrscheinlichkeit einer Mutation pro Grenze eines Individuums: Die Mutation ersetzt eine untere Grenze durch den Wert 0 bzw. eine obere Grenze durch den Wert 100.
- ***Single-Point-Cross-Over***: Der Parameter p bestimmt die Wahrscheinlichkeit, dass eine *Single-Point-Cross-Over*-Operation für die gesamte Population durchgeführt wird.

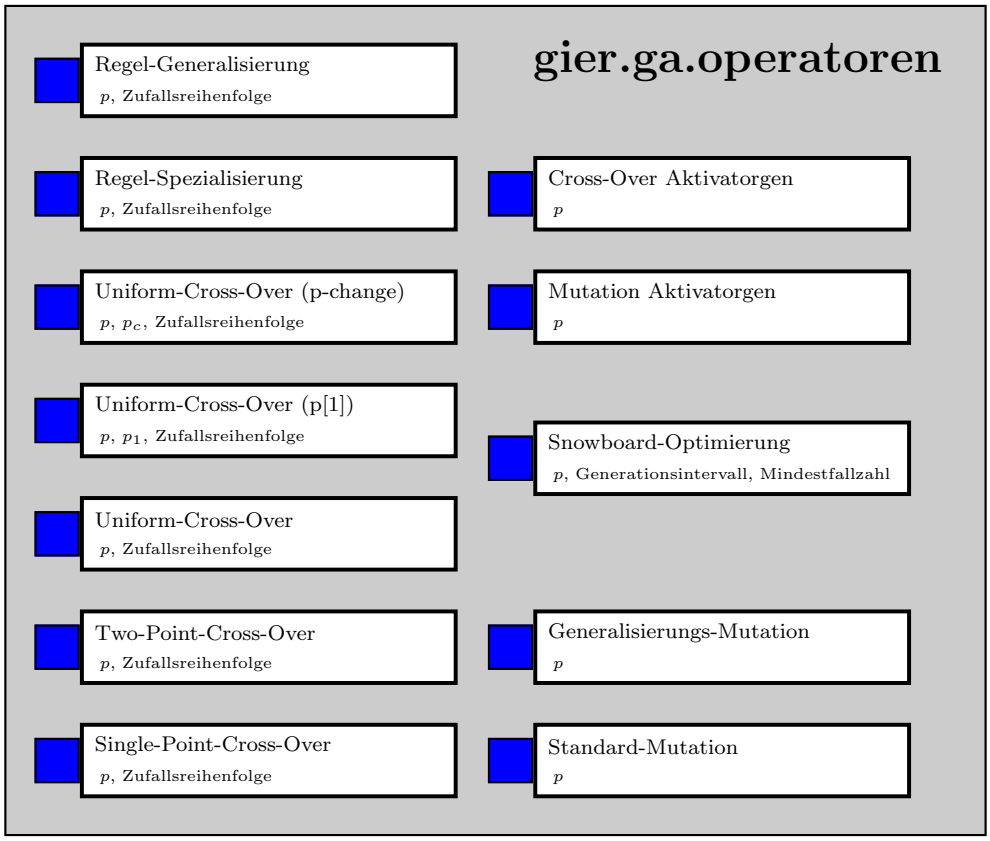


Abbildung C.22: VBk-Darstellung der genetischen Operatoren des Algorithmus GIER$_{[1.0]}$.ga: Es kann eine beliebige Anzahl von Operatoren ausgewählt werden auch bei Mehrfachverwendung einzelner Operatoren. Die beiden Aktivatorgen-Operatoren kommen nur bei Verwendung der Aktivierungsgen-Option zum Einsatz.

- ***Two-Point-Cross-Over***: Der Parameter p bestimmt die Wahrscheinlichkeit, dass eine *Two-Point-Cross-Over*-Operation für die gesamte Population durchgeführt wird.

- **Uniform-Cross-Over**: Der Parameter p bestimmt die Wahrscheinlichkeit, dass eine *Uniform-Cross-Over*-Operation für die gesamte Population durchgeführt wird.

- ***Uniform-Cross-Over (p[1])***: Der Parameter p bestimmt die Wahrscheinlichkeit, dass eine *Uniform-Cross-Over-Operation* für die gesamte Population durchgeführt wird. Der Parameter p_1 gibt die Wahrscheinlichkeit an, dass der erste Nachkomme die betreffende Genomstelle vom ersten Elternteil erbt.

- ***Uniform-Cross-Over (p-change)***: Der Parameter p bestimmt die Wahrscheinlichkeit, dass eine *Uniform-Cross-Over*-Operation für die gesamte Population durchgeführt wird. Der Parameter p_c gibt die Wahrscheinlichkeit an, dass die Zuordnung von Elterngenomen zu Filialgenomen geändert wird. Zu Anfang erbt das erste Filialgenom Genomstellen vom ersten Elternteil, das zweite Filialgenom vom zweiten Elternteil, beim ersten Wechsel wird die Zuordnung vertauscht, beim zweiten wiederhersgestellt usw.

- **Regel-Spezialisierung**: Der Parameter p bestimmt die Wahrscheinlichkeit, dass eine Regel-Spezialisierungs-Cross-Over-Operation für die gesamte Population durchgeführt wird. Das erste Nachkommengenom erhält dabei die größere untere Grenze und die kleinere obere Grenze von seinen Eltern.

- **Regel-Generalisierung**: Der Parameter p bestimmt die Wahrscheinlichkeit, dass eine Regel-Generalisierungs-Cross-Over-Operation für die gesamte Population durchgeführt wird. Das erste Nachkommengenom erhält dabei die kleiner untere Grenze und die größere obere Grenze von seinen Eltern.

- **Snowboard-Optimierung**: Der Parameter p bestimmt die Wahrscheinlichkeit, dass eine Snowboard-Optimierungs-Operation für ein Individuum durchgeführt wird. Der Parameter *Generationsintervall* gibt dabei das Intervall an, in dem die Prüfung für jedes Individuum erneut durchgeführt wird. Der Parameter *Mindestfallzahl* bestimmt die Funktionsweise des Snowboard-Operators.
- ***Cross-Over* (Aktivatorgen)**: Der Parameter p bestimmt die Wahrscheinlichkeit, dass eine *Cross-Over*-Operation für ein Aktivator-Gen durchgeführt wird, wenn die Ativierungsgenregel gewählt worden ist.
- **Mutation Aktivatorgen**: Der Parameter p bestimmt die Wahrscheinlichkeit, dass eine Mutations-Operation für ein Aktivator-Gen durchgeführt wird, wenn die Aktivierungsgenregel gewählt worden ist.

Reparatur

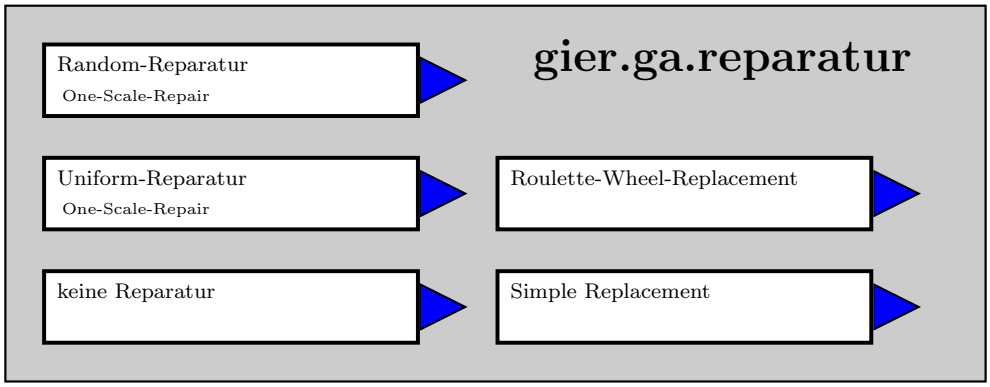


Abbildung C.23: VBk-Darstellung der Komponente „reparatur“ des Algorithmus GIER$_{[1.0]}$.ga

Für die in der Arbeit beschriebenen Reparaturmechanismen und ihre Alternativen ist eine eigene Komponente eingeführt worden. Die Varianten sind die folgenden (vgl. Abbildung C.23):

- **keine Reparatur**: Es wird keine Reparatur durchgeführt.
- ***Uniform*-Reparatur**: Es wird eine *Uniform*-Reparatur durchgeführt. Der Parameter *One-Scale-Repair* gibt an, ob zuvor eine *One-Scale-Repair*-Operation versucht wird.
- ***Random*-Reparatur**: Es wird eine *Random*-Reparatur durchgeführt. Der Parameter *One-Scale-Repair* gibt an, ob zuvor eine *One-Scale-Repair*-Operation versucht wird.
- ***Simple Replacement***: Die ungültige Lösung wird durch eine zufällig ausgewählte andere Lösung ersetzt.
- ***Roulette-Wheel-Replacement***: Die ungültige Lösung wird durch eine nach dem Prinzip der *Roulette-Wheel-Selection* ausgewählte andere Lösung ersetzt.

Die letzten beiden Alternativen sind keine Reparaturlösungen, sondern Ersetzungsregelungen, was einem Entfernen der problematischen Lösung aus der Population gleichkommt.

Abbruchbedingung

Es stehen fünf verschiedene Komponenten für die Abbruchbedingung zur Auswahl (vgl. Abbildung C.24):

- **Anzahl Generationen**: Der Algorithmus wird nach einer festen Anzahl von Generationen beendet, die durch den Parameter *Generationszahl* angegeben wird.
- **Minimale Klassifikationsgenauigkeit**: Der Algorithmus wird nach Erreichen einer Mindestklassifikationsgenauigkeit beendet, die durch den Parameter *Mindestwert* angegeben wird. Dabei gibt es eine untere und eine obere Grenze für die Laufzeit in Generationen, die durch die beiden Parameter *Mindestzahl* und *Maximalzahl* spezifiziert werden.

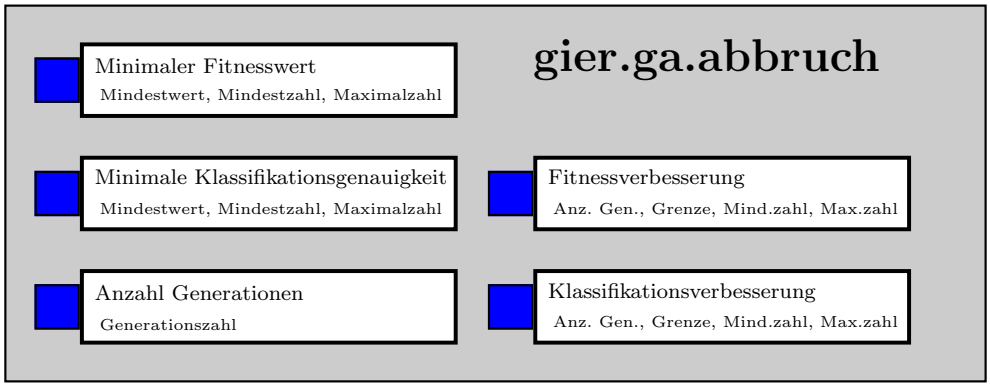


Abbildung C.24: VBk-Darstellung der Komponente „abbruch“ des Algorithmus GIER$_{[1.0]}$.ga

- **Minimaler Fitnesswert**: Der Algorithmus wird nach Erreichen eines Mindestfitnesswerts beendet, der durch den Parameter *Mindestwert* angegeben wird. Dabei gibt es eine untere und eine obere Grenze für die Laufzeit in Generationen, die durch die beiden Parameter *Mindestzahl* und *Maximalzahl* spezifiziert werden.

- **Klassifikationsverbesserung**: Der Algorithmus wird beendet, nachdem für eine durch den Parameter *Anz. Gen.* bestimmte Anzahl von Generationen die Verbesserung der besten Klassifikationsgenauigkeit in der Population einen durch den Parameter *Grenze* angegebenen Wert nicht überschreitet. Dabei gibt es zudem eine untere und eine obere Grenze für die Laufzeit in Generationen, die durch die beiden Parameter *Mindestzahl* und *Maximalzahl* spezifiziert werden.

- **Fitnessverbesserung**: Der Algorithmus wird beendet, nachdem für eine durch den Parameter *Anz. Gen.* bestimmte Anzahl von Generationen die Verbesserung der besten Fitness in der Population einen durch den Parameter *Grenze* angegebenen Wert nicht überschreitet. Dabei gibt es zudem eine untere und eine obere Grenze für die Laufzeit in Generationen, die durch die beiden Parameter *Mindestzahl* und *Maximalzahl* spezifiziert werden.

Damit ist die Darstellung des GIER$_{[1.0]}$.ga-Algorithmus abgeschlossen.

C.3.3 Charakterisierung des gier.gaset-Algorithmus

Der in Abbildung C.25 zusammengefasste GIER$_{[1.0]}$.gaset-Algorithmus stellt eine Variation des GIER$_{[1.0]}$.ga Algorithmus dar. Im Unterschied zum oben dargestellten Algorithmus wird nicht für jede Regel ein eigener Durchlauf gestartet. Statt dessen kodiert jedes Individuum eine vollständige hierarchische Regelmenge. Dementsprechend besteht das Genom aus einzelnen Teilen für jede enthaltene Regel. Die allgemeinen Parameter für den GIER$_{[1.0]}$.gaset Algorithmus sind die Folgenden:

- **Name**: Der Name des Prototypen.

- **Populationsgröße**: Die Anzahl der Individuen in der Population.

- **Maximale Regelzahl**: Der Parameter spezifiziert die maximal mögliche Regelzahl für jedes Individuum.

- **Elitismus**: Die Regel mit dem höchsten Fitnesswert einer Generation wird mit einer mindestens einer Kopie in die Folgegeneration aufgenommen, so dass diese Lösung gesichert wird.

- **Aktivatorgen, maximale Variablenzahl und Fitnessordnung**: Bei Verwendung des Aktivatorgens gibt die maximale Variablenanzahl an, wie viele Variablen maximal aktiviert werden dürfen, so dass effektiv eine Beschränkung der berücksichtigten Information pro Regel erreicht wird. Die Fitnessordnung entscheidet darüber, in welcher Reihenfolge die Variablen Anzahl der Variablen, Anzahl der Regeln und durch die Regel abgedeckte Stichprobe bei gleicher Fitness über die Rangfolge der Lösungen entscheiden.

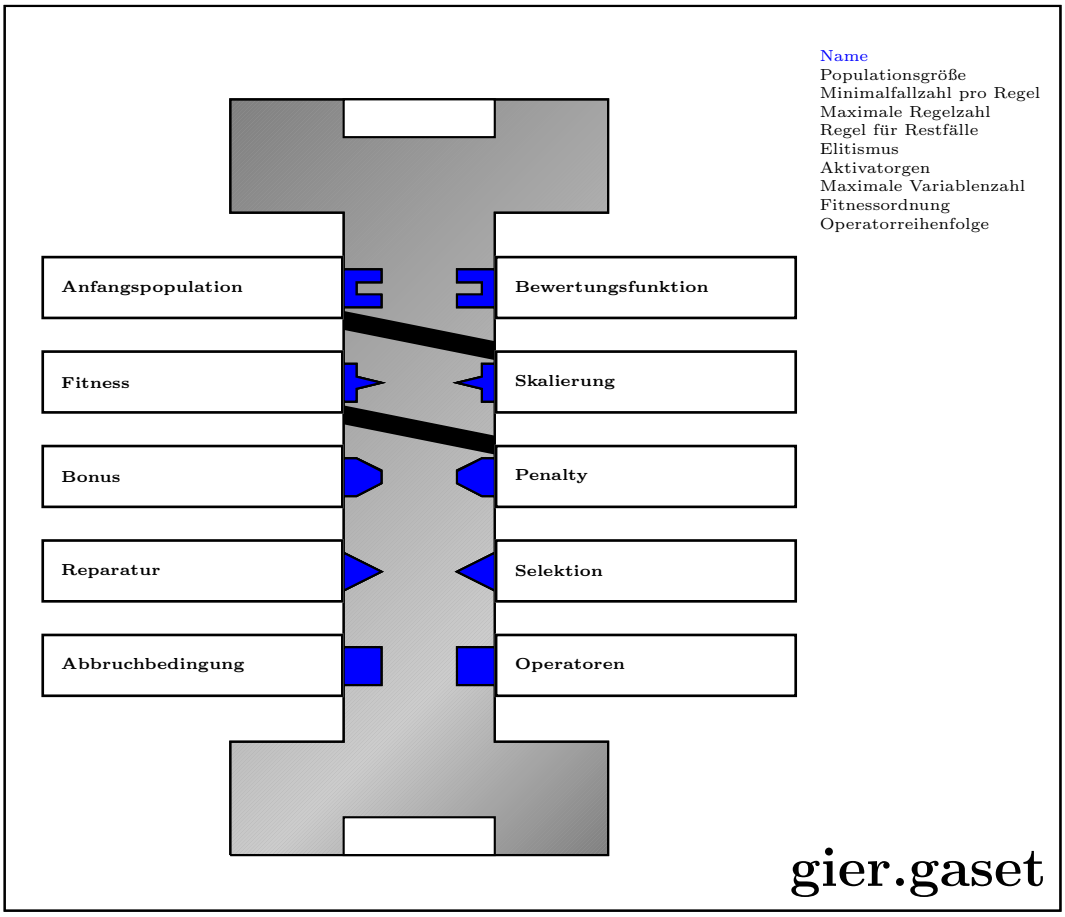


Abbildung C.25: VBk-Darstellung des Algorithmus GIER$_{[1.0]}$.gaset: Die Abbildung zeigt die abstrakten Platzhalter für variable Komponentenklassen und Parameter für den Gesamtalgorithmus.

- **Operatorreihenfolge**: In der Operatorreihenfolge wird die Reihenfolge der Anwendung der ausgesuchten Operatoren spezifiziert. Sie ist deshalb als genereller Parameter aufgeführt, weil die Reihenfolge formal rein über Identifikationszahlen der Operatoren ohne die Notwendigkeit von Informationen über die Operatorenauswahl definiert wird.

Prinzipiell haben die Komponenten des GIER$_{[1.0]}$.gaset-Algorithmus die gleichen Namen wie die Komponenten des GIER$_{[1.0]}$.ga-Algorithmus, aus der Umstellung der Repräsentation erwachsen allerdings einige notwendige Umstellungen in Einzelkomponenten, so dass für die Module „creation“ andere Varianten zur Verfügung stehen und einige Operatoren abgeändert wurden. Zudem musste die „reparatur“-Komponente angepasst werden, so dass Reparaturoperationen nur dann auf die erste Regel angewendet werden, wenn keine gültige Regel im Regelsatz vorliegt. Zudem wird bei dieser Reparatur kein One-Scale-Repair-Versuch unternommen.

Beschreibung der modularen Wiederverwendung

Am Beispiel des GIER$_{[1.0]}$.gaset-Algorithmus lässt sich das Prinzip der modularen Wiederverwendung demonstrieren: Alle oben beschriebenen Komponenten des GIER$_{[1.0]}$.ga-Algorithmus mit Ausnahme spezifischer genetischer Operatoren finden sich auch in diversen anderen Algorithmen wieder. Durch die objektorientierte Programmierung können Klassen auch in andere Algorithmenklassen als Komponentenklassen eingebunden werden, so dass der Programmieraufwand gesenkt werden kann. Abbildung C.26 zeigt, welche Komponenten von welchen Algorithmen wiederverwendet werden.

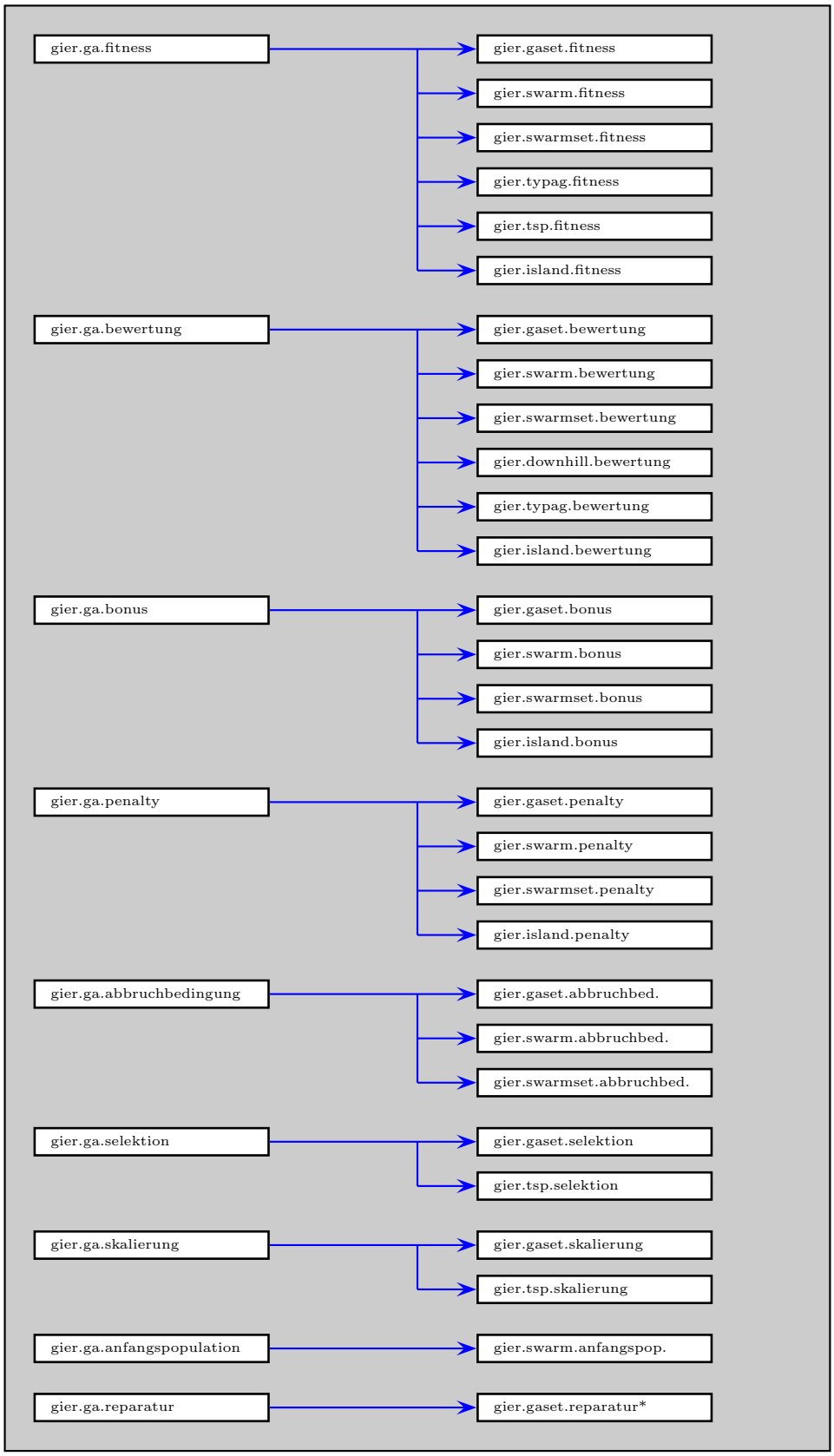


Abbildung C.26: VBk-Darstellung des Prinzips der modularen Wiederverwendung für GIER$_{[1.0]}$.ga-Komponenten: Die Pfeile geben an, welche Komponenten anderer Algorithmen die angegebenen Komponenten des GIER$_{[1.0]}$.ga-Algorithmus verwenden können

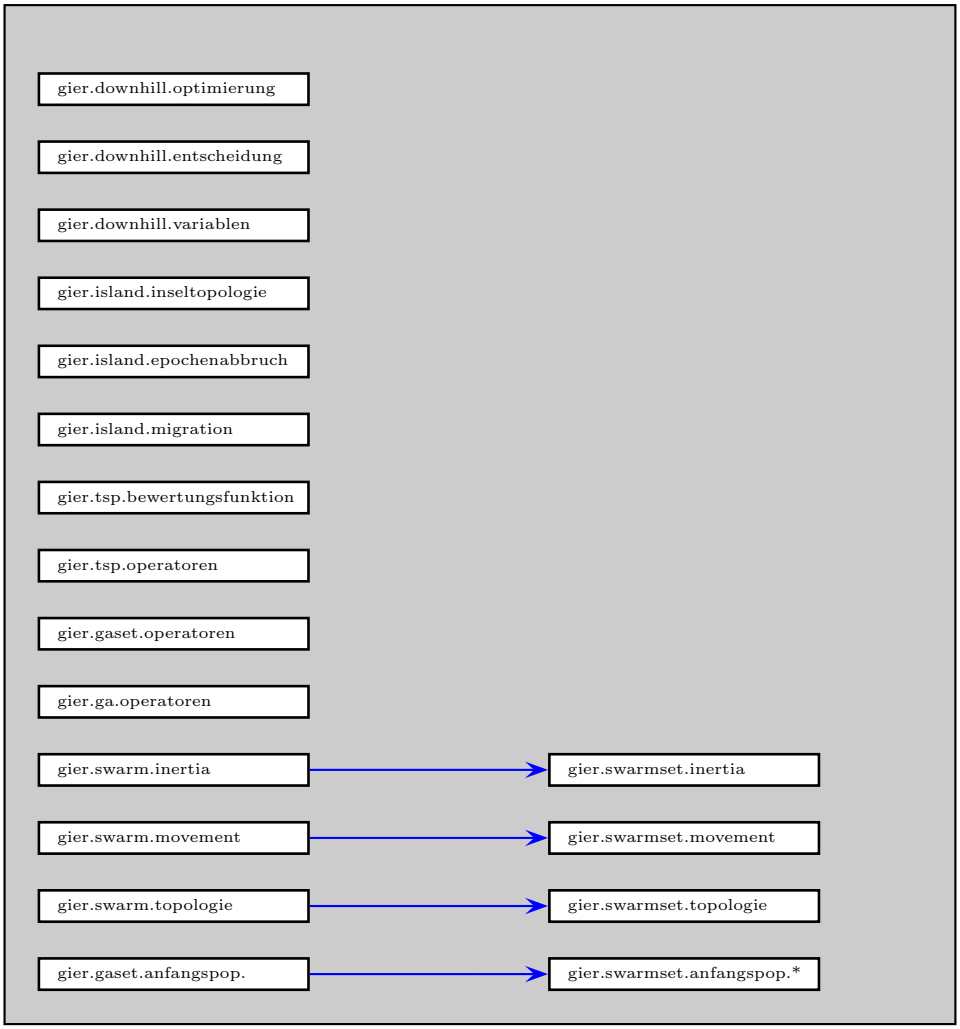


Abbildung C.27: VBk-Darstellung der nicht in $\text{GIER}_{[1.0]}$.ga enthaltenen Komponenten: Die dargestellten Komponenten sind nicht in $\text{GIER}_{[1.0]}$.ga enthalten. Eine modulare Wiederverwendung wird dort durch Pfeile angezeigt, wo sie realisiert worden ist.

Um einen Überblick über die gesamte Komponentensammlung zu geben, sind die nicht in $\text{GIER}_{[1.0]}$.ga enthaltenen Komponenten und ihre Wiederverwendungen in Abbildung C.27 aufgeführt. Komponenten auf der linken Seite der Abbildung, von denen keine Pfeile ausgehen, sind spezifisch für die Algorithmen, die sie enthalten.

Erstellung neuer Individuen

In Abbildung C.28 sind die Komponenten der Klasse $\text{GIER}_{[1.0]}$.gaset.anfangspopulation aufgeführt. Auch diese Komponenten besteht aus zwei Teilen, von denen der erste die Art der Grenzintervallbestimmung, der zweite aus der Bestimmung der Prädiktionswerte besteht. Eine Besonderheit stellen fünf Komponenten dar, die einen vollständigen zweiten Algorithmus nutzen, um die erste Population zu erstellen. Dabei sind diese Algorithmen im Allgemeinen Algorithmen mit kürzerer Laufzeit, sie liefern damit eine vorbearbeitete Regelmenge, die für den $\text{GIER}_{[1.0]}$.gaset-Algorithmus die Grundlage bildet (vgl. auch Abschnitt C.3.8).

- ▷ **Zufallserstellung**: Die Regelgrenzen werden für jede Variable jeder Regel jeder Regelmenge zufällig erstellt, wobei bei unterschiedlichen Grenzen die kleinere Grenze als untere Grenze gewählt wird, der größere Wert als obere Grenze.

- ▷ **Zufallserstellung mit Reparatur (*Random*)**: Die Regelgrenzen werden für jede Variable jeder Regel jeder Regelmenge zufällig erstellt. Anschließend wird jedes ungültige Individuum mit der *Random*-Reparatur behandelt.

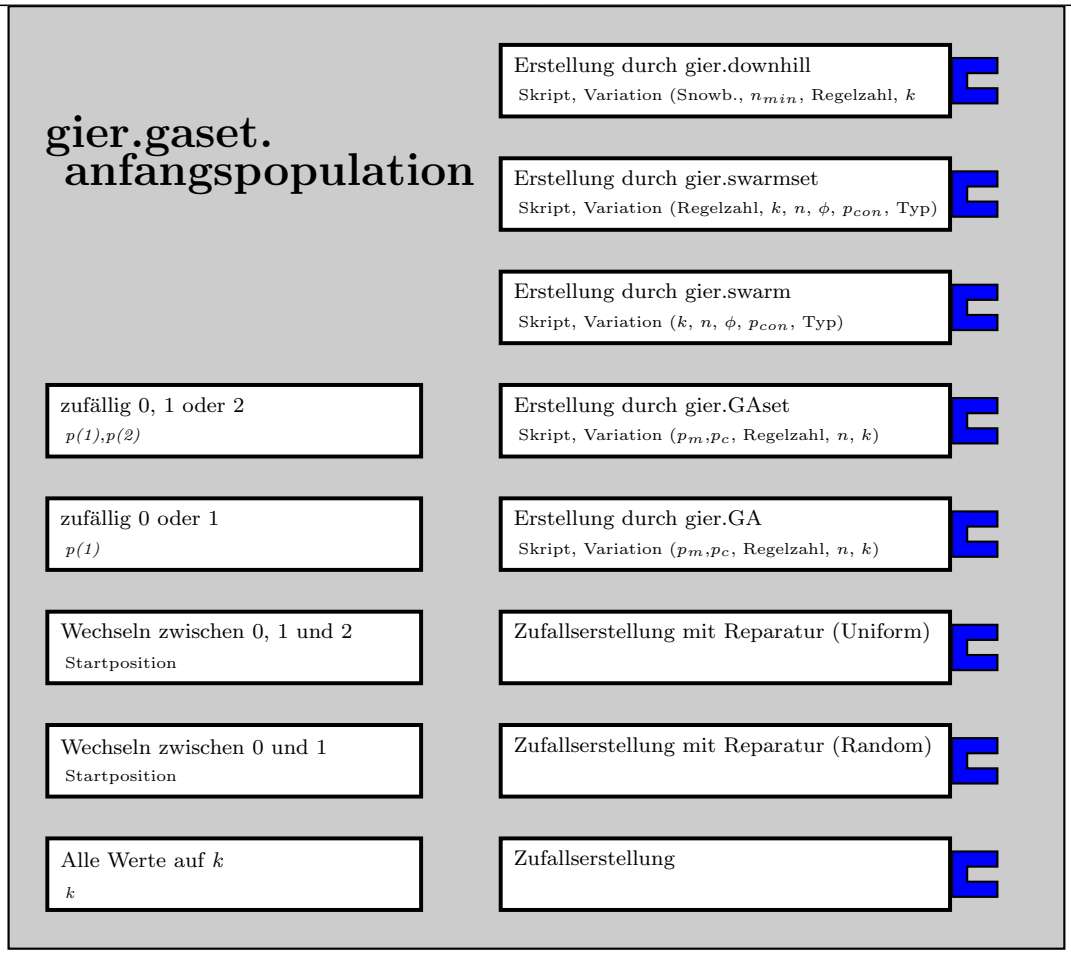


Abbildung C.28: VBk-Darstellung der Komponente „creation“ des Algorithmus GIER$_{[1.0]}$.gaset: Eine legitime Auswahl besteht aus der Kombination eines Elements der linken Seite mit einem Element der rechten Seite. Der erste Bestandteil beschreibt das Festsetzen der Prädiktionsklassen, das zweite die Art der Algorithmenerstellung

- **Zufallserstellung mit Reparatur (*Uniform*)**: Die Regelgrenzen werden für jede Variable jeder Regel jeder Regelmenge zufällig erstellt. Anschließend wird jedes ungültige Individuum mit der *Uniform*-Reparatur behandelt.

- **Erstellung durch Gier$_{[1.0]}$.downhill**: Die Regelmenge für jedes Individuum der Population wird durch einen Durchlauf des GIER$_{[1.0]}$.downhill-Algorithmus kreiert. Die dafür notwendigen Parameter sind der *Scriptname*, sowie Angaben zur Variation des Prototypen für einzelne Individuen. Dies sind im einzelnen Variationsangaben für die Verwendung der Snowboard-Komponente (*Snowb.*) die minimale Fallzahl pro Schritt (n_{min}), die *Regelzahl* und die Maximalzahl an Variablen (k). Zusätzlich sind neben der Angabe über Verwendung der Variation mit Ausnahme der Snowboard-Variation Angaben zu Startposition, Minimum, Maximum und Schrittzahl bzw. Zufallsauswahl der Werte notwendig.

- **Erstellung durch Gier$_{[1.0]}$.ga**: Die Regelmenge für jedes Individuum der Population wird durch einen Durchlauf des GIER$_{[1.0]}$.ga-Algorithmus kreiert. Die dafür notwendigen Parameter sind der *Scriptname*, sowie Angaben zur Variation des Prototypen für einzelne Individuen. Dies sind im einzelnen Variationsangaben für die Größe der Population (n), die *Regelzahl*, die Maximalzahl an Variablen (k), die Wahrscheinlichkeit einer Mutation (p_m) und die Wahrscheinlichkeit eines *Cross-Over* (p_c). Zusätzlich sind neben der Angabe über Verwendung der Variation Angaben zu Startposition, Minimum, Maximum und Schrittzahl bzw. Zufallsauswahl der Werte notwendig.

- **Erstellung durch Gier$_{[1.0]}$.gaset**: Die Regelmenge für jedes Individuum der Population wird durch einen Durchlauf des GIER$_{[1.0]}$.gaset-Algorithmus kreiert. Die dafür notwendigen Parameter sind der *Scriptname*, sowie Angaben zur Variation des Prototypen für einzelne Individuen. Dies sind im einzelnen Variationsangaben für die Größe der Population (n), die *Regelzahl*, die Maximalzahl

an Variablen (k), die Wahrscheinlichkeit einer Mutation (p_m) und die Wahrscheinlichkeit eines *Cross-Over* (p_c). Zusätzlich sind neben der Angabe über Verwendung der Variation Angaben zu Startposition, Minimum, Maximum und Schrittzahl bzw. Zufallsauswahl der Werte notwendig.

- **Erstellung durch Gier$_{[1.0]}$.swarm**: Die Regelmenge für jedes Individuum der Population wird durch einen Durchlauf des GIER$_{[1.0]}$.swarm-Algorithmus kreiert. Die dafür notwendigen Parameter sind der *Scriptname*, sowie Angaben zur Variation des Prototypn für einzelne Individuen. Dies sind im einzelnen Variationsangaben für die Größe des Schwarms (n), diw *Regelzahl*, die Maximalzahl an Variablen (k), den ϕ-Parameter, den p_{con}-Parameter und den Topologie-Typ. Zusätzlich sind neben der Angabe über Verwendung der Variation Angaben zu Startposition, Minimum, Maximum und Schrittzahl bzw. Zufallsauswahl der Werte notwendig.

- **Erstellung durch Gier$_{[1.0]}$.swarmset**: Die Regelmenge für jedes Individuum der Population wird durch einen Durchlauf des GIER$_{[1.0]}$.swarmset-Algorithmus kreiert. Die dafür notwendigen Parameter sind der *Scriptname*, sowie Angaben zur Variation des Prototypen für einzelne Individuen. Dies sind im einzelnen Variationsangaben für die Größe des Schwarms (n), der *Regelzahl*, die Maximalzahl an Variablen (k), den ϕ-Parameter, den p_{con}-Parameter und den Topologie-Typ. Zusätzlich sind neben der Angabe über Verwendung der Variation Angaben zu Startposition, Minimum, Maximum und Schrittzahl bzw. Zufallsauswahl der Werte notwendig.

Hinzu kommt eine Bestimmung der Vergabe von Prädiktionswerten für die Regeln der Regelmenge jedes Individuums. Bei der Erstellung durch einen anderen Algorithmus ist hier die Option zu wählen, alle Prädiktionswerte auf 2 zu setzen. Es stehen die folgenden Varianten zur Verfügung:

- **Zufällig 0, 1 oder 2**: Es werden zusätzlich die Parameter $p(1)$ und $p(2)$ mit der Auflage dass $p(1) \geq 0$, $p(2) \geq 0$ und $p(1) + p(2) \leq 1$ bestimmt. Der Entscheidungswert jeder Regel in jeder Regelmenge wird zufällig bestimmt mit $p(0) = 1 - (p(1) + p(2))$, $p(1)$ und $p(2)$ als Wahrscheinlichkeiten für die Auswahl der Entscheidungscodes 0, 1 und 2.

- **Zufällig 0 oder 1**: Zusätzlich wird der Parameter $p(1)$ angegeben mit $0 \leq p(1) \leq 1$. Der Entscheidungswert jeder Regel in jeder Regelmenge wird zufällig auf 0 oder 1 gesetzt mit $p(0) = 1 - p(1)$.

- **Wechseln zwischen 0, 1 und 2**: Zusätzlich wird ein Parameter für die Startposition angegeben. Dann werden die Werte beginnend mit der Startposition der Reihe nach an die Regeln eines neuen Individuums verteilt. Nach einer Vergabe des Wertes 2 wird an die nächste Regel wieder der Wert 0 verteilt.

- **Wechseln zwischen 0 und 1**: Zusätzlich wird ein Parameter für die Startposition angegeben. Dann werden die Werte 0 und 1 beginnend mit der Startposition abwechseln an die Regeln eines neuen Individuums der Population verteilt.

- **Alle Werte auf k**: Der Entscheidungswert wird für alle Regeln jedes Individuums auf den Wert k gesetzt. Zulässige Parameterwerte für k sind dementsprechend 0, 1 und 2.

Genetische Operatoren

Die genetischen Operatoren für die Regelmengenkodierung unterscheiden sich teilweise von denen bei einer Einzelregelkodierung. Es gibt hier die folgenden Auswahlmöglichkeiten (s. auch Abbildung C.29):

- **Standard-Mutation**: Der Parameter p bestimmt die Wahrscheinlichkeit einer Mutation pro Bit pro Individuum.

- **Generalisierungs-Mutation**: Der Parameter p bestimmt die Wahrscheinlichkeit einer Mutation pro Grenze eines Individuums: Die Mutation ersetzt eine untere Grenze durch den Wert 0 bzw. eine obere Grenze durch den Wert 100.

- **Regeltausch-Mutation**: Der Parameter p bestimmt die Wahrscheinlichkeit einer Vertauschung zweier zufällig ausgewählter Regeln des Individuums.

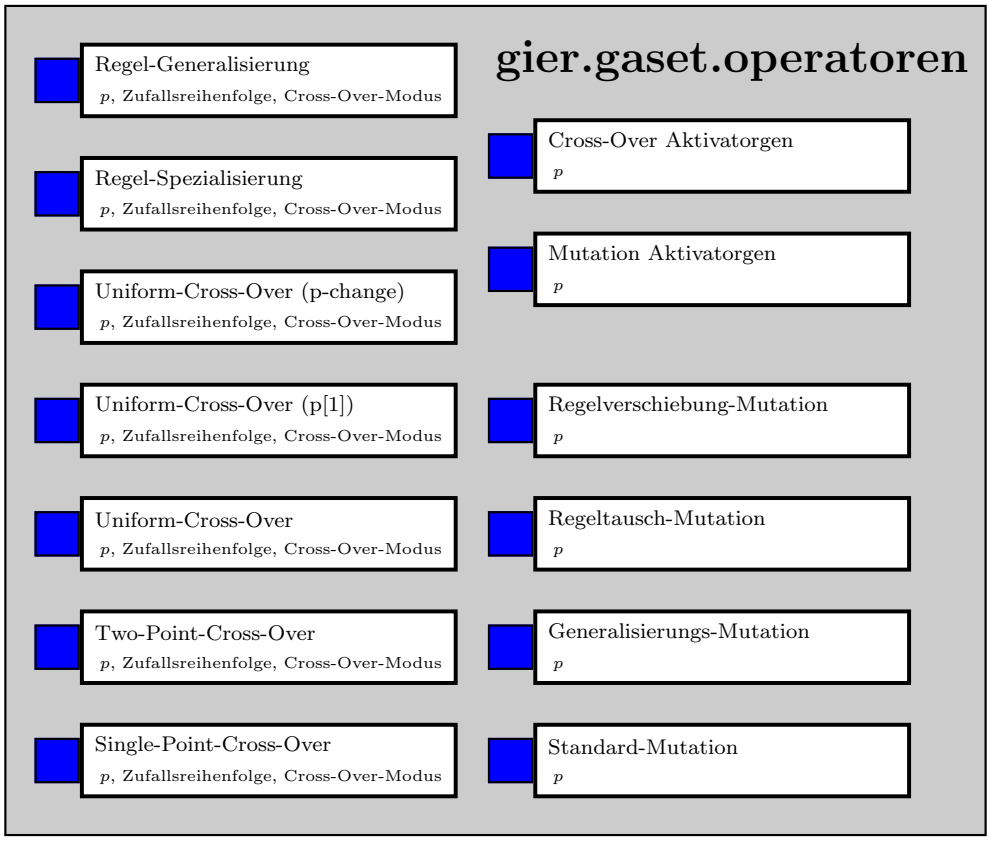


Abbildung C.29: VBk-Darstellung der genetischen Operatoren des Algorithmus GIER$_{[1.0]}$.gaset: Es kann eine beliebige Anzahl von Operaotrn ausgewählt werden bei Mehrfachverwendung einzelner Operatoren. Die beiden Aktivatorgen-Operatoren kommen nur bei Verwendung der Aktivierungsgen-Option zum Enisatz.

- **Regelverschiebung-Mutation**: Der Parameter p bestimmt die Wahrscheinlichkeit einer zufällig ausgewählten Regel des Individuums an eine zufällig ausgewählte Stelle.
- ***Single-Point-Cross-Over***: Der Parameter p bestimmt die Wahrscheinlichkeit, dass eine *Single-Point-Cross-Over*-Operation für die gesamte Population durchgeführt wird. Der Parameter *Cross-Over-Modus* bestimmt die Art des *Cross-Over* (s.u.).
- ***Two-Point-Cross-Over***: Der Parameter p bestimmt die Wahrscheinlichkeit, dass eine *Two-Point-Cross-Over*-Operation für die gesamte Population durchgeführt wird. Der Parameter *Cross-Over-Modus* bestimmt die Art des *Cross-Over*.
- ***Uniform-Cross-Over***: Der Parameter p bestimmt die Wahrscheinlichkeit, dass eine *Uniform-Cross-Over*-Operation für die gesamte Population durchgeführt wird. Der Parameter *Cross-Over-Modus* bestimmt die Art des *Cross-Over*.
- ***Uniform-Cross-Over* (p[1])**: Der Parameter p bestimmt die Wahrscheinlichkeit, dass eine *Uniform-Cross-Over*-Operation für die gesamte Population durchgeführt wird. Der Parameter p_1 gibt die Wahrscheinlichkeit an, dass der erste Nachkomme die betreffende Genomstelle vom ersten Elternteil erbt. Der Parameter *Cross-Over-Modus* bestimmt die Art des *Cross-Over*.
- ***Uniform-Cross-Over* (p-change)**: Der Parameter p bestimmt die Wahrscheinlichkeit, dass eine *Uniform-Cross-Over*-Operation für die gesamte Population durchgeführt wird. Der Parameter p_c gibt die Wahrscheinlichkeit an, dass die Zuordnung von Elterngenomen zu Filialgenomen geändert wird. Zu Anfang erbt das erste Filialgenom Genomstellen vom ersten Elternteil, das zweite Filialgenom vom zweiten Elternteil, beim ersten Wechsel wird die Zuordnung vertauscht, beim zweiten wiederhergestellt usw. Der Parameter *Cross-Over-Modus* bestimmt die Art des *Cross-Over*.
- **Regel-Spezialisierung**: Der Parameter p bestimmt die Wahrscheinlichkeit, dass eine Regel-Spezialisierungs-*Cross-Over*-Operation für die gesamte Population durchgeführt wird. Das erste

Nachkommengenom erhält dabei die größere untere Grenze und die kleinere obere Grenze von seinen Eltern. Der Parameter *Cross-Over-Modus* bestimmt die Art des *Cross-Over*.

- **Regel-Generalisierung**: Der Parameter p bestimmt die Wahrscheinlichkeit, dass eine Regel-Generalisierungs-*Cross-Over*-Operation für die gesamte Population durchgeführt wird. Das erste Nachkommengenom erhält dabei die kleiner untere Grenze und die größere obere Grenze von seinen Eltern. Der Parameter *Cross-Over-Modus* bestimmt die Art des *Cross-Over*.
- **Mutation Aktivatorgen**: Der Parameter p bestimmt die Wahrscheinlichkeit, dass eine Mutations-Operation für ein Aktivator-Gen durchgeführt wird, wenn die Ativierungsgenregel gewählt worden ist.
- ***Cross-Over* Aktivatorgen**: Der Parameter p bestimmt die Wahrscheinlichkeit, dass eine *Cross-Over*-Operation für ein Aktivator-Gen durchgeführt wird, wenn die Ativierungsgenregel gewählt worden ist.

Zum *Cross-Over*-Typ gehört neben der Angabe, ob eine Zufallsreihenfolge der Individuen gebildet wird die Spezifizierung einer von drei möglichen Vorgehensweisen beim *Cross-Over* (*Cross-Over-Modus)*:

- **Jede Regel**: Die C*ross-Over*-Operation wird für jede einzelne Regel durchgeführt.
- **Pro Regel**: Für jede Regel wird einzeln geprüft, ob die *Cross-Over*-Operation durchgeführt wird.
- **Gesamtes Genom**: Die *Cross-Over*-Operation wird für das gesamte Genom durchgeführt.

C.3.4 Charakterisierung des gier.downhill-Algorithmus

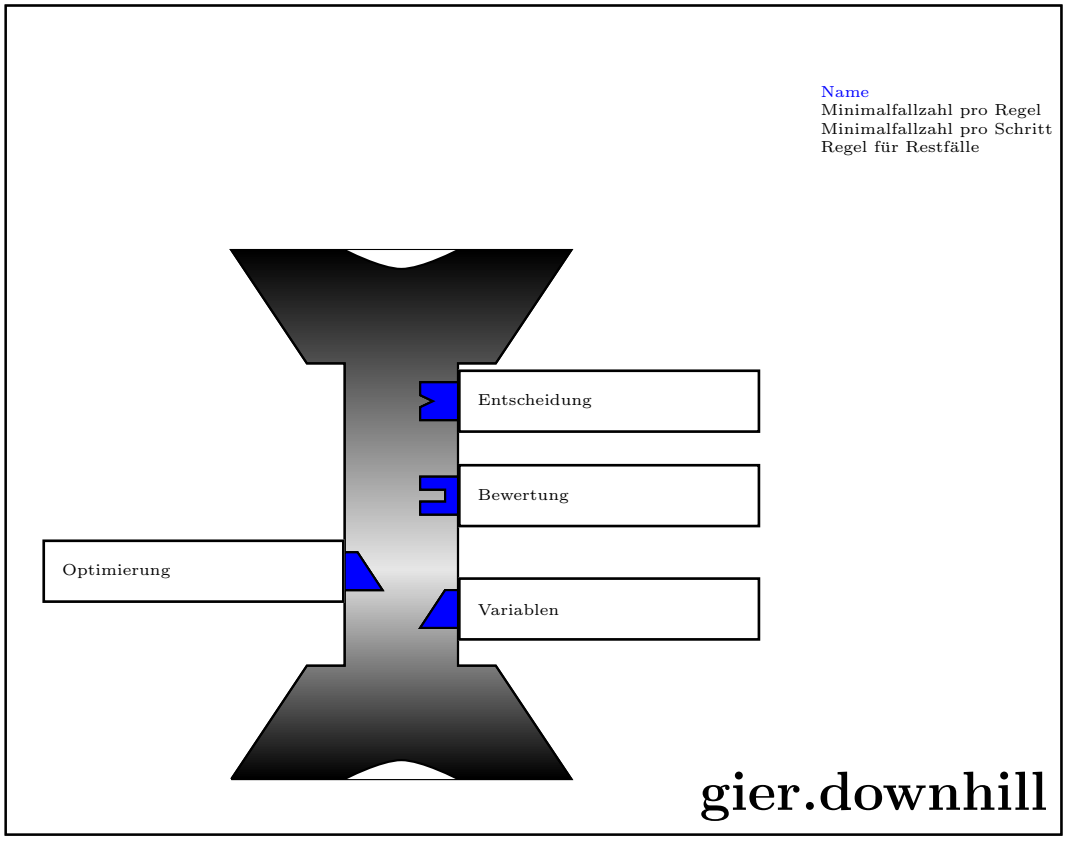


Abbildung C.30: VBk-Darstellung des Algorithmus $\text{GIER}_{[1.0]}$.downhill: Die Abbildung zeigt die abstrakten Platzhalter für variable Komponentenklassen und Parameter für den Gesamtalgorithmus.

Die Umsetzung des Downhill-Algorithmus in $\text{GIER}_{[1.0]}$ist in Abbildung C.30 zusammengefasst. Folgende allgemeine Angaben sind notwendig:

- **Name**: Der Name des Prototypen.
- **Minimalfallzahl pro Schritt**: Dieser Parameter bestimmt die minimale Schrittgröße.

Die Komponente „Bewertung“ wird vom GIER$_{[1.0]}$.ga-Algorithmus übernommen, für die übrigen drei Komponenten sind spezifische Varianten notwendig.

Prädiktionswerte

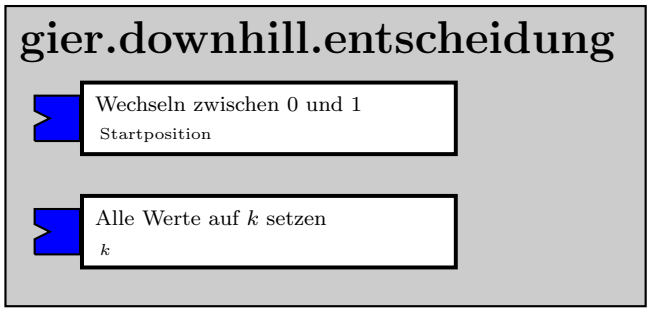


Abbildung C.31: VBk-Darstellung der Komponente „entscheidung“ des Algorithmus GIER$_{[1.0]}$.downhill

Zur Verteilung der Prädiktionswerte auf die Regeln eines Individuums gibt es die in Abbildung C.31 dargestellten beiden Möglichkeiten:

- **Wechseln zwischen 0 und 1**: Es werden abwechselnd die Prädiktionswerte 0 und 1 verteilt, dabei gibt der Parameter *Startposition* die Startposition an.
- **Alle Werte auf k setzen**: Alle Prädiktionswerte werden durch den Wert des Parameters k bestimmt.

Optimierung

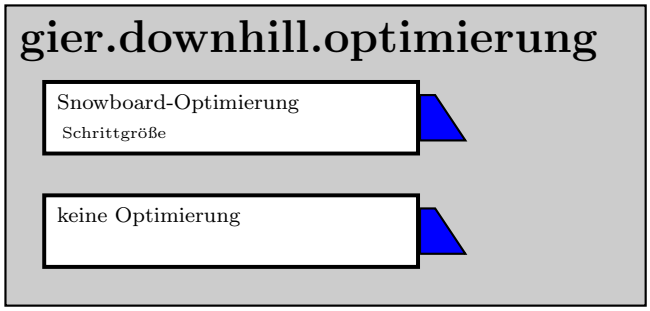


Abbildung C.32: VBk-Darstellung der Komponente „optimierung“ des Algorithmus GIER$_{[1.0]}$.downhill

Für die Komponente GIER$_{[1.0]}$.downhill.optimierung gibt es ebenfalls zwei Varianten (vgl. Abbildung C.32):

- **keine Optimierung**: Es wird keine Optimierung des Ergebnisses durchgeführt.
- **Snowboard-Optimierung**: Es wird eine Snowboard-Optimierung durchgeführt, der Parameter *Schrittgröße* bestimmt die Minimalfallzahl pro Schritt.

Abbildung C.33: VBk-Darstellung der Komponente „variablenreduktion“ des Algorithmus GIER$_{[1.0]}$.downhill

Variablenreduktion

Zur Reduzierung der Variablenzahl gibt es schließlich noch einmal zwei Möglichkeiten (s. Abbildung C.33):

- **keine Reduktion**: Die Variablenzahl wird nicht künstlich reduziert.
- **Modelle vergleichen**: Modelle mit unterschiedlicher Variablenzahl werden verglichen. Die Parameter *Min. Variablenz.* und *Max. Variablenz.* bestimmen das untersuchte Intervall, zudem wird mit dem Parameter *Reihenf. Größe-Var.-Regeln* die Reihenfolge der Kriterien „abgedeckte Stichprobe“, „Anzahl der Variablen“ und „Anzahl der Regeln“ bei der Rangfolgebildung für Ergebnisse mit gleicher Bewertung festgelegt.

Mit diesen Angaben ist der GIER$_{[1.0]}$.downhill-Algorithmus vollständig bestimmt.

C.3.5 Charakterisierung des gier.swarm-Algorithmus

Der GIER$_{[1.0]}$.swarm Algorithmus setzt die Prinzipien der Partikelschwarmoptimierung um. Dabei bewegen sich die Partikel bei k Variablen durch einen $2k$-dimensionalen Suchraum, dessen Koordinaten den unteren und oberen Grenzen der Antezedenzbedingung einer Regel entsprechen. Zudem sind die Partikel von einem bestimmten Typ, der dem Prädiktionswert der Regel entspricht. Durch den Umstand der eingeschränkten Wertebereiche sind einige Modifikationen des Standardalgorithmus notwendig. Die Grundform des Algorithmus entspricht der Darstellung in Abbildung C.34. Von den neun Komponenten sind sechs bereits für den GIER$_{[1.0]}$.ga-Algorithmus diskutiert worden, die übrigen drei werden unten dargestellt. Als generelle Parameter braucht der Algorithmus folgende Werte:

- **Name**: Der Name des Prototypen.
- **Schwarmgröße**: Die Anzahl der Schwarmpartikel.
- **Max. Anfangsgeschwindigkeit**: Die maximale Anfangsgeschwindigkeit für Partikel.

Schwarmtopologie

Für die Erstellung der Kommunikationsstruktur von Schwärmen stehen vier Komponenten bereit (s. Abbildung C.35):

- ***Total Network***: Es wird eine *Total-Network*-Topologie eingesetzt.
- ***Circle***: Es wird eine *Circle*-Topologie eingesetzt.
- ***Wheel***: Es wird eine *Wheel*-Topologie eingesetzt.
- ***Random Connections***: Es wid eine *Random-Connections*-Topologie eingesetzt. Der Wert des Parameters p_{con} gibt die Wahrscheinlichkeit für das Vorhandensein einer Verbindung zwischen zwei beliebigen Partikeln an.

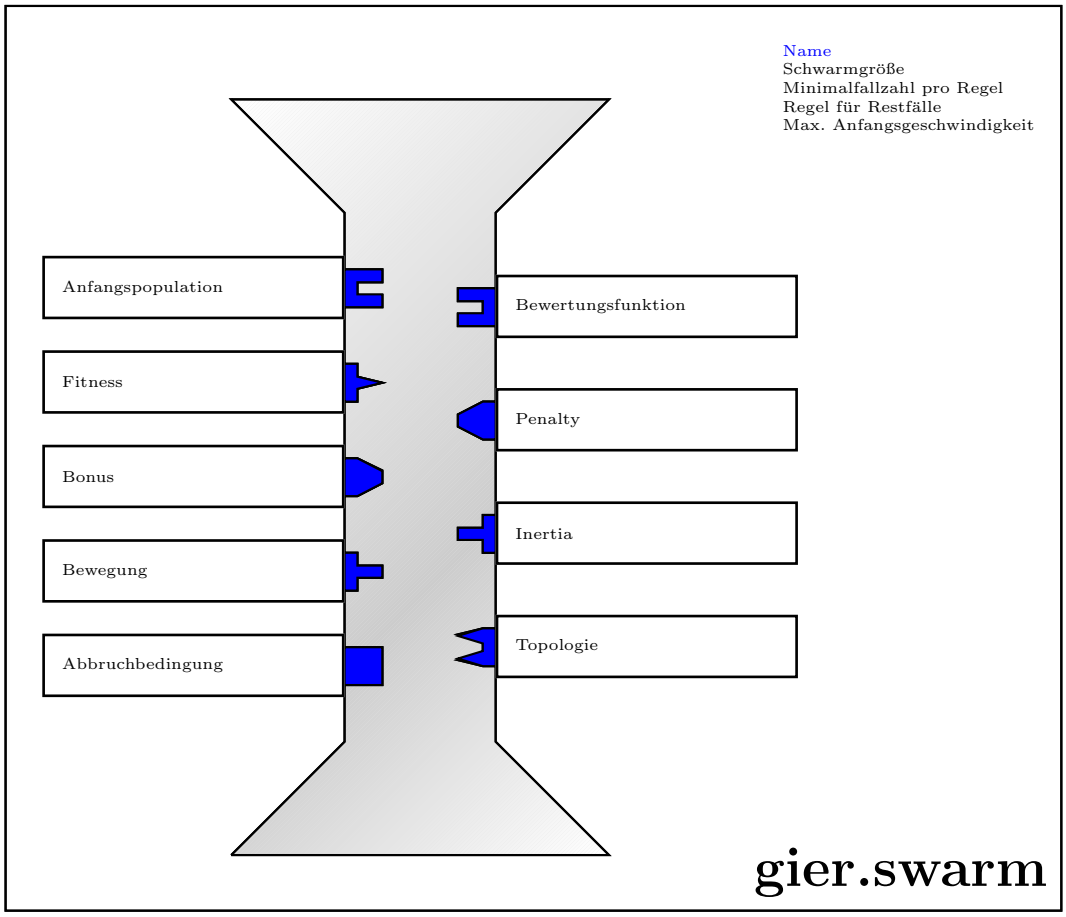


Abbildung C.34: VBk-Darstellung des Algorithmus GIER$_{[1.0]}$.swarm: Die Abbildung zeigt die abstrakten Platzhalter für variable Komponentenklassen und Parameter für den Gesamtalgorithmus.

Schwarmbewegung

Für die Bewegung der Partikel gibt es zwei mögliche Komponenten (vgl. Abbildung C.36):

- ***Random Movement***: Der Partikelschwarm wird mit der üblichen *Random Movement* Regel bewegt. Dabei steht der Parameterwert von ϕ für den Wert der gleichnamigen Schwarmvariablen und V_{max} für die Maximalgeschwindigkeit des Schwarms, während der Parameter *Seitenbegrenzung* angibt, ob die Ränder des erlaubten Wertebereichs für Prädiktoren von Schwarmpartikeln überschritten werden dürfen.

- ***Random Movement (Species)***: In Ergänzung zu den ebenfalls vorhandenen Eigeschaften der *Random Movement*-Variante, werden bei der Bewegungsbestimmung nur Schwarmpartikel mit gleichem Prädiktionswert zugelassen.

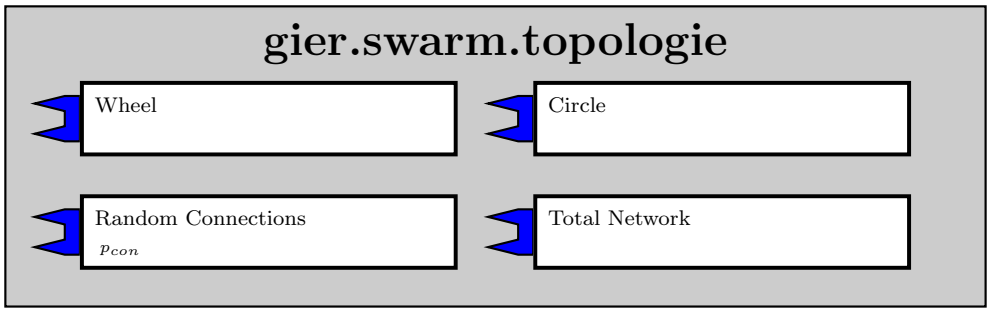


Abbildung C.35: VBk-Darstellung der Komponente „topologie“ des Algorithmus GIER$_{[1.0]}$.swarm

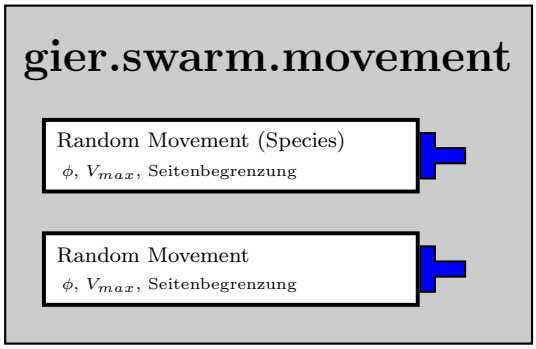


Abbildung C.36: VBk-Darstellung der Komponente „movement“ des Algorithmus GIER$_{[1.0]}$.swarm

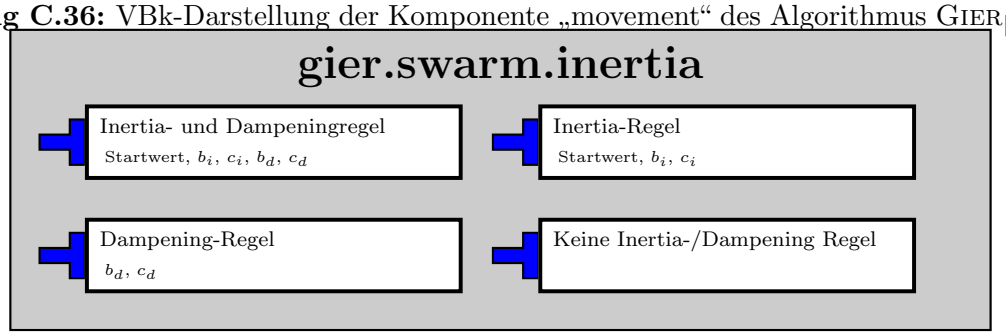


Abbildung C.37: VBk-Darstellung der Komponente „inertia“ des Algorithmus GIER$_{[1.0]}$.swarm

Schwarm-Inertia

Für das Zusammenspiel aus *Inertia*- und *Dampening*-Regeln sind vier Kombinationen möglich (s. Abbildung C.37):

- **Keine *Inertia-/Dampening* Regel**:
- ***Inertia*-Regel**: Es wird die *Inertia*-Regel zur Anwendung gebracht, wobei der Parameter *Startwert* den Anfangswert bestimmt und die Parameter b_i und c_i den Verlauf.
- ***Dampening*-Regel**: Es wird die *Dampening*-regel verwendet mit den Parametern b_d und c_d.
- ***Inertia*- und *Dampening*-Regel**: Es wird zum einen die *Inertia*-Regel zur Anwendung gebracht, wobei der Parameter *Startwert* den Anfangswert bestimmt und die Parameter b_i und c_i den Verlauf. Zum anderen wird die *Dampening*-Regel verwendet mit den Parametern b_d und c_d.

Die übrigen Komponenten werden übernommen. Dabei folgt die Abbruchbedingung beispielsweise dem Sprachgebrauch von Generationen, so dass ein Bewegungszyklus des Schwarms als eine Generation interpretiert wird, um so GIER$_{[1.0]}$.ga.abbruch unverändert übernehmen zu können.

C.3.6 Charakterisierung des gier.swarmset-Algorithmus

Der GIER$_{[1.0]}$.swarmset-Algorithmus steht im gleichen Verhältnis zum GIER$_{[1.0]}$.swarm-Algorithmus wie der GIER$_{[1.0]}$.gaset-Algorithmus zum GIER$_{[1.0]}$.ga-Algorithmus: Der Unterschied findet sich in einer unterschiedlichen Art der Kodierung insofern, dass bei den Set-Varianten ein Individuum eine Regelmenge und keine Einzelregel verkörpert. Für den GIER$_{[1.0]}$.swarmset-Algorithmus bedeutet dies eine Ausweitung der Dimensionalität des Suchraums: Bei m Regeln und k Variablen wird hier jeder Partikel in einem $2mk$-dimensionalen Suchraum verortet, wobei jeweils $2k$ Dimensionen eine einzelne Regel der Menge beschreiben. Dabei ist die Reihenfolge der Bezüge strikt festgelegt.

Die neun Komponenten des Algorithmus (vgl. Abbildung C.38) werden teilweise von GIER$_{[1.0]}$.swarm und teilweise von GIER$_{[1.0]}$.ga, in einem Fall auch von GIER$_{[1.0]}$.gaset wiederverwendet, so dass keine neuen Komponenten genannt werden müssen. Bei der Erstellung der ersten Schwarmstruktur muss allerdings im Gegensatz zur Erstellung durch GIER$_{[1.0]}$.gaset.creation auch die Anfangsgeschwindigkeit der Partikel festgelegt werden, die durch einen allgemeinen Parameter nach oben eingeschränkt wird. Die allgemeinen Parameter des Algorithmus sind der Reihe nach:

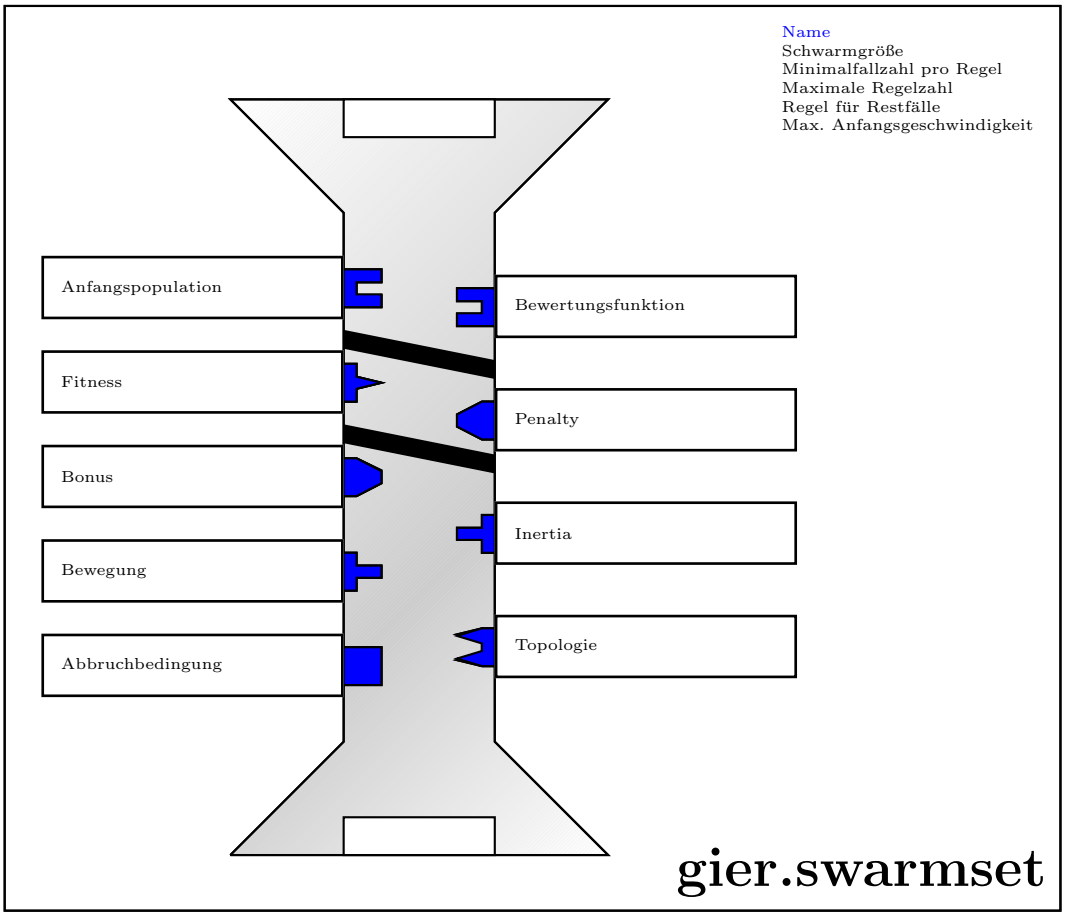


Abbildung C.38: VBk-Darstellung des Algorithmus GIER$_{[1.0]}$.swarmset: Die Abbildung zeigt die abstrakten Platzhalter für variable Komponentenklassen und Parameter für den Gesamtalgorithmus.

- **Name**: Der Name des Prototypen.
- **Schwarmgröße**: Die Anzahl der Schwarmpartikel.
- **Maximale Regelzahl**: Die maximale Anzahl von Regeln legt die Dimensionalität des Suchraums fest.
- **Max. Anfangsgeschwindigkeit**: Die maximale Anfangsgeschwindigkeit für Partikel.

Wie auch beim GIER$_{[1.0]}$.gaset-Algorithmus muss die maximale Regelzahl nicht der Regelzahl für ein Individuum entsprechen, da Regeln im Genom kodiert sein können, die nicht die erforderliche Anzahl von Fällen abdecken.

C.3.7 Charakterisierung des gier.island-Algorithmus

Der GIER$_{[1.0]}$.island-Algorithmus ist ein konsequenter Meta-Algorithmus, da er keine eigenständigen Konzepte zur Generierung von Lösungen mitbringt. Er dient zur Bündelung der Ergebnisse einfacherer Algorithmen. Sein Ablauf entspricht dem vorgestellten Inselalgorithmus, seine Komponentenstruktur ist in Abbildung C.39 festgehalten. Es wurde zudem die Konvention fixiert, dass alle Algorithmen auf den Inseln GIER$_{[1.0]}$.gaset-Agorithmen sein müssen. Diese Konvention ist weniger rigide als sie scheint, da bei der Erstellung der Erstpopulationen andere Algorithmen zum Einsatz kommen können. Die Komponenten zur Bewertung von Lösungen werden von GIER$_{[1.0]}$.ga geerbt, eigene Komponenten gibt es für die Inseltopologie, die Migration und den Epochenabbruch. Die generellen Parameter für den Algorithmus sind:

- **Name**: Der Name des Prototypen.

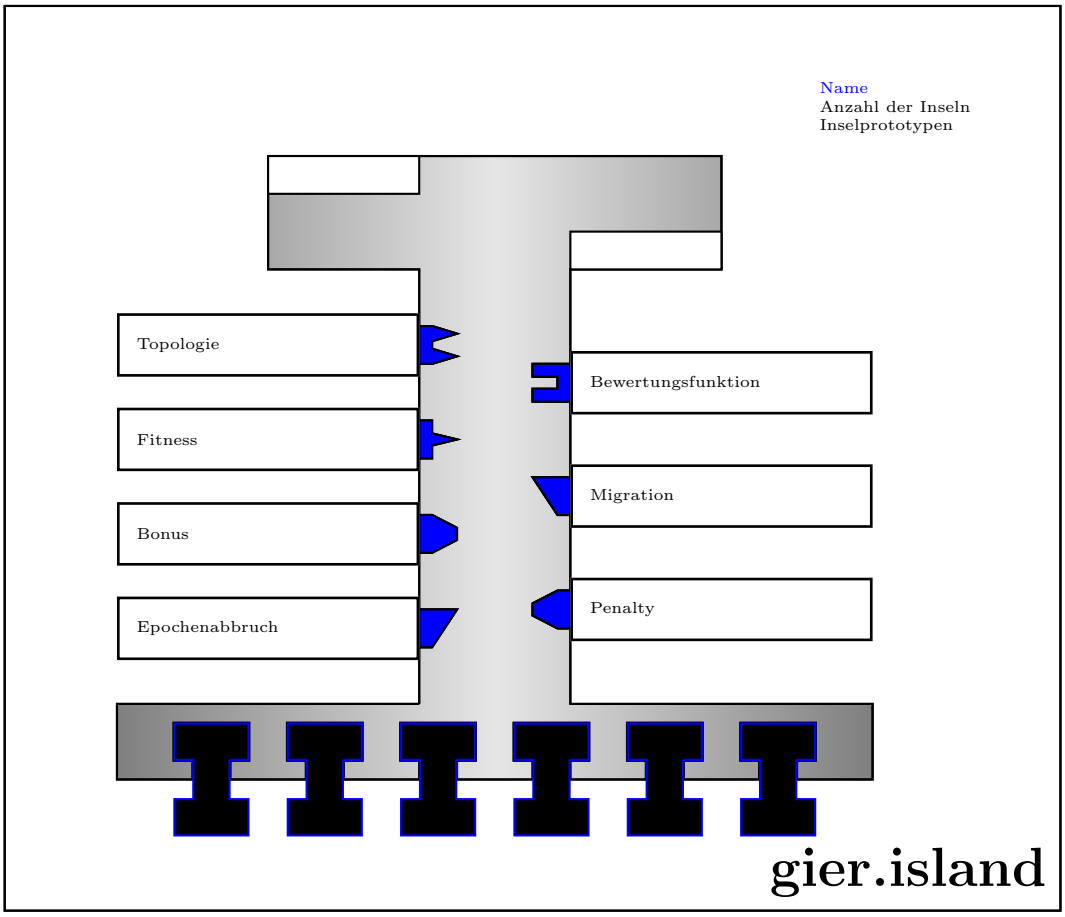


Abbildung C.39: VBk-Darstellung des Algorithmus GIER$_{[1.0]}$.island: Die Abbildung zeigt die abstrakten Platzhalter für variable Komponentenklassen und Parameter für den Gesamtalgorithmus. Die schwarzen Symbole stehen stellvertretend für angeschlossene GIER$_{[1.0]}$.gaset-Algorithmen

- **Inselzahl**: Die Anzahl der Inseln mit Einzelalgorithmen.
- **Inselprototypen**: Die Namen der Prototypen für die einzelnen Inseln legen die eingesetzten Algorithmen fest, wobei der gleiche Prototyp auf mehreren Inseln eingesetzt werden kann, da Prototypen Klassen und nicht Instanzen von Algorithmen darstellen.

Inseltopologie

In Abbildung C.40 sind die Komponenten des Moduls GIER$_{[1.0]}$.island.topologie zusammengestellt. Die Varianten sind im Einzelnen:

- **Ring**: Die Inseln bilden einen Ring. Der Parameter *bidirektional* gibt an, ob die Migration zwischen zwei benachbarten Inseln unidirektional oder in beide Richtungen vorgenommen wird.
- ***Stepping Stones***: Die Inseln bilden eine *Stepping-Stones*-Struktur. Der Parameter *bidirektional* gibt an, ob die Migration zwischen zwei benachbarten Inseln unidirektional oder in beide Richtungen vorgenommen wird.
- **2D-*Grid*** $(4, 9, 16, \ldots)$: Die Inseln bilden ein zweidimensionales Gitter. Zulässige Inselzahlen sind die Quadrate natürlicher Zahlen.
- **Kreise**: Die Inseln bilden eine Kreise-Struktur. Der Parameter *Linienzahl* bestimmt die Anzahl der Kreise. Der Parameter *bidirektional* gibt an, ob die Migration zwischen zwei benachbarten Inseln unidirektional oder in beide Richtungen vorgenommen wird.

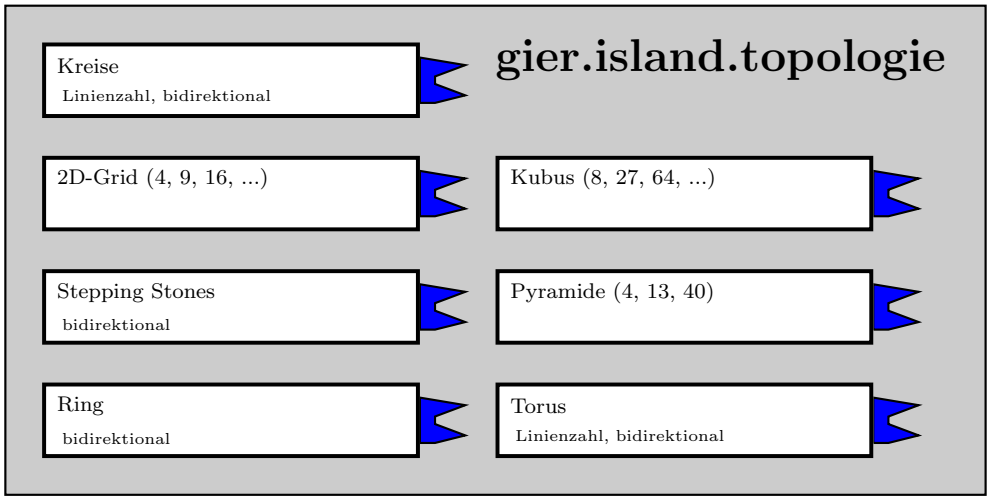


Abbildung C.40: VBk-Darstellung der Komponente „topologie“ des Algorithmus GIER$_{[1.0]}$.island

- **Torus**: Die Inseln bilden eine Torus-Struktur. Der Parameter *Linienzahl* bestimmt die Anzahl der Linien in eine Richtung, die der anderen ergeben sich als Quotient aus Gesamtzahl und *Linienzahl*. Der Parameter *bidirektional* gibt an, ob die Migration zwischen zwei benachbarten Inseln unidirektional oder in beide Richtungen vorgenommen wird.
- **Pyramide (4, 13, 40)**: Die Inseln bilden eine pyramidale Struktur. Zuläsige Inselzahlen sind 4, 13 und 40.
- **Kubus** $(8, 27, 64, \ldots)$: Die Inseln bilden ein zweidimensionales Gitter. Zulässige Inselzahlen sind alle n^3 mit $n \in \mathbb{N}_+$.

Migration

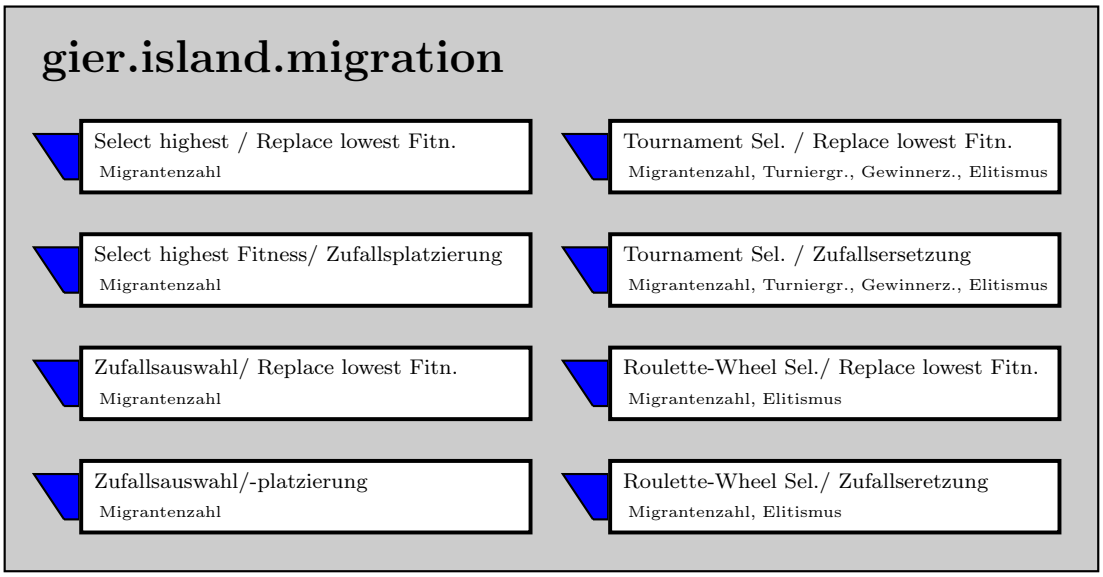


Abbildung C.41: VBk-Darstellung der Komponente „migration“ des Algorithmus GIER$_{[1.0]}$.island

Die Migration besteht aus einer Kombination eines Auswahlmechanismus für Migranten und eines Ersetzungsmechanismus für die Zielpopulationen der Migranten. Dafür stehen insgesamt acht Komponentenvarianten zur Verfügung(s. Abbildung C.41):

- **Zufallsauswahl/-platzierung**: Die Migranten werden zufällig ausgewählt und zufällig ermittelte Individuen in der Zielpopulation werden durch sie ersetzt. Der Parameter *Migrantenzahl* bestimmt die Anzahl der Migranten.

- **Zufallsauswahl/ *Replace lowest Fitn.***: Die Migranten werden zufällig ausgewählt und die Individuen mit der niedrigsten Fitness in der Zielpopulation werden durch sie ersetzt. Der Parameter *Migrantenzahl* bestimmt die Anzahl der Migranten.

- ***Select highest Fitness*/ Zufallsplatzierung**: Es werden die Individuen mit der höchsten Fitness als Migranten ausgewählt. Es werden zufällig ausgewählte Individuen in der Zielpopulation ersetzt. Der Parameter *Migrantenzahl* bestimmt die Anzahl der Migranten.

- ***Select highest / Replace lowest Fitn.***: Es werden die Individuen mit der höchsen Fitness als Migranten ausgewählt. Es werden die Individuen mit der niedrigsten Fitness in der Zielpopulation ersetzt. Der Parameter *Migrantenzahl* bestimmt die Anzahl der Migranten.

- ***Roulette-Wheel-Sel.*/ Zufallsersetzung**: Die Migranten werden durch das Verfahren der *Roulette-Wheel-Selection* bestimmt. Es werden zufällig ausgewählte Individuen in der Zielpopulation ersetzt. Der Parameter *Migrantenzahl* bestimmt die Anzahl der Migranten. Der Wert des Parameters *Elitismus* bestimmt darüber, ob das Individuum mit höchster Fitness mit Sicherheit ausgewählt wird oder wie alle anderen Individuen behandelt wird

- ***Roulette-Wheel-Sel.*/ *Replace lowest Fitn.***: Die Migranten werden durch das Verfahren der *Roulette-Wheel-Selection* bestimmt. Es werden die Individuen mit der niedrigsten Fitness in der Zielpopulation ersetzt. Der Parameter *Migrantenzahl* bestimmt die Anzahl der Migranten. Der Wert des Parameters *Elitismus* bestimmt darüber, ob das Individuum mit höchster Fitness mit Sicherheit ausgewählt wird oder wie alle anderen Individuen behandelt wird.

- ***Tournament Sel.* / Zufallsersetzung**: Die Migranten werden durch das Verfahren der *Tournament Selection* bestimmt. Die Parameter *Turniergr* und *Gewinnerz* bestimmten Turniegröße und Gewinneranzahl pro Turnier. Es werden zufällig ausgewählte Individuen in der Zielpopulation ersetzt. Der Parameter *Migrantenzahl* bestimmt die Anzahl der Migranten. Der Wert des Parameters *Elitismus* bestimmt darüber, ob das Individuum mit höchster Fitness mit Sicherheit ausgewählt wird oder wie alle anderen Individuen behandelt wird

- ***Tournament Sel.* / *Replace lowest Fitn.***: Die Migranten werden durch das Verfahren der *Tournament Selection* bestimmt. Die Parameter *Turniergr.* und *Gewinnerz.* bestimmen Turniergröße und Gewinneranzahl pro Turnier. Es werden die Individuen mit der niedrigsten Fitness in der Zielpopulation ersetzt. Der Parameter *Migrantenzahl* bestimmt die Anzahl der Migranten. Der Wert des Parameters *Elitismus* bestimmt darüber, ob das Individuum mit höchster Fitness mit Sicherheit ausgewählt wird oder wie alle anderen Individuen behandelt wird

Epochenabbruch

Die Komponenten zur Bestimmung des Abbruchs des Inselalgorithmus (s. Abbildung C.42) ähneln den GIER$_{[1.0]}$.ga.abbruch-Komponenten. . Die zur Auswahl stehenden Komponenten sind:

- **Anzahl Epochen**: Der Inselalgorithmus wird nach einer festen Anzahl von Epochen beendet, die durch den Parameter *Epochenzahl* festgelegt wird.

- **Minimale Klassifikationsgenauigkeit**: Der Inselalgorithmus wird nach Erreichen eines Mindestklasifikationsgenauigkeitswerts auf einer beliebigen Insel beendet, der durch den Parameter *Mindestwert* angegeben wird. Dabei gibt es eine untere und eine obere Grenze für die Laufzeit in Epochen, die durch die beiden Parameter *Mindestzahl* und *Maximalzahl* spezifiziert werden.

- **Minimaler Fitnesswert**: Der Inselalgorithmus wird nach Erreichen eines Mindestfitnesswerts auf einer beliebigen Insel beendet, der durch den Parameter *Mindestwert* angegeben wird. Dabei gibt es eine untere und eine obere Grenze für die Laufzeit in Epochen, die durch die beiden Parameter *Mindestzahl* und *Maximalzahl* spezifiziert werden.

- **Klassifikationsverbesserung**: Der Inselalgorithmus wird beendet, nachdem für eine durch den Parameter *Anz. Ep.* bestimmte Anzahl von Epochen, die Verbesserung der besten Klassifikationsgenauigkeit in den Populationen aller Inseln einen durch den Parameter *Grenze* angegebenen Wert

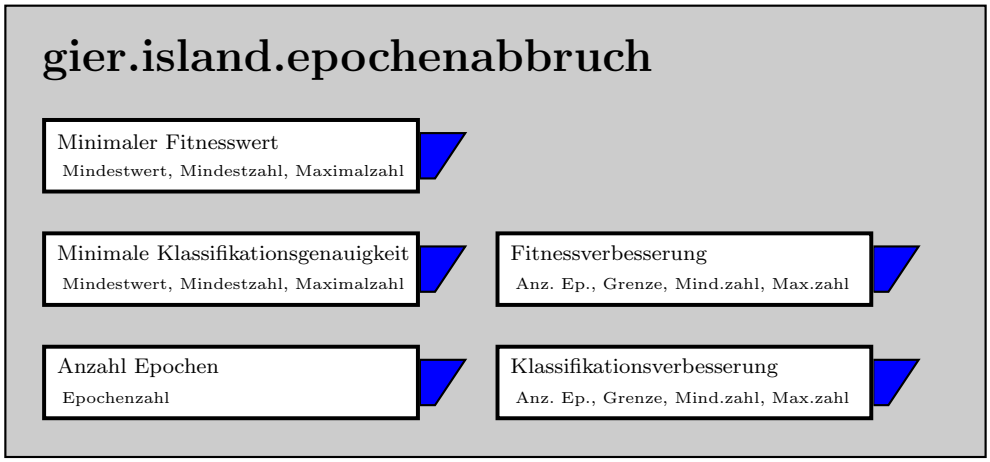


Abbildung C.42: VBk-Darstellung der Komponente „epochenabbruch“ des Algorithmus GIER$_{[1.0]}$.island

nicht überschreitet. Dabei gibt es zudem eine untere und eine obere Grenze für die Laufzeit in Epochen, die durch die beiden Parameter *Mindestzahl* und *Maximalzahl* spezifiziert werden.

- **Fitnessverbesserung**: Der Inselalgorithmus wird beendet, nachdem für eine durch den Parameter *Anz. Ep.* bestimmte Anzahl von Epochen, die Verbesserung des besten Fitnesswerts in den Populationen aller Inseln einen durch den Parameter *Grenze* angegebenen Wert nicht überschreitet. Dabei gibt es zudem eine untere und eine obere Grenze für die Laufzeit in Epochen, die durch die beiden Parameter *Mindestzahl* und *Maximalzahl* spezifiziert werden.

C.3.8 Die Verschachtelung von Algorithmen

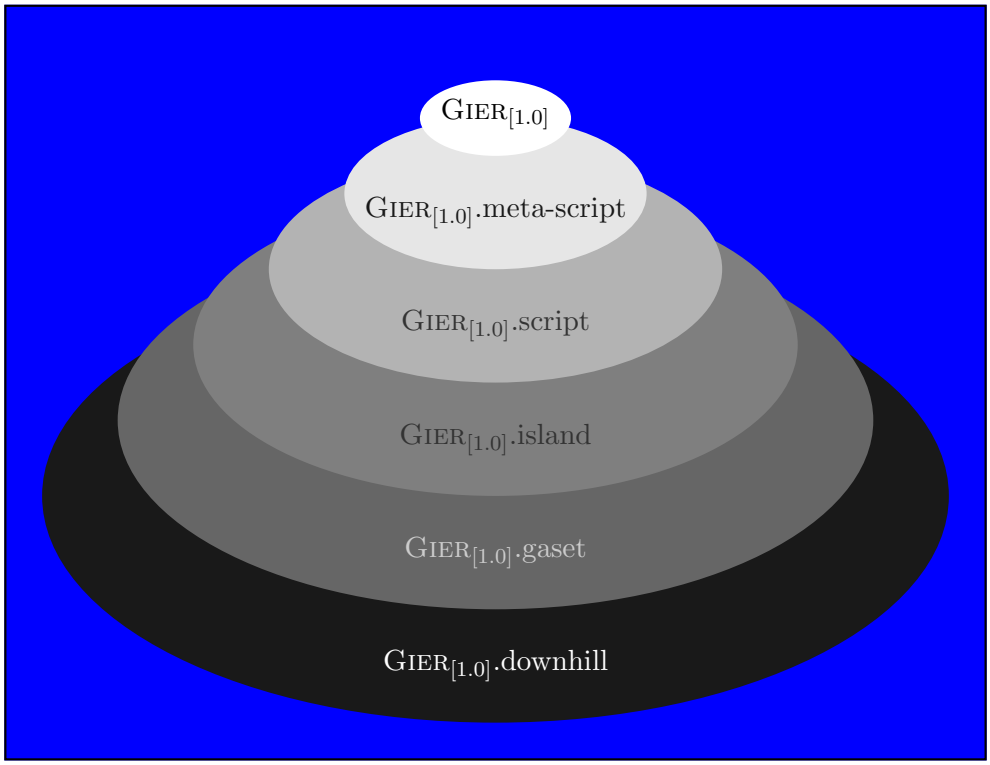


Abbildung C.43: Klassenverschachtelung bei Aufruf eines GIER$_{[1.0]}$.island-Algorithmus aus einem Skript, das Teil eines Meta-Skripts ist, wobei auf einer Insel des GIER$_{[1.0]}$.island-Algorithmus ein GIER$_{[1.0]}$.gaset-Algorithmus einen GIER$_{[1.0]}$.downhill-Algorithmus aufruft, um die Erstpopulation zu erstellen.

Bei der Verwendung von GIER$_{[1.0]}$.island-Algorithmen kann das höchste Ausmaß der Verschachtelung

erreicht werden. Abildung C.43 zeigt ein erreichbares Niveau. Verschachtelung wird im Rahmen der GIER$_{[1.0]}$-Toolbox auf verschiedene Arten erreicht:

- Die Software ermöglicht die Ausführung von Meta-Scripten: Die Rahmensoftware stellt die Möglichkeiten bereit, Meta-Scripte auszuführen.
- Meta-Scripte rufen Scripte auf: Scripte bilden dabei einen Teil der abzuarbeitenden Befehle.
- Scripte erzeugen Algorithmen: Die Algorithmen bilden dabei eine durch das Script erzeugte Struktur, die zur Umsetzung des Berechnungsauftrages dient.
- Meta-Algorithmen verwalten Algorithmen: Algorithmen werden koordiniert und Ergebnisse werden zusammengeführt, um lokale Minima zu bekämpfen.
- Algorithmen rufen einfachere Algorithmen auf: Einfache Algorithmen werden genutzt, um Daten vorzubearbeiten, damit zeitaufwändigere Algorithmen, die breiter suchen, auf dieser Grundlage beginnen können. Eine Gefahr dabei ist die mögliche Vergrößerung der Gefahr des Erreichens lokaler Minima durch die Vorverarbeitung, der Vorteil ist die unter Umständen essenzielle Reduktion des Zeitbedarfs.
- Algorithmen verwalten ihre Komponenten: Die Routinen der Komponentenklassen werden von den Algorithmen angesteuert und ausgeführt.
- Komponenten verwalten ihre Methoden: Die Methoden der Komponenten werden durch Hauptroutinen der Komponenten verwaltet und ausgeführt.
- Methoden führen Operationen aus: Die Methoden der Komponenten enthalten ausführbare Anweisungen. Diese Ebene stellt die unterste Ebene der Programmierung dar, darunter liegen noch maschinennahe Kodierungsformen.

Die Komponentenebene und die Möglichkeit, Algorithmen zu nutzen, die andere Algorithmen nutzen, könnten die in der Abbildung gezeigte Struktur weiter ausbauen, doch das Prinzip der Rekombination auch auf der Ebene programmierter Klassen sollte mit diesen Beispielen deutlich geworden sein.

C.3.9 Visualisierungen

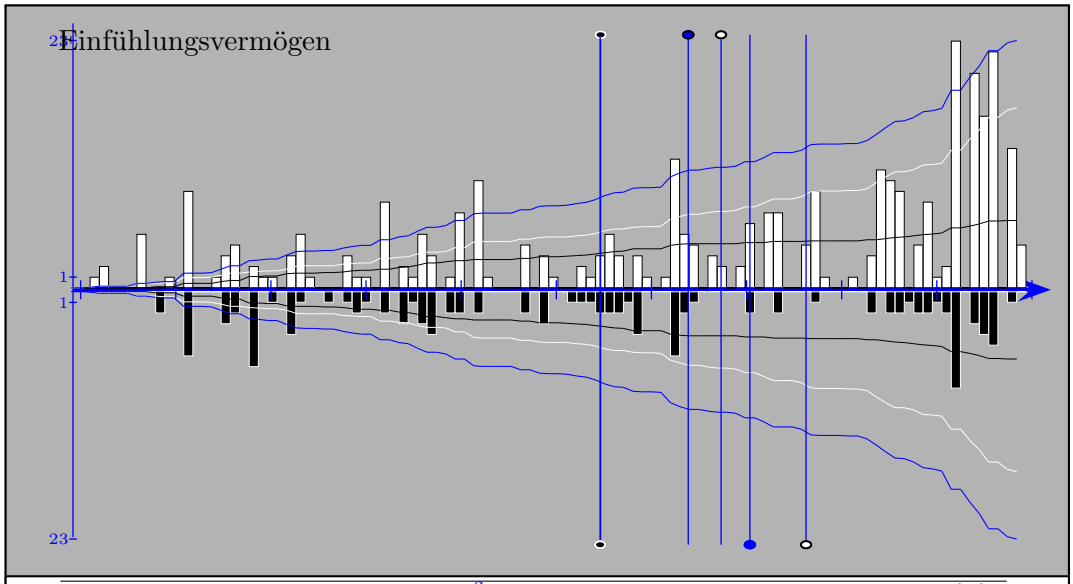
Die Software unterstützt mehrere Möglichkeiten, Regeln, Datenverteilungen und Ergebnisse zu visualisieren. Unter anderem sind folgende Möglichkeiten der Visualisierung umgesetzt worden:

- Zweidimensionale und dreidimensionale Streudiagramme für Variablen mit Anzeige der Klassen
- Schattierungsdiagramme für die zusammenfassende Darstellung von Datenverteilungen
- Matrixanordnungen von Streudiagrammen für dne paarweisen Vergleich einer Variablenmenge
- Liniendiagramme für die Verbesserung von Lösungen über Epochen und Generationen
- Farbschattierungsdiagramme für eine nach Klassen geordnete oder ungeordnete Visualisierung aller Daten einer Stichprobe
- Doppelhistogramme zur Charakterisierung von Prädiktoren mit mehreren statistischen Prüfgrößen (s.u.)

Weiterhin gibt es zwei Möglichkeiten zur Visualisierung von Ergebnisregelmengen. Zum einen kann in einer Kombination aus Balken- und Liniendiagramm jede einzelne Regel als Balkendiagramm visualisiert werden, wobei Fälle als Linienverlauf darüber gelegt werden können. Damit wird die Analyse typischer Fehler möglich, indem Gruppen von Abweichungen identifiziert werden.

Zum anderen lässt sich jede hierarchische Regelmenge - in einem relativ aufwändigen Algorithmus - auch als Entscheidungsbaum ausgeben. Dabei stehen verschiedene Form der Darstellungsoptimierung zur Verfügung, um entweder Knotenzahl, Endknotenzahl oder Baumtiefe zu minimieren. Dabei werden verschiedene Permutationen der Abfrage geprüft und unnötige Äste werden dabei in jedem Fall entfernt. Es

entsteht dabei ein Entscheidungsbaum, dessen Entscheidungen mit denen der hierarchischen Regelmenge äquivalent ist. Die Knoten können visuell oder als Text um knotenspezifische Klassenverteilungsinformationen oder Entscheidungsfehleranteile ergänzt werden.



Klasse	n	$\bar{x}$	σ	σ^2	Md_x	R_x	t	U	$z(U)$
1:	318	67,915	28,705	823,996	77,0	74473,5	4,134	23752,500	4,114
0:	119	55,050	29,634	878,184	55,0	21229,5			
alle:	437	64,412	29,490	869,642	71,0	95703,0			

Abbildung C.44: Beispiel für ein Doppelhistogramm in GIER$_{[1.0]}$: Für den Datensatz *Beratungsunternehmen* ist der Prädiktor *Einfühlungsvermögen* als Histogramm dargestellt: Eingetragen ist die Verteilung der Prädiktorwerte für Objekt der positiven Klasse (oben) und Objekte der negativen Klasse (unten), sowie Mittelwert (Kreis oben) und Median (Kreis unten) für die Klassen zusammengenommen (blauer Kreis) und einzeln (weißer und schwarzer Kreis). Zusätzlich sind kumulative Verteilungsfunktionen als Linie für die Klassen getrennt (weiße und schwarze Linie) und zusammengefasst (blaue Linie) eingezeichnet.

In Abbildung C.44 ist ein beispielhaftes Doppelhistogramm abgebildet, das durch zahlreiche Parameter nach Belieben angepasst werden kann. Die Darstellung ermöglicht in der Analyse eine genauere Vorstellung davon aufzubauen, was es heißt, einen Prädiktor für die Klassendiskriminierung als geeignet zu betrachten. Mittelwerts- und Medianunterschiede können in ihrer Bedeutung an der Gesamtverteilung ohne Informationsverlust relativiert werden. Eine Reihe von zusätzlichen Statistiken kann weiterhin jeden Prädiktor deskriptiv zuammenfassen, so dass die für die Aufgabenstellung relevanten Eigenschaften der - hier isoliert betrachteten - Variablen deutlich zu Tage treten können.

Insgesamt steht damit ein unterstützender visueller Werkzeugkasten zur Exploration von Daten und Ergebnissen zur Verfügung, der des weiteren aufgrund der Struktur der Programmierung leicht um weitere Komponenten ergänzt werden könnte. Auf weitere visuelle Demonstrationen soll an dieser Stelle verzichtet werden.

C.4 Beschreibung zusätzlicher Komponenten

In diesem Abschnitt sollen einige Erweiterungen der Werkzeugkiste in GIER$_{[1.0]}$vogestellt werden, die zu einem großen Teil zu pädagogischen Demosntrationszwecken aufgenommen worden sind und im Hauptteil der Arbeit deshalb keine Rolle spielen.

C.4.1 Charakterisierung des gier.tsp-Algorithmus

Der GIER$_{[1.0]}$.tsp-Algorithmus ist ein einacher genetischer Algorithmus, der als Zielaufgabe nicht die Klassifikation von Objekten, sondern die Wegminimierung durch Reihenfolgebildung von Wegpunkten hat. Als Eingangsdaten werden räumliche Koordinaten von Objekten, oder anschaulicher von Städten benötigt, die eine beliebige Dimensionalität haben können. Die Koordinaten müssen eine gültige GRundlage für

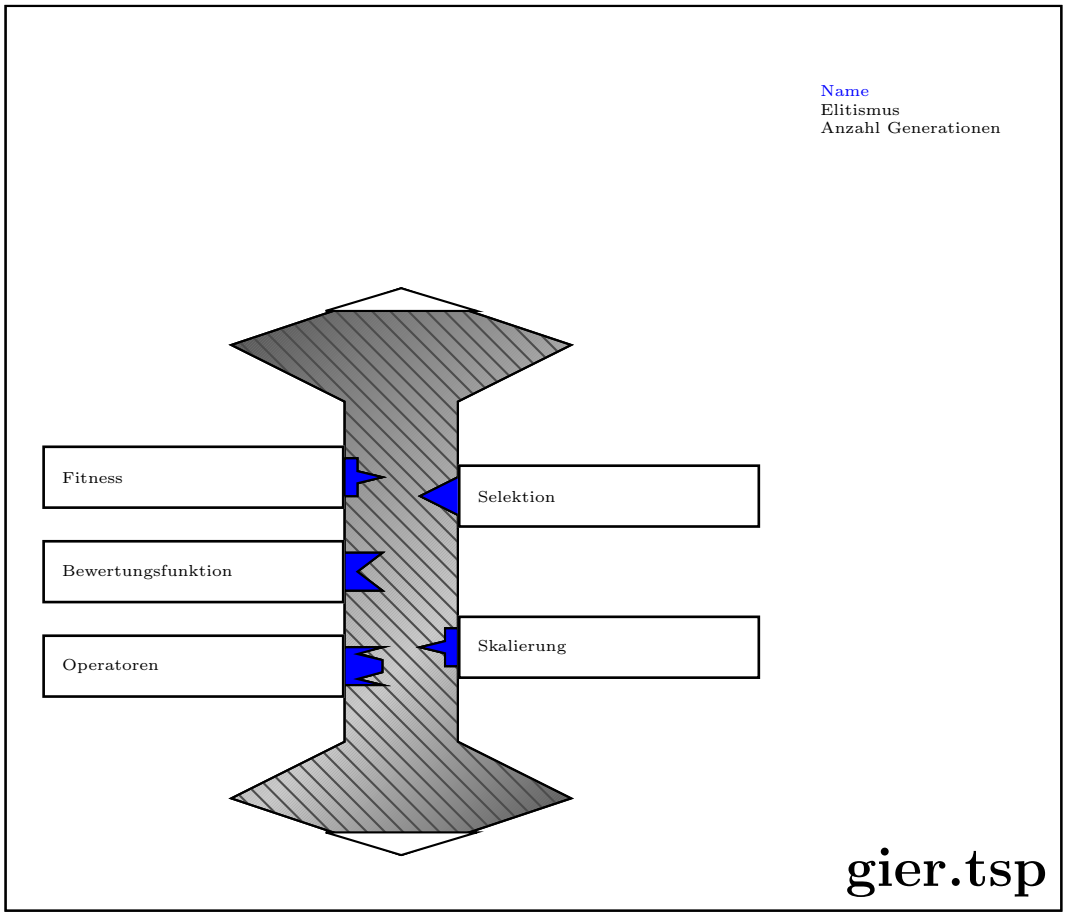


Abbildung C.45: VBk-Darstellung des Algorithmus GIER$_{[1.0]}$.tsp: Die Abbildung zeigt die abstrakten Platzhalter für variable Komponentenklassen und Parameter für den Gesamtalgorithmus.

die Etablierung einer Distanzmetrik darstellen. Da das erklärte Ziel der Entwicklung dieses Algorithmus lediglich die Möglichkeit darstellt, Erfahrungen mit einem genetischen Algorithmus außerhalb von Klassifizierungsproblemen zu ermöglichen, wurde lediglich eine basale Form umgesetzt (vgl. Abbildung C.45). Dementsprechend gibt es keine Alternativen für die Komponenten „anfangspopulation“ und „abbruch“, es gibt keine „bonus“- und keine „penalty“- oder „reparatur“-Komponenten.

Die Anfangspopulation wird zufällig erzeugt, wobei nur gültige Permutationen kreiert werden können. Der Abbruch erfolgt nach einer festgelegten Zahl von Generationen. Auf Mechanismen für den Umgang mit ungültigen Lösungen kann verzichtet werden, weil nur Operatoren zugelassen werden, die aus gültigen Lösungen keine ungültigen produzieren. Für die „fitness“-, „selektion“- und die „skalierung“-Komponenten werden die Module von GIER$_{[1.0]}$.ga wiederverwendet. Als allgemeine Parameter verbleiben:

- **Name**: Der Name des Prototypen.
- **Anzahl Generationen**: Die Anzahl der Generationen, die der Algorithmus vor dem Abbruch durchläuft.
- **Elitismus**: Legt die Verwendung der Elitismus-Regel fest.

Bewertungsfunktion

Zur Implementierung von Bewertungsfunktionen stehen die in Abbildung C.46 dargestellten drei Komponenten zur Verfügung:

- **Euklidische Distanz**: Als Zielkriterium wird die Minimierung der Summe der euklidischen Distanzen zwischen den Städten festgelegt, die auf dem Weg von Anfangspunkt zu Anfangspunkt überwunden werden müssen.

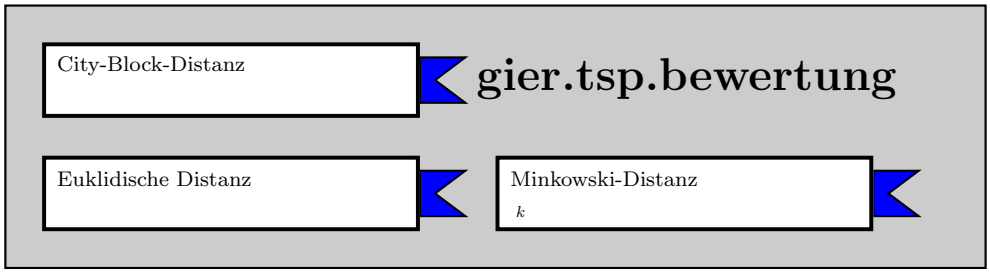


Abbildung C.46: VBk-Darstellung der Komponente „bewertung“ des Algorithmus $\text{GIER}_{[1.0]}$.tsp

- ***City-Block*-Distanz**: Als Zielkriterium wird die Minimierung der Summe der *City-Block*-Distanzen zwischen den Städten festgelegt, die auf dem Weg von Anfangspunkt zu Anfangspunkt überwunden werden müssen.
- **Minkowski-Distanz**: Als Zielkriterium wird die Minimierung der Summe der Distanzen zwischen den Städten festgelegt, die auf dem Weg von Anfangspunkt zu Anfangspunkt überwunden werden müssen. Das Distanzmaß wird durch die Minkowskimetrik mit dem Parameter k gebildet.

Operatoren

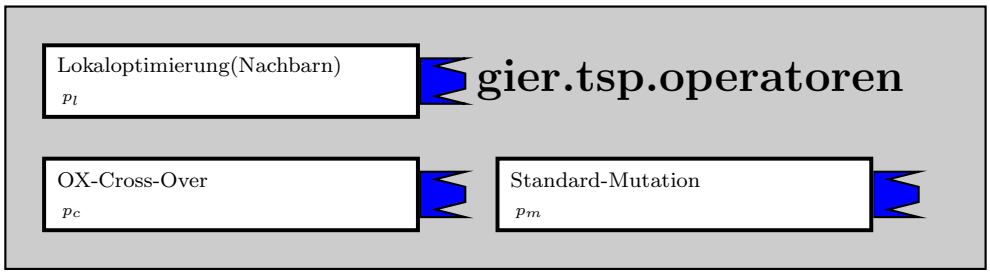


Abbildung C.47: VBk-Darstellung der genetischen Operatoren des Algorithmus $\text{GIER}_{[1.0]}$.tsp: Es kann eine beliebige Anzahl von Operatoren ausgewählt werden auch bei Mehrfachverwendung einzelner Operatoren.

Die Operatoren müssen aufgrund fehlender Reparaturmechanismen die Eigenschaft erfüllen, keine ungültigen Lösungen auszugeben, wenn gültige Lösungen eingegeben werden. Dies trifft für die in Abbldung C.47 gezeigte Auswahl von Operatoren zu:

- **Standard-Mutation**: Die Standard-Mutation bei Permutationen besteht aus dem Tausch zweier zufällig ausgewählter Elemente. Der Parameter p_m gibt die Wahrscheinlichkeit für eine Mutation im Genom an.
- ***OX-Cross-Over***: Der *OX-Cross-Over*-Operator wird mit einer dem Parameter p_c entsprechenden Wahrscheinlichkeit auf die gesamte Population angewendet.
- **Lokaloptimierung(Nachbarn)**: Ein Lokaloptimierungs-Operator wird mit der Wahrscheinlichkeit p_l auf ein Genom angewendet: Der Operator überprüft, ob die Distanz einer Wegstrecke durch Austausch zweier benachbarter Positionen reduziert werden kann damit werden einfache Kreuzungen von Wegstrecken aus der Lösung entfernt, die die Eigenschaft haben, eine Schleife mit 2 Elementen abzutrennen.

Der so beschriebene TSP-Algorithmus kann als Prototyp gespeichert und aufgerufen werden. Er ist zudem so in die Software integriert worden, dass er in die TSP-Demonstration integriert werden kann.

C.4.2 Die TSP-Demonstration

Die TSP-Demonstration hat ein didaktisches Ziel: Der hohe Abstraktionsgrad einiger Algorithmenklassen sollte durch eine Visualisierung von Grundprinzipien erfahrbar gemacht werden. Einige Anwendungen der

Demonstration bei Präsentation haben die Sinnhaftigkeit einer derartigen Ergänzung bestätigt. In der TSP-Demonstration kommen drei Klassen von Algorithmen zum Einsatz:

1. GIER$_{[1.0]}$.tsp-Algorithmen: Es kann ein beliebiger Prototyp des GIER$_{[1.0]}$.tsp-Algorithmus in die Demonstration eingebunden werden.

2. GIER$_{[1.0]}$.*ACO*-Algorithmen: Diese Algorithmenklasse ist nicht-modular in die Software integriert worden und wird unten beschrieben.

3. Ein visueller, interaktiver Algorithmus: Dieser Algorithmus ist nicht in der Software implementiert, sondern erfordert die visuelle Einschätzung eines Menschen, der eine seiner Meinung nach optimale Permutation über ein Interface eingeben kann.

Diese drei Algorithmenklassen sollen in der Demonstration miteinander konkurrieren.

Charakterisierung der Gier$_{[1.0]}$.ACO-Komponente

Die GIER$_{[1.0]}$.ACO-Algorithmen setzen Ameisenalgorithmen (Cordón, Herrera & Stützle, 2002; Dorigo & Di Caro, 1999; Guntsch, 2004; Merkle, 2002) um, die den - etwa beim Nestbau von Termiten zu beobachtenden - *Stigmergy*-Effekt (Guntsch, 2004) nutzen durch Simulation von Ameisenkolonien mit biologisch plausibler Pheromonausschüttungen (Beckers, Deneubourg & Goss, 1992; Goss, Aron, Deneubourg & J.Pasteels, 1989). Erfolgsversprechende Wege werden verstärkt und erhöhen die Wahrscheinlichkeit der Wiederbenutzung durch nachfolgende Ameisen. Ameisenalgorithmen sind bereits erfolgreich angewendet worden auf Probleme der Bildanalyse (Broggi, Cellario, Lombardi & Porta, 2003), zur Analyse chemischer Prozessdaten (Shelokar, Jayaraman & Kulkarni, 2004), zum *Web Mining* (Holden & Freitas, 2004), zur medizinischen Regelinduktion (Parpinelli, Lopes & Freitas, 2001), für Probleme des *Flowshop Scheduling* (Rajendran & Ziegler, 2005) und der Instandhaltung (Samrout, Yalaoui, Châtelet & Chebbo, 2005), auf das Platzieren von Maschinen (Solimanpur, Vrat & Shankar, 2005) und die Entwicklung von Wirkstoffen durch das Screening chemischer Datenbanken (Solmajer & Zupan, 2004). Hier wird der ACO-Algorithmus auf das TSP-Problem angewendet (vgl. Dorigo & Gambardella, 1997; Mazzeo & Loiseau, 2004). Folgende Parameter können durch den Anwender eingegeben werden:

- **Anzahl Ameisen**: Die Anzahl der verwendeten Ameisen.
- **Anzahl Läufe**: Die Anzahl der verwendeten Läufe.
- α **und** β: Duch die beiden Parameter α und β wird das Explorationsverhalten der Ameisen bestimmt, wobei α den Einfluss der Pheromonverstärkung steuert und β den Einfluss der Städtedistanzen.
- **Pheromonstartwert**: Bestimmung des Parameters *Pheromonstartwert*.
- **Evaporationsfaktor**: Bestimmung des Parameters *Evaporationsfaktor*.
- **Pheromonverstärkung**: Bestimmung des Parameters *Pheromonverstärkung*.
- **konstante Verstärkung**: Wahl der Option *konstante Verstärkung*.
- **Alle Ameisen verstärken**: Wahl der Option *Alle Ameisen verstärken*.
- **Lokaloptimierung**: Wendet den für den GIER$_{[1.0]}$.tsp-Algorithmus beschriebenen Lokaloptimierungsoperator an.

Wird die Variante der populationsbasierten ACO (PACO) verwendet, gibt es folgende zusätzliche Einstellungen:

- **Pheromonmenge**: Bestimmung des Parameters *Pheromonmenge*.
- **Größe Markierpopulation**: Bestimmung der Größe der Markierpopulation.

- ▷ ***Replacement***: Bestimmung des Ersetzungsverfahrens. Es besteht die Wahl zwischen den Optionen *Replace Worst*, *Replace Oldest* und *Replace Random*, wobei entweder die schlechtesten, die ältesten oder zufällig ausgewählte Ameisen ersetzt werden.

Der Algorithmus wird direkt aus dem Einstellungsmenü heraus gestartet, da er nicht außerhalb der Demonstration verwendet wird, weshalb eine Prototypendatei unnötig ist.

Beschreibung des Demonstrationsablaufs

Die Demonstration ermöglicht den Leistungsvergleich zwischen menschlicher visueller Wahrnehmung und Planung und zwei Algorithmenklassen. Dazu können menschliche „Kontrahenten“ durch sukzessive button-gestützte Eingabe Lösungen für TSP-Probleme erstellen, die mit den Ergebnissen automatischer Heuristiken verglichen werden. Zunächst muss deshalb ein geeigneter Datensatz geladen werden, der ein zweidimensionales TSP-Problem als zeilenweise aufgeführte Städtekoordinaten liefert. Dann wird ein beliebiger $\text{GIER}_{[1.0]}$.tsp-Protoyp als Konkurrent gewählt. Aus dem Menü heraus können nun wahlweise Läufe des $\text{GIER}_{[1.0]}$.tsp-Algorithmus gestartet werden, oder Läufe eines $\text{GIER}_{[1.0]}$.ACO-Algorithmus bzw. eines hybriden Algorithmus, der zunächst einen $\text{GIER}_{[1.0]}$.tsp-Algorithmus startet und das Ergebnis durch einen $\text{GIER}_{[1.0]}$.ACO-Algorithmus weiterverarbeiten lässt.

Die Ausführung der Algorithmen wird visuell in einem Fenster begleitet durch Angabe der generationsweise aktualisierten Fitness-Statistiken und Anzeige der aktuell besten Lösung für den $\text{GIER}_{[1.0]}$.tsp-Algorithmus sowie eine Anzeige der Pheromonaktivität für den $\text{GIER}_{[1.0]}$.ACO-Algoithmus. Die Ergebnisse der Algorithmen werden in einer Rangliste nach Ergebnis sortiert und mit dem Namen des Algorithmus vermerkt. Zudem ist es möglich, manuelle Lösungen unter dem Namen des Bearbeiters einzugeben, die ebenfalls in der Rangliste vermerkt werden, so dass eine gemeinsame Rangliste aller Ansätze entstehen kann.

Das didaktische Ziel der Demonstration ist zunächst die Vertrautmachung mit der sprunghaften Operationsweise der Heuristiken sowie ihres Zeitverhaltens. Schnellen anfänglichen Fortschritten folgen lange Phasen ohne Lösungsverbesserung. Zudem ist der ACO-Algorithmus im Allgemeinen unvorhersehbarer als der genetische Algorithmus. Bei einfachen Aufgaben mit einer kleinen Zahl an Städten sollten Algorithmen und Menschen die optimale Lösung erreichen, bei Aufgaben mittlerer Schwierigkeit sind im Allgemeinen die menschlichen Lösungen überlegen, was die Leistungsstärke des menschlichen visuellen Systems unterstreicht, obwohl nicht immer die optimale Lösung gefunden wird. Bei Aufgaben hoher Schwierigkeit können sich die Algorithmen durchsetzen, da die Aufgabenstellungen für Menschen nicht mehr dem im ökologischen Sinne natürlichen Rahmen entsprechen.

C.4.3 Darstellung von $\text{Gier}_{[1.0]}$.typag

Der $\text{GIER}_{[1.0]}$.typag-Algorithmus ist eine Umsetzung eines Teilschritts der von Hollmann (1991) entwickelten TYPAG-Algorithmus. Dieser Algorithmus ist ein konfigurales Verfahren zur Diskrimination zwischen zwei Gruppen, also damit auch ein binäres Klassifikationsverfahren und ist ein Akronym für Typ-Agglutination, inspiriert durch die Verfahren zur Hypothesenagglutination (Künzel & Wottawa, 1985; Wottawa, 1987a, 1987b). HYPAG besteht aus einer Sammlung explorativer Verfahren zur Suche nach Variablenzusammenhängen, für TYPAG direkt relevant ist ein Unterprogramm des Programms HYPAG/Search, das von Hollmann (1991) als „heuristischer Ansatz“ bezeichnet wird: Im Programmteil HYPAG/Search-Disc werden multiple Zusammenhänge „im Sinne der Regressions- oder Diskriminanzanalyse“ aufgedeckt. Da auch Prädiktoren mit wenigen Stufen untersucht werden können, liegt der Vergleich mit hypothesenprüfenden Erweiterungen der Konfigurationsfrequenzanalyse und log-linearen Modellen nahe (vgl. Krauth, 1980).

Der TYPAG-Algorithmus

Der ursprüngliche TYPAG-Algorithmus geht so vor, dass alle Teilmengen vorgegebener Größe aus der Menge der Prädiktoren daraufhin untersucht werden, wie gut durch das Profil eines Objekts hinsichtlich der in der Teilmenge vorhandenen Prädiktoren das Kriterium vorhergesagt werden kann. Für dichotome Kriterien und Prädiktoren beschreibt Hollmann (1991) das Vorgehen so, dass für jedes distinkte Profil die Vorhersage getroffen wird, die die beste Fitting-Genauigkeit erbringt, bei Gleichheit der Kriteriumsklassen

für ein Prädiktorenprofil wird hingegen keine Entscheidung getroffen. Als Vorhersagegüte wird wahlweise ein χ^2-Wert oder die Differenz zwischen der Anzahl korrekt und falsch klassifizierter Objekte herangezogen, was Hollmann (1991) zu der Kritik bewegt, dass die Differenzierung der Fehlerarten verloren geht. Zum anderen kritisiert er, dass einzelne Zellen einer Prädiktorkombination im ursprünglichen Verfahren nur dann genutzt werden können, wenn die übrigen Zellen den Vorhersagewert der Kombination nicht zu stark schmälern.

Mit diesen Modifikationen als Rahmen wird deshalb hier folgendermaßen vorgegangen: Zunächst wird eine Mindestbesetzungszahl von Zellen vorgegeben, die eine Interpretation von Zellen möglich macht, zum anderen wird eine Mindesttrefferquote vorgegeben, unter der eine Zelle nicht weiter in der Interpretation berücksichtigt wird (die Grundquote bietet sich hier als Maßstab an). Mit diesen Vorgaben werden alle Zellen aller Prädiktorkombinationen einer vorgegebenenn Maximalgröße hinsichtlich eines Kriteriums untersucht, dass unter zahlreichen Möglichkeiten gewählt werden kann: Trefferquote, Sensitivität, Spezifität, positive und negative Validität, spezifische Treffer-Fehler-Differenz oder Treffer-Fehler-Quotient etc. Dann wird die in diesem Sinne beste Zelle herausgegriffen, um die erste Entscheidungsregel zu bilden, mit der alle Objekte klassifiziert werden, die das entsprechende Prädiktorprofil aufweisen, um mit den verbleibenden eine erneute Analyse vorzunehmen.

Das auch von Wottawa (1987a) benannte Problem des fehlenden Fehlermodells kann allerdings auch in TYPAG nicht gelöst werden. Ähnliches gilt für die Notwendigkeit großer Stichproben und die potenzielle Unentscheidbarkeit zwischen ähnlich guten Modellen. Auf den Einwand, dass die Gültigkeit gefundener Strukturen nicht abgesichert wird, kann - abgesehen von der prinzipiellen Unmöglichkeit der endgültigen Klärung bei induktiven Schlüssen - auch in analoger Weise reagiert werden, wie es für HYPAG formuliert wurde:

> The results can not be seen as any kind of „truth“, but as optimal descriptions of the connection between independent variables and the target criteria. The usefulness of these descriptions depends on the external goals of the relevant study. therefore HYPAG is to be seen as a kind of supplement to conventional methods of analysis, especially in cases of weak scales or if technical assumptions of other methods are only usatisfactorily fulfilled. (Wottawa, 1987a, S.92)

Umsetzung in Gier$_{[1.0]}$.typag

Das Verfahren Gier$_{[1.0]}$.typag stellt eine eingeschränkte Version des TYPAG-Verfahrens dar, insofern als dass in der jetzigen Form nicht mehr als ein Durchlauf des Verfahrens gleichzeitig möglich ist und als Kriterien nur die Klassifikationsgenauigkeit und das KNMK zur Verfügung stehen. Die Integration des Verfahrens wurde in erster Linie motiviert durch die Beobachtung, dass zur Zeit keine windowsbasierte rechnergestützte Implementierung von TYPAG verfügbar ist und das Verfahren einen Überblick über konfigurale Zusammenhänge in der Stichprobe liefert.

Dementsprechend besteht die aktuelle Implementierung aus folgenden Schritten:

- Die minimale Fallzahl pro Regel und die maximale Kombinationstiefe werden zusammen mit dem Bewertungskriterium spezifiziert
- Das Verfahren wird mit den gewählten Parametern für einen eingelesenen Datensatz gestartet
- Die Auswertung wird als Text ausgegeben.

Die Auswertung besteht für einen Datensatz aus einer nach Güte der Regel geordneten Liste von Antezedenzbedingungen und Konsequenzen sortiert nach Verknüpfungstiefe. Dabei werden Regel, abgedeckte Fallzahl und Kriterium spezifiziert.

Q-TYPAG als Erweiterungsmöglichkeit

Das Verfahren gliedert sich nicht in die übrigen Verfahren von Gier$_{[1.0]}$ein, da nicht die vollständige Objektmenge klassifiziert wird. Für die Nutzung zur Lösung der Basisaufgabe wäre eine deutliche Modifizierung des Verfahrens notwendig, da die Anzahl möglicher Prädiktorenwerte zu einer Inflation von

zu untersuchenden Profilen führen würde und die Besetzungszahlen zu klein würden. Eine solche Erweiterung könnte die folgende Form annehmen, die als Q-TYPAG (Quantil-Typag) bezeichnet werden soll:

Als Voraussetzung werden erneut Klassifikationsgütemaß, Mindestbesetzung pro Profil und maximale Verknüpfungstiefe angegeben. Zusätzlich werden Auflösungsgrad und Regelzahlmultiplikator m (mit $0 < m \leq 1$) spezifiziert. Der erste Wert bestimmt die Anzahl der Gruppen, in die Objekte hinsichtlich jedes Prädiktors aufgeteilt werden. Bei einem Auflösungsgrad von zwei wird hierzu der Median herangezogen, bei drei Gruppen sind es die Terzil-Werte, bei vier Gruppen die Quartile etc. Der Regelzahlmultiplikator dient zur Berücksichtigung des Umstands, dass eine Entscheidung aufgrund weniger Variablen in den meisten Fällen stabiler sein sollte als eine Entscheidung, die von vielen Variablen abhängig ist. Dabei wird der Wert für die Klassifikationsgüte einer Regel mit einer Verknüpfungstiefe von t mit $m^{(t-1)}$ multipliziert, nachdem dieser Wert gegebenenfalls so transformiert worden ist, dass höhere Werte eine höhere Güte bedeuten. Das Produkt wird dann zum Ranking aller Regeln bis zur spezifizierten Verknüpfungstiefe verwendet. Die Regel mit dem höchsten Wert wird ausgewählt und alle klassifizierten Objekte werden aus der Stichprobe entfernt, bevor das Verfahren eine weitere Iteration durchläuft so lange bis keine Regel mehr gefunden werden kann, die der Mindestbesetzungsbedingung genügt. Die restlichen Fälle werden der Klasse zugeordnet, die für die letzte Regel das Klassifikationsgütekriterium maximiert. Durch die Wahl des Auflösungsgrades kann sichergestellt werden, dass die Anzahl der untersuchten Profile im Rahmen bleibt, zudem sind höhere Besetzungen zu erwarten. Durch Verwendung des Multiplikators wird zusätzlich sichergestellt, dass einfachen Regeln gegenüber komplexeren der Vorzug gegeben wird, so dass diese Modifikation hier als Empfehlung ausgesprochen wird.

Charakterisierung des Gier$_{[1.0]}$.typag-Algorithmus

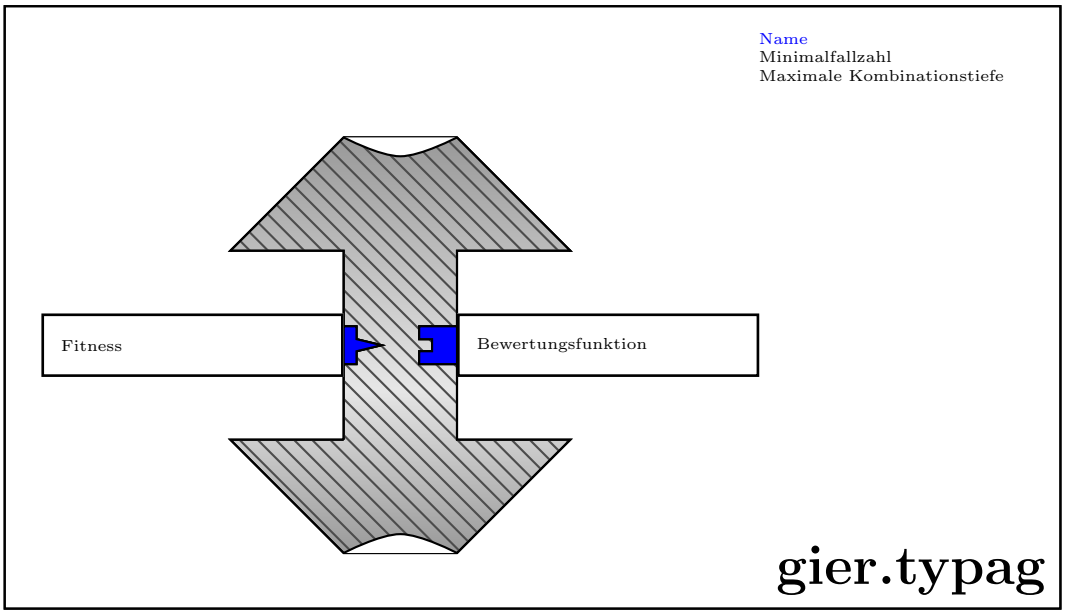


Abbildung C.48: VBk-Darstellung des Algorithmus Gier$_{[1.0]}$.typag: Die Abbildung zeigt die abstrakten Platzhalter für variable Komponentenklassen und Parameter für den Gesamtalgorithmus.

Der Gier$_{[1.0]}$.typag-Algorithmus stellt wird ebenfalls in die Gier$_{[1.0]}$-Toolbox aufgenommen. Aufgrund der einfachen Struktur des Algorithmus, wird er durch nur zwei Komponentenklassen charakterisiert (vgl. Abbildung C.48), die zudem von Gier$_{[1.0]}$.ga übernommen werden können. Als allgemeine Parameter verbleiben:

- **Name**: Der Name des Prototypen.
- **Minimalfallzahl**: Die minimale Fallzahl zur Berücksichtigung einer Regel.
- **Maximale Kombinationstiefe**: Es werden Lösungen mit einer verknüpften Variablenzahl generiert, deren Höhe zwischen 1 und dem Wert dieses Parameters liegt.

C.4.4 Integration von r2-f2 in die Toolbox

Ein durch das ABC-Programm inspirierter Algorithmus soll die Regelinduktionstoolbox von GIER$_{[1.0]}$ergänzen. Grundidee, Struktur und Integration des Algorithmus sollen im Folgenden vorgestellt werden.

Grundstruktur des r2-f2-Algorithmus

Der r2-f2-Algorithmus orientiert sich an den Prinzipien de fast and frugal heuristics, allerdings ist er nicht vollständig fast and frugal, weshalb sich der Name „**r**ather **f**ast and **r**ather **f**rugal"-Algorithmus erklärt, was zusammengefasst wird zu r2-f2. Der Algorithmus ist darauf abgestimmt, das gleiche Problem zu lösen wie etwa der GIER$_{[1.0]}$.ga-Algorithmus, also eine binäre Klassifikation vorzunehmen unter Berücksichtigung einer Menge von Prädiktoren mit Werten im Intervall $[0, 100]$ Der Algorithmus ist ebenfalls parametrisierbar und daher eher als eine Klasse von Verfahren zu verstehen.

Der Algorithmus übernimmt das Prinzip des *One Reason Decision Making*, allerdings sind die Prädiktoren nicht wie im Fall von TTB oder lexikographischer Bäume binäre Prädiktoren. Die Ergebnisse von r2-f2 lassen sich als Entscheidungsbaum darstellen mit drei Knoten und zwei Endknoten pro Stufe, die der Wurzel folgt. Dabei gibt es zwei Intervalle von Prädiktorwerten, für die eine Entscheidung in die eine bzw. die andere der beiden Richtungen getroffen wird, sowie einen Indifferenzbereich, der die beiden Intervalle trennt, die jeweils ein Skalenende miteinschließen.

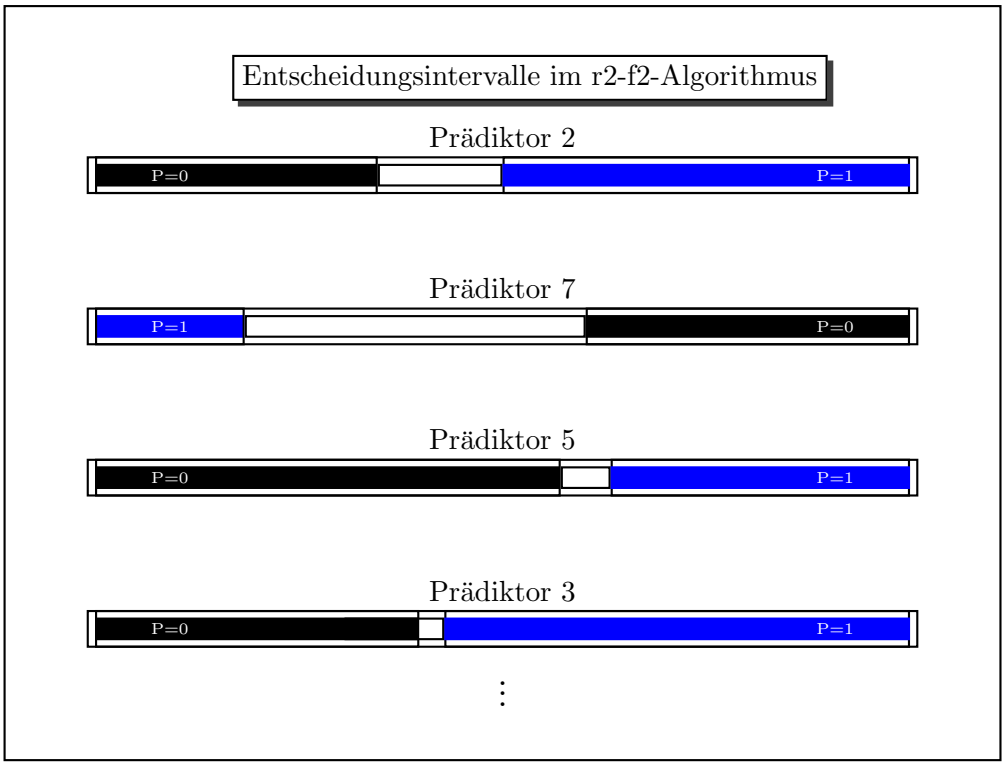


Abbildung C.49: Beispielergebnis für den r2-f2-Algorithmus: Die Prädiktoren sind in eine Reihenfolge gebracht worden (x_2, x_7, x_5, x_3, ...) und die Intervallgrenzen für die Prädiktionen sind zusammen mit der Bestimmung der Prädiktionswerte vorgenommen worden. Der eingezeichnete weiße Abschnitt der Balken entspricht dem Indifferenzbereich.

Die Option eines Indifferenzbereiches wurde für *Take The Best* durchaus erwogen, aber aufgrund der Notwendigkeit der Parametersetzung fallen gelassen:

> There are two major ways lexicographic strategies such as Take The Best can be extended to make inferences from quantitative values ... . In the first version, search continues until the first cue is found in which the two values are different; in the second, search continues until the difference between cue values exceeds a threshold or „ just noticeable difference". To avoid the arbitrariness in defining how large the threshold should be, we went with the first version. (Czerlinski et al., 1999, S. 111)

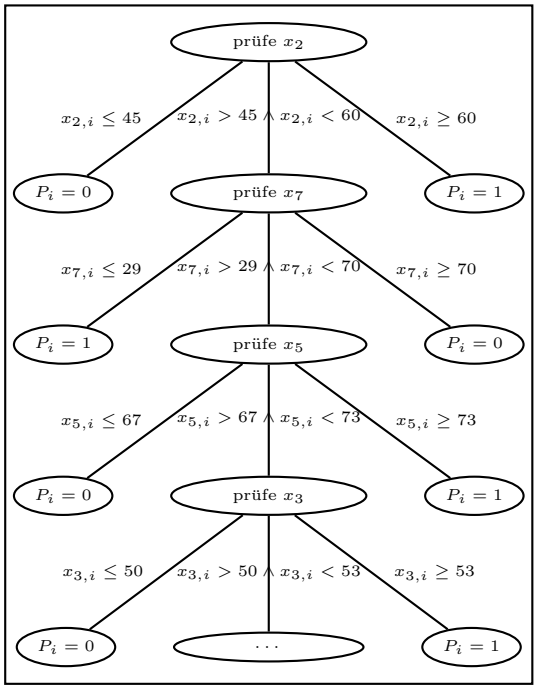


Abbildung C.50: Ergebnisäquivalenter Entscheidungsbaum zum Beispielergebnis des r2-f2-Algorithmus: Der Baum entspricht der in Abbildung C.49 gezeigten Ergebnisstruktur. Prädiktorwerte werden sukzessive überprüft, alle Werte außerhalb des Indifferenzintervalls führen zu einer sofortigen Entscheidung.

Im r2-f2-Algorithmus soll diese Variante verfolgt werden. Der Algorithmus durchläuft zwei Phasen:

1. Die Variablen werden in eine Rangfolge gebracht.
2. Für jede Variable werden Entscheidungsgrenzwerte festgelegt.

In diesen zwei Phasen wird als Ergebnis ein Regelsatz erstellt, der in der Lage ist, Objekte aufgrund ihres Prädiktorprofils zu klassifizieren. Dabei wird in jeder Regel des Satzes ausschließlich ein einziger Prädiktor verwendet. Pro Prädiktor gibt es eine oder zwei Regeln, wobei im zweiten Fall die Regeln unterschiedliche Prädiktionswerte haben. Ist der Prädiktorwert im Indifferenzbereich, wird keine der beiden entsprechenden Regelantezedenzbedingungen erfüllt, so dass die Regeln für den im Ranking nachfolgenden Prädiktor überprüft werden.

Abbildung C.49 zeigt ausschnittsweise ein mögliches Ergebnis nach einer Anwendung des r2-f2-Algorithmus. Die Prädiktoren sind in eine Reihenfolge gebracht worden, Vorhersagewerte und Intervallgrenzen sind bestimmt. Das Entscheidungsprinzip ist im Entscheidungsbaum von Abbildung C.50 für dieses Beispiel visualisiert. Prädiktoren, die in der Reihenfolge später stehen, werden nur überprüft, wenn die Werte auf den in der Reihgenfolge vorangegangenen Prädiktoren alle im Indifferenzbereich liegen. An dieser Stelle sollen Varianten für die beiden genannten Schritte des Algorithmus diskutiert werden.

Erster Schritt: Ranking der Variablen

Zum Erstellen der Reihenfolge, in der die Variablen für die spätere Entscheidung herangezogen werden, stehen sieben verschiedene Basisverfahren zur Verfügung, deren Details im Folgenden erläutert werden sollen. Die ersten fünf dieser Verfahren berechnen statistische Kenngrößen, die zur Rangfolebildung herangezogen werden:

1. Die Mittelwertsdifferenz ($\Delta\bar{x_j}$) wird als Betrag der Differenz der Mittelwerte der beiden Gruppen von Objekten gebildet, deren Kriteriumswert 0 bzw. 1 ist, also als

$$\Delta\bar{x_j} = \left| \frac{\sum_{i:y_i=0} x_{j,i}}{\sum_{i:y_i=0} 1} - \frac{\sum_{i:y_i=1} x_{j,i}}{\sum_{i:y_i=1} 1} \right| . \tag{C.2}$$

Die Prädiktoren werden in absteigender Reihenfolge der $\Delta\bar{x}$-Werte sortiert.

2. Die Mediandifferenz (ΔMd_x) bezeichnet den Betrag der Differenz der Medianwerte der beiden Teilgruppen, also
$$\Delta Md_{x_j} = \left| Md_{x_{j,1}} - Md_{x_{j,0}} \right| . \tag{C.3}$$
Die Prädiktoren werden in absteigender Reihenfolge der ΔMd_{x_j}-Werte sortiert.

3. Die Medianrangdifferenz ($\Delta R_{Md_{x_j}}$) berzeichnet die Differenz der Rangplätze der Teilgruppenmediane:
$$\Delta R_{Md_{x_j}} = \left| R_{Md_{x_{j,1}}} - R_{Md_{x_{j,0}}} \right| . \tag{C.4}$$
Die Prädiktoren werden in absteigender Reihenfolge der $\Delta R_{Md_{x_j}}$-Werte sortiert.

4. Es werdenen t-Tests zum Vergleich der Teilgruppenmittelwerte für die Prädiktoren berechnet und vom Betrag her verglichen ($\left|t_{x_j}\right|$) . Dabei ist (vgl. Bortz, 1993)
$$\left|t_{x_j}\right| = \left| \frac{\bar{x}_{j,1} - \bar{x}_{j,0}}{\hat{\sigma}_{\bar{x}_{j,1} - \bar{x}_{j,0}}} \right| , \tag{C.5}$$
mit
$$\bar{x}_{j,1} = \frac{\sum_{i:y_i=1} x_{j,i}}{\sum_{i:y_i=1} 1} \tag{C.6}$$
und
$$\bar{x}_{j,0} = \frac{\sum_{i:y_i=0} x_{j,i}}{\sum_{i:y_i=0} 1} . \tag{C.7}$$
Die Prädiktoren werden in absteigender Reihenfolge der $\left|t_{x_j}\right|$-Werte sortiert.

5. Es werden Mann-Whitney-U-Tests (mit Verbundwertkorrektur) berechnet und die U-Werte in z-Werte transformiert. Bezeichnet man die Summe der Rangwerte in den Teilgruppen als
$$T_{j,0} = \sum_{i:y_i=0} R_{x_{j,i}} \tag{C.8}$$
bzw.
$$T_{j,1} = \sum_{i:y_i=1} R_{x_{j,i}} , \tag{C.9}$$
und sind weiterhin $n_0 = \sum_{i:y_i=0} 1$ und $n_1 = \sum_{i:y_i=1} 1$, dann ist
$$U_j = n_1 \cdot n_0 + \frac{n_0 \cdot (n_0 + 1)}{2} - T_{j,0} . \tag{C.10}$$
Die Transformation in z-Werte (vgl. Bortz, 1993) erfolgt über
$$z_j(U_j) = \frac{U_j - \frac{n_0 \cdot n_1}{2}}{\sqrt{\frac{n_0 \cdot n_1 \cdot (n_0 + n_1 + 1)}{12}}} . \tag{C.11}$$
Als Kriterium für die Rangordnung werden die Beträge der ermittelten z-Werte ($|z_j(U_j)|$) miteinander verglichen und die Prädiktoren werden in absteigender Reihenfolge dieser Werte sortiert.

Zwei weitere Verfahren vergleichen verschiedene Möglichkeiten der Entscheidungsbildung aufgrund des Prädiktors und nehmen die Rangfolgebildung aufgrund eines maximal erreichbaren Gütewerts vor:

1. Es werden alle möglichen Spaltungen der Prädiktorwerte in zwei Intervalle betrachtet. Für jede Spaltung werden für die beiden Intervalle unterschiedlichen Vorhersagen getroffen und zwar so, dass das gewählte Gütekriterium (entweder Klassifikationsgenauigkeit oder KNMK) maximiert wird. Für jeden Prädiktor wird der auf diesem Weg erreichte maximal hohe Kriteriumswert als Vergleichswert berücksichtigt, um die Prädiktoren in absteigender Reihenfolge dieses Kriteriums in eine Rangordnung zu bringen. Dieses Kriterium soll mit $split2_{max}$ abgekürzt werden.

2. Das zweite Kriterium ähnelt dem ersten, nur wird das Intervall der Prädiktorvariablen nicht in zwei, sondern in drei Teile gespalten, wobei das mittlere Intervall auch leer sein kann. Dieses mittlere Intervall wird als Indifferenzzone verstanden, für die keine Vorhersagen getroffen werden. Es werden nun alle möglichen Aufteilungen des Wertebereichs untersucht, um das auf diesem Weg erreichbare Maximum des Gütekriteriums als Vergleichswert zu nutzen, um alle Prädiktoren in absteigender Reihenfolge dieses Werts zu ordnen. Dieses Kriterium soll mit $split3_{max}$ abgekürzt werden.

Um zu verhindern, dass bei den letzten beiden Varianten solche Lösungen als optimal gefunden werden, die nur sehr wenige Fälle (richtig) klassifizieren, kann eine Mindestanzahl von betroffenen Fällen angegeben werden, oder die Verwendung einer von mehreren Korrekturformeln etabliert werden, die die Anzahl der betroffenen Fälle in die Bestimmung des Fitnesswertes mit einfließen lassen.

Zweiter Schritt: Bestimmung der Cut-Off-Grenzen

Sind im ersten Schritt die Prädiktoren für jede Regel in der Ergebnisregelmenge bestimmt worden, werden im zweiten Schritt anschließend Cut-Off-Werte für jeden einzelnen Prädiktor bestimmt. Dazu stehen zehn Grundmöglichkeiten zur Auswahl, die teilweise weitere Parameter enthalten. Dabei entsprechen vier Verfahren einer einfachen Festlegung der Intervalle durch Statistiken, sechs weitere Verfahren nutzen einen Vergleich verschiedener Modelle.

Zu den ersten vier Verfahren gehören:

1. Festlegung durch Mittelwerte ohne Indifferenzbereich: Es wird kein Indifferenzbereich festgelegt, also wird eine Entscheidung durch den ausgewählten Prädiktor erzwungen. Dementsprechend muss nur ein Cut-Off-Wert festgelegt werden. Die Festlegung durch Mittelwert setzt diesen Wert wahlweise auf den Prädiktormittelwert in der Gesamtstichprobe oder den Mittelwert der ungewichteten Gruppenmittelwerte. Die Prädiktionswerte für die beiden Intervalle werden so festgelegt dass das verwendete Gütemaß maximal wird.

2. Festlegung durch Mittelwerte mit Indifferenzbereich: In diesem Fall müssen zwei Cut-Off-Werte festgelegt werden, um den Indifferenzbereich zu markieren. Dazu wird erneut entweder der Stichprobenmittelwert oder der Mittelwert der Gruppenmittelwerte berechnet. Dann wird ein Indifferenzintervall symmetrisch um diesen Wert gelegt, das einer durch einen Parameter vorgegebenen Größe entspricht.

3. Festlegung durch Mediane ohne Indifferenzbereich: Das Vorgehen entspricht dem in der erstgenannten Methode, nur dass nicht Mittelwerte, sondern Mediane eingesetzt werden. In ähnlicher Weise kann demenstprechend entweder der Stichprobenmedian (Md_{x_j}) genutzt werden oder der Mittelwert der Gruppenmediane

$$\bar{Md}_{x_j} = \frac{1}{2}(Md_{x_{j,1}} + Md_{x_{j,0}}) \tag{C.12}$$

bzw. der mittlere Rangplatz zwischen den Rangplätzen der Gruppenmediane

$$\bar{R}_{Md_{x_j}} = \frac{1}{2}(R_{Md_{x_{j,1}}} + R_{Md_{x_{j,0}}}). \tag{C.13}$$

4. Festlegung durch Mediane mit Indifferenzbereich: Es wird wie im Fall ohne Indifferenzbereich ein mittlerer Wert bestimmt, um den symmetrisch ein in der Größe parametrisiertes Indifferenzintervall gelegt wird.

Zusätzlich zu diesen Verfahren gibt es sechs modellvergleichende Verfahren. Dabei wird durch die Methode vorgegeben, in welchem Intervall nach Cut-Offs gesucht werden soll. Sind dieser Bereich und die Anzahl der Cut-Offs bestimmt, werden in jedem Fall alle möglichen Modelle erstellt und hinsichtlich eines Gütekriteriums in der Klassifikationsleistung verglichen. Allerdings werden alle Regeln, deren Antezedenzbedingung von weniger als einer festgelegten Anzahl von Objekten erfüllt wird, mit einem Gütewert von 0 bewertet. Das Modell mit dem höchsten Wert hinsichtlich dieses Kriteriums wird zur Grenzbestimmung verwendet. Die sechs Methoden sind im Einzelnen:

1. Festlegung durch Mittelwerte ohne Indifferenzbereich mit Optimierung: Durch die Gruppenmittelwerte als Randpunkte wird das Intervall vorgegeben, innerhalb dessen nach der hinsichtlich es Kriteriums optimalen Cut-Off-Setzung gesucht wird.

2. Festlegung durch Mittelwerte mit Indifferenzbereich mit Optimierung: Die Mittelwerte geben erneut die Suchgrenzen vor, wobei der Indifferenzbereich nun den Bereich zwischen den Mittelwerten überdecken muss. Bei der Wahl eines symmetrischen Bereichs wird der Mittelpunkt des Indifferenzbereichs zwischen den Mittelwerten gesucht, wird eine feste Größe vorgegeben, wird der Indifferenzbereich mit dieser Einschränkung gesucht.

3. Festlegung durch Mediane ohne Indifferenzbereich mit Optimierung: Durch die Gruppenmediane als Randpunkte wird das Intervall vorgegeben, innerhalb dessen nach der hinsichtlich des Kriteriums optimalen Cut-Off-Setzung gesucht wird.

4. Festlegung durch Mediane mit Indifferenzbereich mit Optimierung: Es wird analog zum zweiten Verfahren, vorgegangen, wobei die Grenzen des Suchbereichs durch die Mediane vorgegeben werden.

5. Freie Optimierung ohne Indifferenzbereich: Es werden alle möglichen Werte für einen Cut-Off untersucht, also bis zu 100 verschiedene Modelle überprüft.

6. Freie Optimierung mit Indifferenzbereich: Es werden alle möglichen Werte für zwei Cut-Offs untersucht, je nach Einstellung von gewünschter Indifferenzbereichsgröße also Modelle im Bereich von bis zu mehreren Tausend betrachtet. Das Modell mit dem höchsten Kriteriumswert wird ausgewählt.

Es sollte angemerkt werden, dass diese Formen der Optimierung lediglich die Optimierungen von Regeln in Isolation für die *Fitting*-Aufgabe darstellen. Aus diesem Grund bleibt zum einen der Suchraum zeitlich bearbeitbar, zum anderen aber ist in keiner Weise gewährleistet, dass sich nicht auch für die *Fitting*-Aufgabe bessere Lösungen für Regelmengen finden lassen, es handelt sich also selbst im Fall der Wahl einer „optimierenden“ Methode um einen von Grund auf heuristischen Ansatz.

Umsetzung in Gier$_{[1.0]}$

Die Umsetzung des oben beschriebenen Verfahrens in Gier$_{[1.0]}$erfolgt über die Definition einer zusätzlichen Algorithmenklasse mit fünf Komponenten (s. Abbildung C.51). Das Prinzip der Wiederverwendung für diese Komponenten ist in Abbildung C.52 veranschaulicht: Drei der Komponenten können von Gier$_{[1.0]}$.ga übernommen werden, zwei der Komponenten verlangen eine eigene Implementation.

Als allgemeine Parameter des Algorithmus werden die folgenden benötigt:

- **Name**: Der Name des Prototypen.
- **Minimalfallzahl verwenden**: Wird diese Regel aktiviert, so werden bei der Optimierung nur Intervallgrenzen berücksichtigt, die zu einer durch den zusätzlichen Parameter *Minimalfallzahl pro Regel* vorgegebenen Anzahl von Entscheidungen führen.

Auch für diesen Algorithmus kann eine Regel für die nicht-klassifizierten Fälle hinzugefügt werden.

Die Ranking-Komponente Abbildung C.53 zeigt die Optionen für den Schritt der Reihenfolgebildung der Prädiktoren. Folgende Optionen stehen zur Auswahl:

- **Mittelwertsdifferenz**: Die Mittelwertsdifferenz ($\Delta\bar{x}_j$) wird als Sortierungskriterium verwendet.
- **Mediandifferenz**: Die Mediandifferenz (ΔMd_x) wird als Sortierungskriterium verwendet.
- **Medianrangunterschied**: Die Medianrangdifferenz ($\Delta R_{Md_{x_j}}$) wird als Sortierungskriterium verwendet.
- **t-Tests**: Die $\left|t_{x_j}\right|$-Werte werden als Sortierungskriterium verwendet.
- **Mann-Whitney-U-Tests**: Die $|z_j(U_j)|$-Werte werden als Sortierungskriterium verwendet.

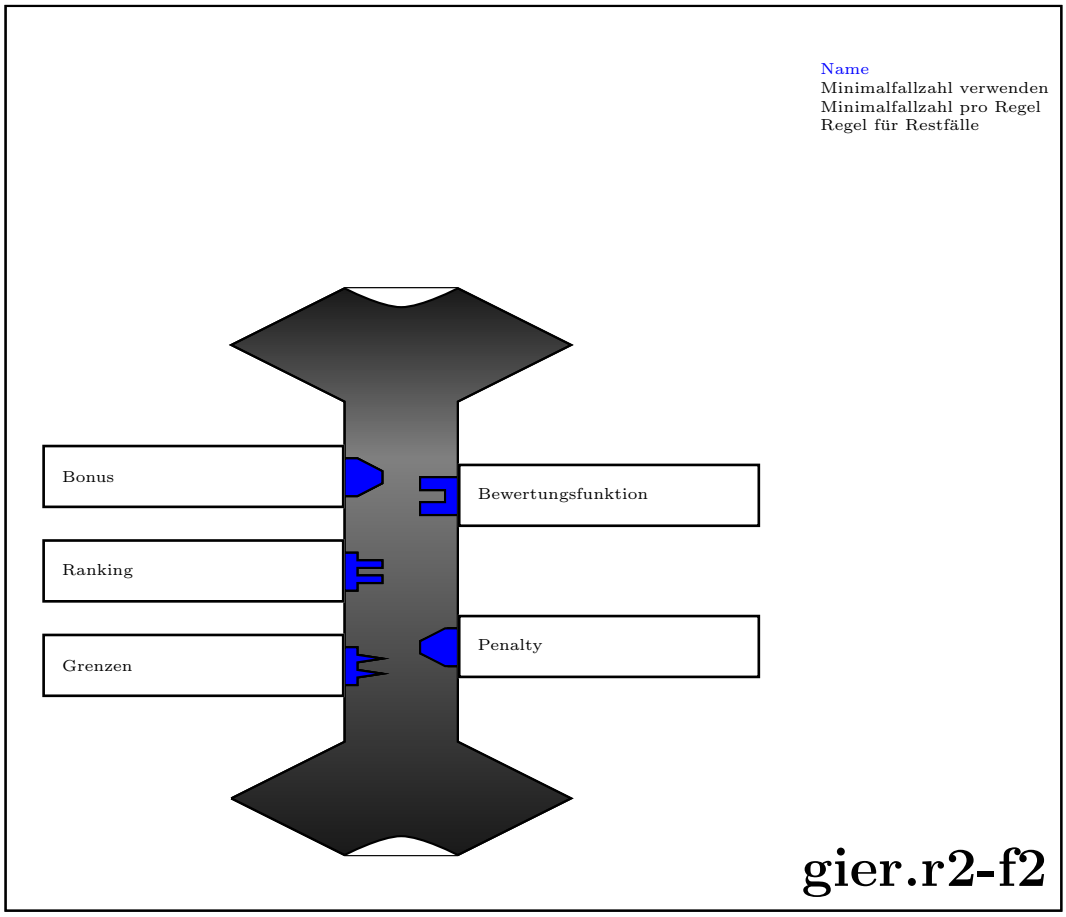


Abbildung C.51: VBk-Darstellung des Algorithmus GIER$_{[1.0]}$.r2-f2: Die Abbildung zeigt die abstrakten Platzhalter für variable Komponentenklassen und Parameter für den Gesamtalgorithmus.

- **Maximale Klassifikationsgüte**: Der Parameter *Indifferenzbereich* entscheidet darüber ob das $split2_{max}$-Kriterium oder das $split3_{max}$-Kriterium als Sortierungskriterium verwendet wird.

Mit diesen Einstellungen kann die Prädiktorreihenfolge bestimmt werden.

Die Grenzen-Komponente Abbildung C.54 zeigt die sechs parametrisierten Komponenten, mit deren Hilfe die oben vorgestellten zehn Verfahren zur Intervallbestimmung implementiert werden. Im Einzelnen sind dies:

- **Mittelwerte**: Die Intervallgrenzen werden durch Mittelwerte ohne Indifferenzbereich festgelegt. Der Parameter *Optim.* bestimmt, ob ein Modellvergleichsverfahren eingesetzt wird oder nicht. Der Parameter *Mw.d.Gruppenmw.* bestimmt, ob der Mittelwert der Gruppenmittelwerte verwendet wird im Gegensatz zum Mittelwert der Stichprobe. Soll ein Cut-Off vorgegeben werden, wird dieser durch den Parameter *Mw.entsch.* angegeben.

- **Mittelwerte und Indifferenzbereich**: Die Intervallgrenzen werden durch Mittelwerte mit Indifferenzbereich festgelegt. Der Parameter *Optim.* bestimmt, ob ein Modellvergleichsverfahren eingesetzt wird oder nicht. Der Parameter *Mw.d.Gruppenmw.* bestimmt, ob der Mittelwert der Gruppenmittelwerte verwendet wird im Gegensatz zum Mittelwert der Stichprobe. Mit dem Parameter *Symm.* wird die Symmetrieeigenschaft des Indifferenzintervalls, mit dem Paramter *Indiff.größe* eine optionale feste Intervallgröße bestimmt.

- **Mediane**: Die Intervallgrenzen werden durch Mediane ohne Indifferenzbereich festgelegt. Der Parameter *Optim.* bestimmt, ob ein Modellvergleichsverfahren eingesetzt wird oder nicht. Der Parameter *Gruppenmed.* bestimmt, ob der Mittelwert der Gruppenmediane (bzw. bei entsprechender Setzung des Paramters *Ränge zählen* der Wert, der dem mittleren Rang zwischen den Rangplätzen der

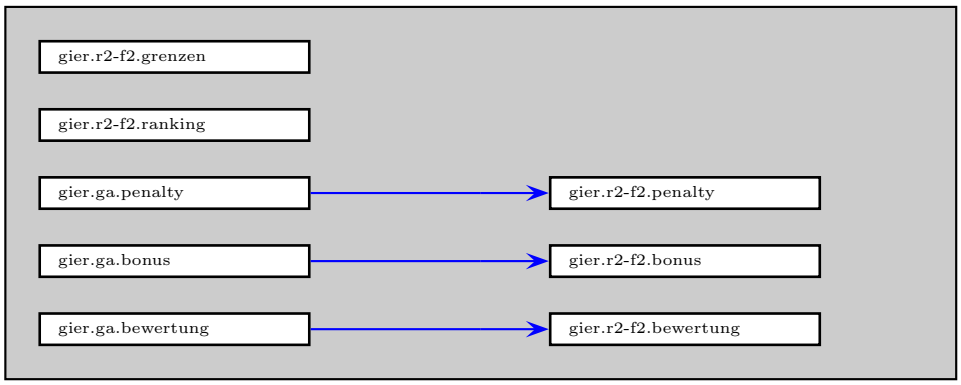


Abbildung C.52: VBk-Darstellung der in GIER$_{[1.0]}$.r2-f2 enthaltenen Komponenten: Die dargestellten Komponenten sind in GIER$_{[1.0]}$.r2-f2 enthalten. Eine modulare Wiederverwendung wird dort durch Pfeile angezeigt, wo sie realisiert worden ist.

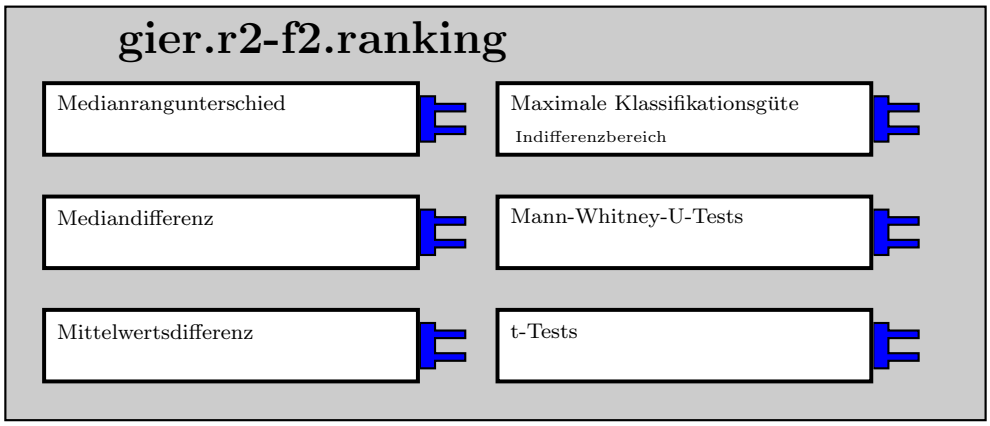


Abbildung C.53: VBk-Darstellung der Komponente „ranking“ des Algorithmus GIER$_{[1.0]}$.r2-f2

Mediane entspricht) verwendet wird im Gegensatz zum Median der Stichprobe. Soll ein Cut-Off vorgegeben werden, wird dieser durch den Parameter *Med.entsch.* angegeben.

- **Mediane und Indifferenzbereich**: Die Intervallgrenzen werden durch Mediane mit Indifferenzbereich festgelegt. Der Parameter *Optim.* bestimmt, ob ein Modellvergleichsverfahren eingesetzt wird oder nicht. er Parameter *Gruppenmed.* bestimmt, ob der Mittelwert der Gruppenmediane (bzw. bei entsprechender Setzung des Paramters *Rangz.* der Wert, der dem mittleren Rang zwischen den Rangplätzen der Mediane entspricht) verwendet wird im Gegensatz zum Median der Stichprobe. Mit dem Parameter *Symm.* wird die Symmetrieeigenschaft des Indifferenzintervalls, mit dem Paramter *Indiff.größe* eine optionale feste Intervallgröße bestimmt.

- **Freie Optimierung**: Es wird eine parameterfreie Optimierung ohne Indifferenzbereich vorgenommen.

- **Freie Optimierung und Indifferenzbereich**: Es wird eine parameterfreie Optimierung mit Indifferenzbereich vorgenommen.

Scriptdateien

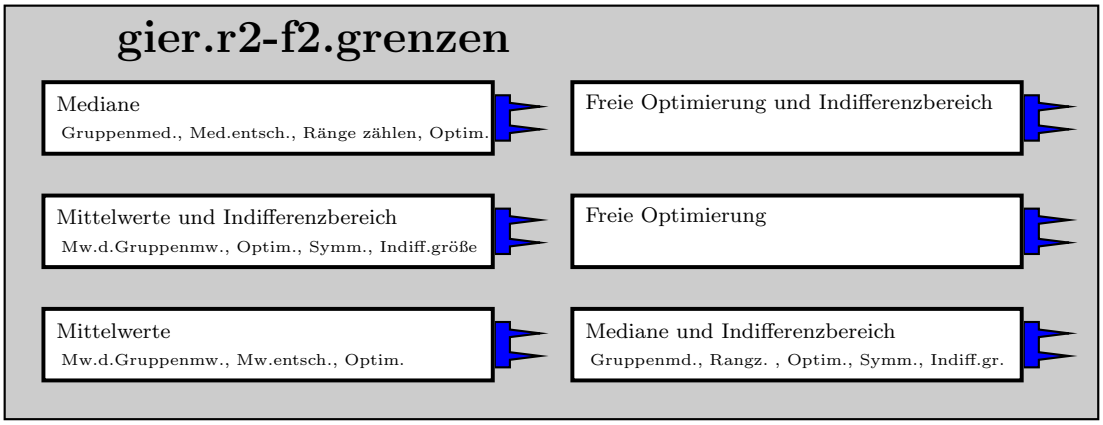


Abbildung C.54: VBk-Darstellung der Komponente „grenzen“ des Algorithmus GIER$_{[1.0]}$.r2-f2

```
r2-f2_beispielprototyp/
(gier.R2-F2_script)
Fallauswahl=False/ n_pro_regel=50/ Restfälle=True/
select_all**
t_var_ranking_rank_diff**
t_cutoff_determination_median_indifference** optimi.=True/ mindestgr.=False/
bonus=False/ penalty=False/ sub_md.=True/ ränge=False/ symm.=True/
Ind.fix=False/ ind.größe=0/
t_objective_set_table_gains** restfälle=False/ false_0=0/ false_1=-1/ true_0=1/ true_1=1/
t_bonus_fitness_no_bonus**
t_penalty_fitness_no_penalty**
```

Listing C.4: Beispiel-Prototyp für GIER$_{[1.0]}$.r2-f2 in GIER$_{[1.0]}$.script

Auch der r2-f2-Algorithmus kann mit allen Einstellungen als Steuerdatei gepeichert werden, und ist damit voll integriert in die Automatisierngsmöglichkeiten. Eine Beispielsteuerdatei hat die in Listing C.4 dargestellte Form. Mit diesem Algorithmus ist die auf das Eingangsproblem zugeschnittene Algorithmensammlung in GIER$_{[1.0]}$vollständig. Durch den modularen Aufbau ist eine spätere Erweiterung allerdings problemlos denkbar. Abbildung C.55 im Anhang B zeigt die vollständige Regelinduktionstoolbox für GIER$_{[1.0]}$. Allerdings sind die Einsteckmodule nur durch jeweils einen Platzhalter markiert. Einige Zusatzmodule und Demonstrationsprogramme wurden weiterhin nicht in die visuelle Sprache integriert. Weiterhin ist das Programm GIER$_{[1.0]}$.simple_heuristics nicht in dieser Darstellung erfasst. Trotzdem ist es mit dieser Darstellung möglich, einen Eindruck über die Verbindungen der Algorithmenkomponenten zu gewinnen und das Konstruktionsprinzip der Algorithmen nachzuvollziehen.

C.4.5 Komponenten zur Simulation künstlichen Lebens

Die GIER$_{[1.0]}$.ROCS-Komponente wurde bereits im Haupttext verwendet. Einige weitere Komponenten der Software sollen hier nur sehr oberflächlich dargestellt werden. Diese Komponenten verbindet die Nutzungsmöglichkeit, Aspekte der Dynamik von Simulationen des künstlichen Lebens erfahrbar zu machen. Im Einzelnen sind folgende weitere Komponenten in GIER$_{[1.0]}$bzw. als Zusatzprogramme umgesetzt worden:

- GIER$_{[1.0]}$.life: Zwei-Farben Version von Conways *Game of Life* (vgl. Gardner, 1970) mit Entwicklungsvarianten, Speichermöglichkeiten und Klassifikationsvariationen.
- GIER$_{[1.0]}$.norms: Eine Umsetzung von Axelrods Meta-Normen-Spiel (Axelrod, 1986, 1997); eine Anwendung der Komponente, die die Bedeutsamkeit von genetischem Drift in der Destabilisierung etablierter Normen demonstriert.
- GIER$_{[1.0]}$.boids: Zweidimensionale Variante der Vorgelschwarmsimuation mit zahlreichen Einstellungsmöglichkeiten und Jäger-Beute-Verhalten zur Demonstration lokaler Bewegungsheuristiken

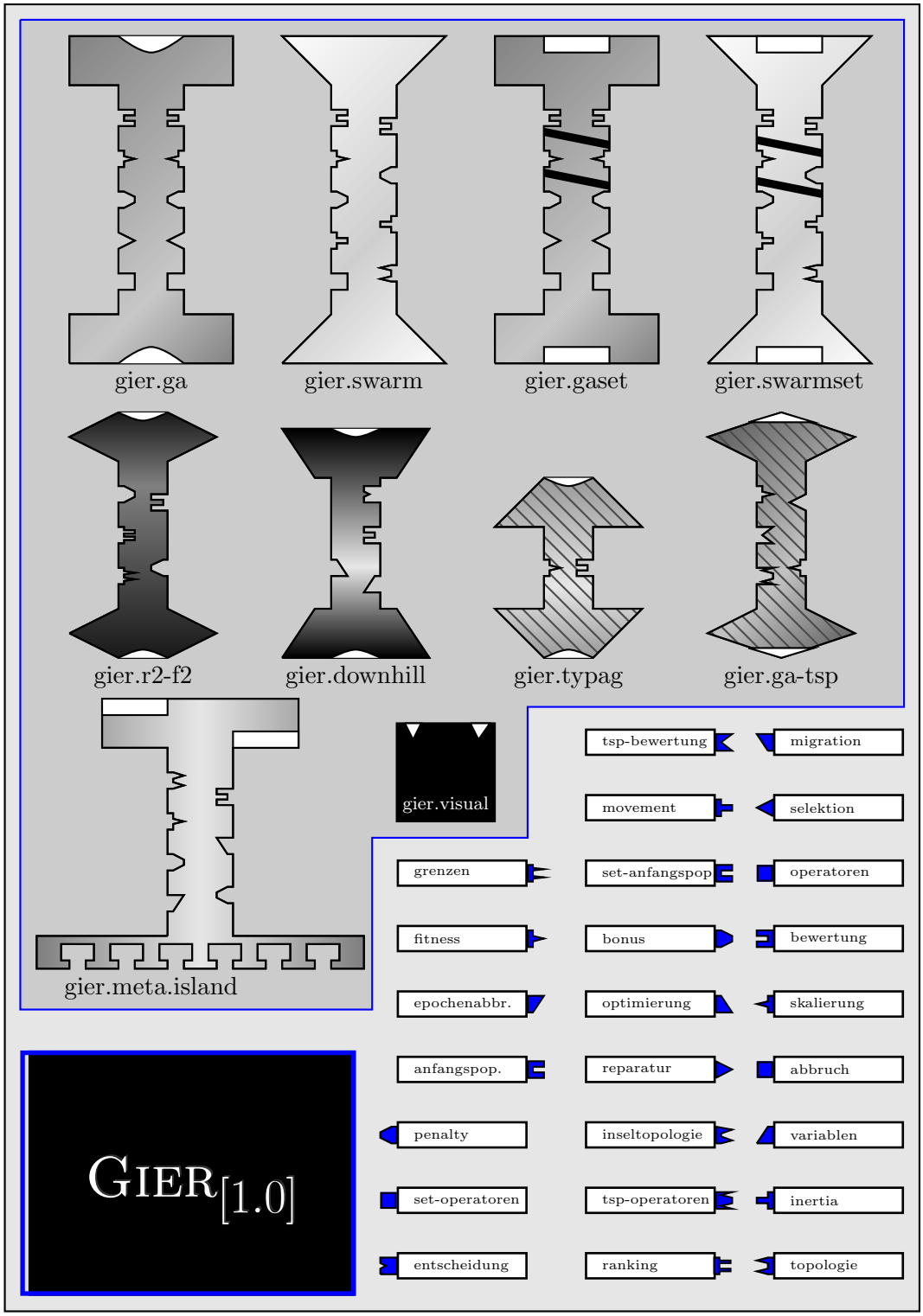


Abbildung C.55: Vollständige Regelinduktionstoolbox in GIER[1.0]

künstlicher Agenten (vgl. Reynolds, 1987). item GIER$_{[1.0]}$.swarmtest: Einige Ergebnisse der Anwendung der Komponente GIER$_{[1.0]}$.swarmtest sind präsentiert worden. Die Komponente ermöglicht die Simulation der Bewegung einzelner Partikel, die mit zwei Stützstellen beschleunigt werden. Neben der Ausgabe der Trajektorien kann die Bewegung der Partikel auch am Bildschirm verfolgt werden.

Ein Großteil der ursprünglichen Faszination des Forschungsprogramms des Künstlichen Lebens ging von der Anziehungswirkung ähnlicher Simulationen aus. Aus diesem Grund werden diese Komponenten zwar nicht als essenzieller Bestandteil der Textform der Arbeit bewertet, aber auf keinen Fall als unnützes Beiwerk betrachtet, ihren Zweck haben sie in der Vermittlung von Konzepten und einem Verständnis für die Ästhetik des Feldes bereits erbracht: Einige der Programme konnten bereits zu pädagogischen Demonstrationszwecken und zur Visualisierung der teilweise abstrakten Konzepte eingesetzt werden.

C.5 Programmierung einfacher Heuristiken: Überblick über den zweiten Werkzeugkasten

An dieser Stelle soll die Entwicklung eines zweiten Werkzeugkastens skizziert werden, der teilweise für die Gewinnung von Simulationsergebnissen genutzt wurde. Der erhoffte Ertrag der Sammlung an Analysewerkzeugen ist aber ein zukünftiger, weshalb im Rahmen dieser Arbeit nur ein kurzer Überblick über einen Teil der Möglichkeiten in diesem Anhang gegeben werden soll.

C.5.1 Komponenten von Gier$_{[1.0]}$.simple_heuristics

In der zweiten Werkzeugkiste befindet sich eine Reihe von Analysetools für die Erforschung einfacher Heuristiken. Dabei sind Werkzeuge für Heuristiken aus unterschiedlichen Bereichen enthalten. Zum einen gibt es Werkzeuge zu Heuristiken für binäre Paarvergleiche, zu Heuristiken für Klassifikationsaufgaben und für Heuristiken zur Schätzung. Dabei bilden die Werkzeuge der ersten Gruppe die größte Teilmenge.

Binäre Paarvergleiche

Beispiele für Heuristiken, die für binäre Paarvergleiche eingesetzt werden können, sind die Rekognitionsheuristik und TTB. Dementsprechend konzentrieren sich die Werkzeuge auf diese Heuristiken. Für die Rekognitionsheuristik ist ein Modul enthalten zur Überprüfung der Manipulierbarkeit der Rekognitionsvalidität für zufällig ausgewählte Objektmengen, die bei simulierten Individuen als bekannt vorausgesetzt werden. Dabei werden Selektionsoperatoren der genetischen Algorithmen auf diesen Bereich übertragen.

Für TTB ist eine größere Zahl spezialisierter Werkzeuge enthalten, wobei TTB als eine Variante lexikografischer Methoden betrachtet wird. Es gibt Werkzeuge zur Überprüfung der von *Fitting*- und *Prediction*-Ergebnissen von allen denkbaren Cueordnungen und den Vergleich mit TTB-Ergebnissen. Dabei ist eine große Zahl von Darstellungsformen und Ergebnisdiagrammen umgesetzt (Streudiagramme, Liniendiagramme, Fehlermatrizen, Shadingplots, Histogrammen etc.). Dabei kann die Qualität der unkonditonalen Sortierung der Cues mit diversen anderen Sortierungsprinzipien verglichen werden, zudem kann die Streuung bei unterschiedlichen Trainings-/Test-Splits beobachtet werden. Es gibt weiterhin spezialisierte Werkzeuge für die Analyse der korrekten und fehlerhaften Wahlen, die eine spezifischere Würdigung der Leistung ermöglichen.

Weiterhin besteht de Möglichkeit, auch an eigenen Datensätzen umfangreichere Simulationsstudien durchzuführen, in denen die datensatzspezifische Leistungsfähigkeit einer größeren Zahl von Heuristiken verglichen werden kann. In die Software integriert sind bisher die folgenden Heuristiken:

- TTB und Varianten (konditionale Validität, *Success Rate* etc.)
- *Take The Last* und *Minimalist*
- *Dawes's Rule* und *Franklin's Rule*
- *Tallying* und gewichtetes *Tallying*
- *Rational Model* bzw. Na'ive Bayes

- Ergänzungen der Verfahren mit der *Recognition Heuristic*
- „Fibonacci-Regression“ für semi-lexikografische Varianten
- „Fibonacci-Tallying“
- *Minimum Incorrect Lexicographic Strategy*
- Lineare Regression als Vergleichsmaßstab

Dabei können eine Reihe von Aufteilungen der Stichprobe untersucht werden und unterschiedlich große Informationsmengen.

Klassifikation

Das Modul zur Klassifikation enthält die in der Arbeit eingesetzten Varianten der *Fast and Frugal Trees*. Ähnlich wie für binär Paarvergleiche kann hier an beliebigen Datensätzen ein Algorithmenvergleich durchgeführt werden. Als Ausgabemöglichkeiten stehen Liniendiagramme, Kastendiagramme für die Analyse von Cuereihenfolgen und ROC-Plots. zur Verfügung. Bisher sind die folgenden Baumvarianten integriert worden:

- FaF MaxVal, FaF MaxSens und FaF MaxAcc
- FaF ZigVal und FaF ZigSens
- FaF ZigVal Plus, FaF ZigVal Majority und FaF ZigVal Fold
- FaF ZigSens Plus, FaF ZigSens Majority und FaF ZigSens Fold
- ein parametrisierbares *General Context Model*
- Varianten für die Konstruktion von *L-Shaped Trees*
- Experimentelle Algorithmen

Auch hier können über die Anzahl der Trials und verschiedene Aufteilungen der Stichprobe umfangreiche Simulationsstudien durchgeführt werden.

Schätzung

Ein letztes Modul enthält Komponenten für Probleme der Schätzung bzw. Regressionprobleme. Ebenfalls an eigenen Datensätzen können Simulationsstudien mit variablen Aufteilungen des Datensatzes vorgenommen werden, wobei Algorithmen hinsichtlich unterschiedlicher Fehlermaße verglichen werden können. Zum Zeitpunkt der Abgabe dieser Arbeit sind die folgenden Heuristiken und Algorithmen enthalten:

- *Estimation Tree*
- QuickEst mit Einstellmöglichkeiten für die Kodierung der Prädiktorwerte und die verwendete *Pruning*-Methode
- Lineare Regression
- Gleichgewichtete Regression
- 32 Varianten zu QuickEst mit vier unterschiedlichen Konstruktionsverfahren (MAX, ZIG, QEST und AQEST), vier unterschiedlichen Schätzverfahren (AV, SL, DT und ET) sowie optionaler Verwendung von Medianen und Mittelwerten.

Diese Bestandteile sollen hier lediglich vorgestellt werden, um deutlich zu machen, in welchem Sinne sich die vorliegende Arbeit als Beitrag zum Programm der Adaptiven Werkzeugkiste versteht. Mit den umgesetzten Algorithmenboxen sind neben pädagogischen Anwendungen auch Analysen der ökologischen Rationalität der Heuristiken möglich sowie Weiter- und Neuentwicklungen von Heuristiken. Die Umsetzung von Algorithmen in allen drei Anwendungsbereichen soll weiterhin die Dimension des Programms deutlich machen. Eine Beschränkung auf binäre Paarvergleiche könnte den falschen Eindruck eines sehr speziellen Zuschnitts der betrachteten Heuristiken entstehen lassen. An einigen Demonstrationen soll zuletzt die Art der visuellen Ergebnisse der zweiten Werkzeugkiste vorgestellt werden.

C.5.2 Einige visuelle Demonstrationen

An einigen Beispieldiagrammen soll ein kleiner Teil der Möglichkeiten der Toolbox GIER$_{[1.0]}$.simple_heuristics veranschaulicht werden. Dabei sollen ein Ausgabeformat für Heuristikenvergleiche, eine unterstützende Anwendung zu Studien zur Rekognitionsheuristik und einige Analysemöglichkeiten für lexikografische Paarvergleiche dargestellt werden. Es wird jeweils durch die Software entweder ein Bitmap oder direkt der für dieses Dokument kompilierte LaTeX-Code ausgegeben.

Heuristikenvergleiche

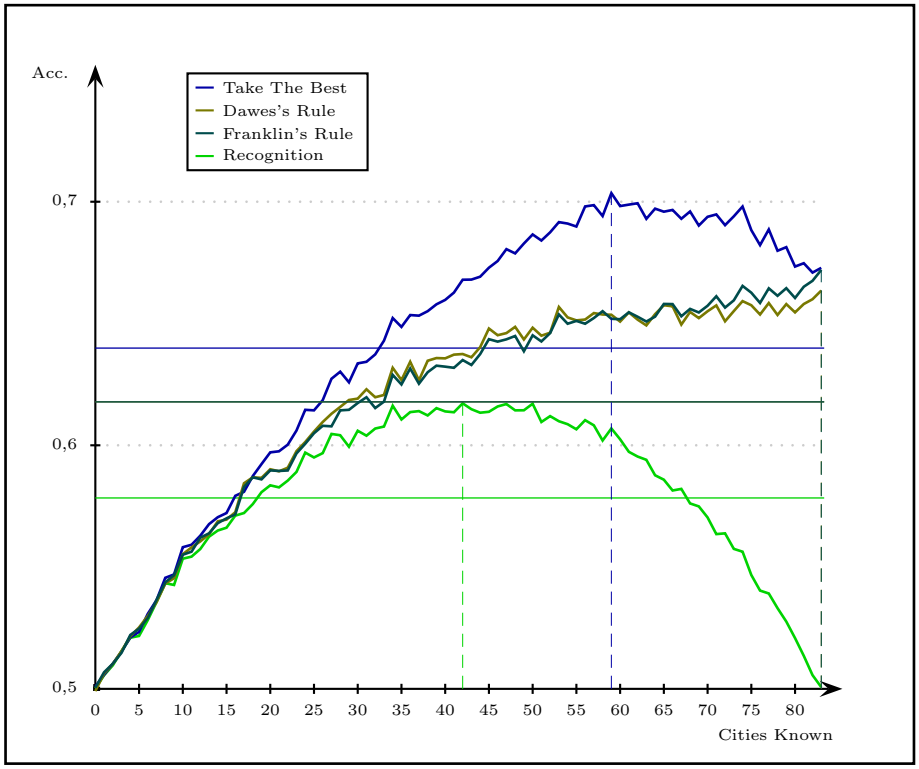


Abbildung C.56: Beispieldarstellung für das Ergebnis eines Vergleichs zwischen Heuristiken hinsichtlich des Vergleichs von Stadtgrößen: Es ist jeweils der über mehrere Trials gemittelte Anteil richtiger Entscheidungen in Abhängigkeit von der Informationsbasis angegeben. Der Mittelwert der Ergebnisse einer Heuristik ist als farbige horizontale Linie eingezeichnet.

Oben sind die Heuristiken geschildert worden, die zu einem Turnier zusammengestellt werden können. Die Ergebnisse eines Turniers lassen sich zum einen als Zahlen bearbeiten, aber auch in verschiedenen grafischen Formaten visualisieren. Als ein Beispiel zeigt Abbildung C.56 das Ergebnis eines Vergleichs von vier Heuristiken (Take The Best, Minimalist, Dawes's Rule und Franklin's Rule) hinsichtlich der Entscheidungsgenauigkeit im *Fitting* bei Variation der Größe der Informationsbasis. Die Ergebnisse sind für jede Wissensbasis über zehn Individuen in jeweils elf Bedingungen (0% bis 100% der Cuewerte bekannt in 10%-Inkrementen) gemittelt.

Zum einen kann ein *Less is More*-Effekt beobachtet werden, zum anderen können Leistungsunterschiede zwischen Algorithmen visuell verdeutlicht werden, die vertikalen Linien bieten die Möglichkeit

zum globalen Leistungsvergleich. Während im Beispiel die einzelnen Kurven Heuristiken darstellen und auf der x-Achse die Abstufungen der Informationsbasis aufgetragen sind, können auch beliebige andere Kombinationen visualisiert werden, wahlweise bei Mittlung über die übrigen Kategorien oder bei Auswahl bestimmter Kategorien.

Abbildung C.57 ermöglicht in diesem Sinne den Vergleich zwischen der Bedingung mit einem geringen Prozentatz bekannter Cuewerte und einem hohen Prozentsatz bekannter Cuewerte. Es wird deutlich, dass der Vorteil von TTB gegenüber den Alternativen bei steigender Anzahl von Cuewerten größer wird. Die Ergebnisse von Minimalist sind zwangsläufig in beiden Bedingungen gleich. Abbildung C.58 ermöglicht den Vergleich zweier Heuristiken hinsichtlich aller simulierten Anteile bekannter Cuewerte. Es zeigt sich, dass der Vorteil von TTB gegenüber Franklin's Rule erst bei 20% auftritt und sich dann vergrößert. Zuletzt ist Abbildung C.59 zu entnehmen, welche Konsequenzen die Anzahl bekannter Städte für die Ergebnisse von Minimalist und Take The Best in Abhängigkeit vom Anteil bekannter Cuewerte hat. Cuewertanteile und Trainings-/Testsplits (im Beispiel wurde immer über Trainingsstichproben von jeweils 60%, 80% und 100% der Städte gemittelt) sind in ihren Inkrementen frei wählbar, die Auswahl der Heuristiken kann aus der umfangreichen Liste frei bestimmt werden und beliebige Datensätze, die für binäre Paarvergleiche geeignet sind, können analysiert werden. Diese skizzenhafte Darstellung soll als Vorstellung des Moduls für Heuristikenvergleiche genügen.

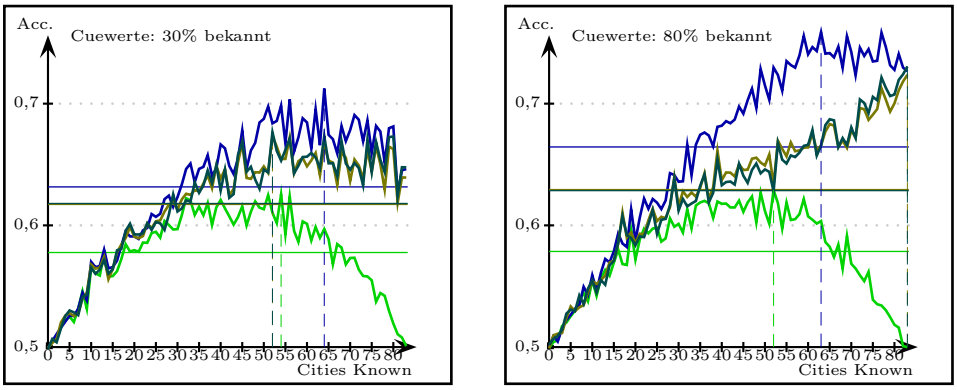


Abbildung C.57: Beispieldarstellung für das Ergebnis eines Vergleichs zwischen Cueprozentwerten: Die beiden Diagramme ermöglichen den Vergleich der Ergebnisse für die Fälle, dass 30% der Cuewerte bekannt sind und dass 80% der Cuewerte bekannt sind.

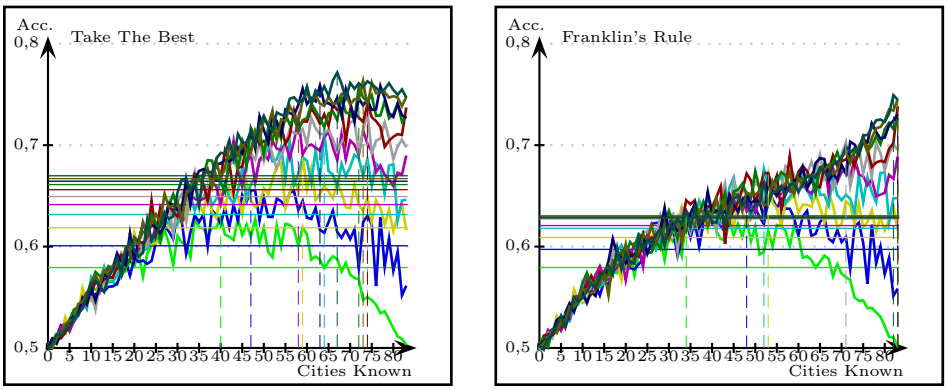


Abbildung C.58: Beispieldarstellung für die Analyse einzelner Heuristiken hinsichtlich der Auswirkung unterschiedlicher bekannter Cuewertanteile: Im Diagramm können Franklin's Rule und Take The Best miteinander verglichen werden. Jede Linie repräsentiert die Ergebnisse für jeden Anteil bekannter Cuewerte von 0% bis 100% in 10%-Inkrementen.

Rekognitionsvalidität

Eine Hilfsanwendung für Studien zur Rekognitionsheuristik kontrolliert die Manipulierbarkeit der Rekognitionsvalidität durch eine *Roulette-Wheel-Selection*-Verfahren. Zum Aufbau einer künstlichen Wissensbasis für simulierte Individuen werden hierbei Städte über den Selektionsoperator genetischer Algorith-

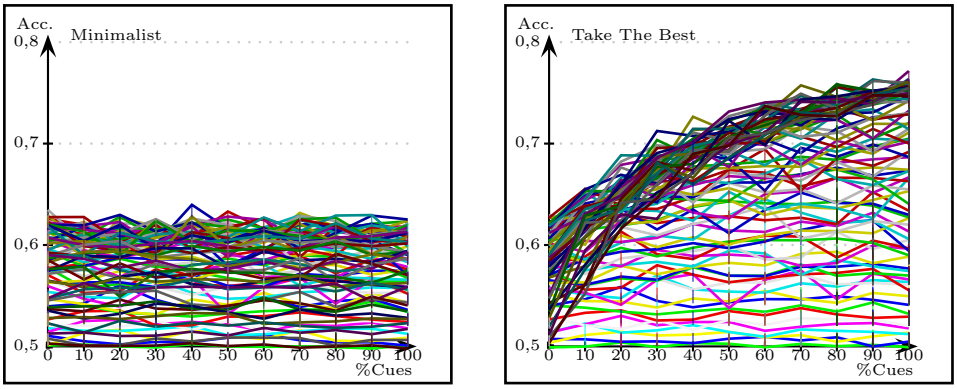


Abbildung C.59: Beispieldarstellung für das Ergebnis eines Vergleichs zwischen Minimalist (links) und Take The Best (rechts) hinsichtlich der Auswirkung der Anzahl bekannter Städte. Für jeden simulierten Anteil bekannter Cuewerte sind die Ergebnisse für jede mögliche Anzahl als Linie eingetragen.

men ausgewählt, wobei das Fitnesskriterium durch Kriteriumsgrößen ersetzt wird, die zusätzlich einer Power-Skalierung unterzogen werden. Für verschiedene Parameterwerte kann dabei die durchschnittliche Rekognitionsvalidität in Abhängigkeit von der Größe der Wissensbasis beobachtet werden. In einer Beispielanwendung mit wenigen Individuen zeigt sich - dargestellt in Abbildung C.60 - der Erfolg der Manipulation: Die erzeugte Rekognitionsvalidität ist abhängig von verschiedenen Paramtersetzungen auf einem unterschiedlichen Niveau, die aber relativ stabil erscheinen über verschieden Wissensbasengrößen.

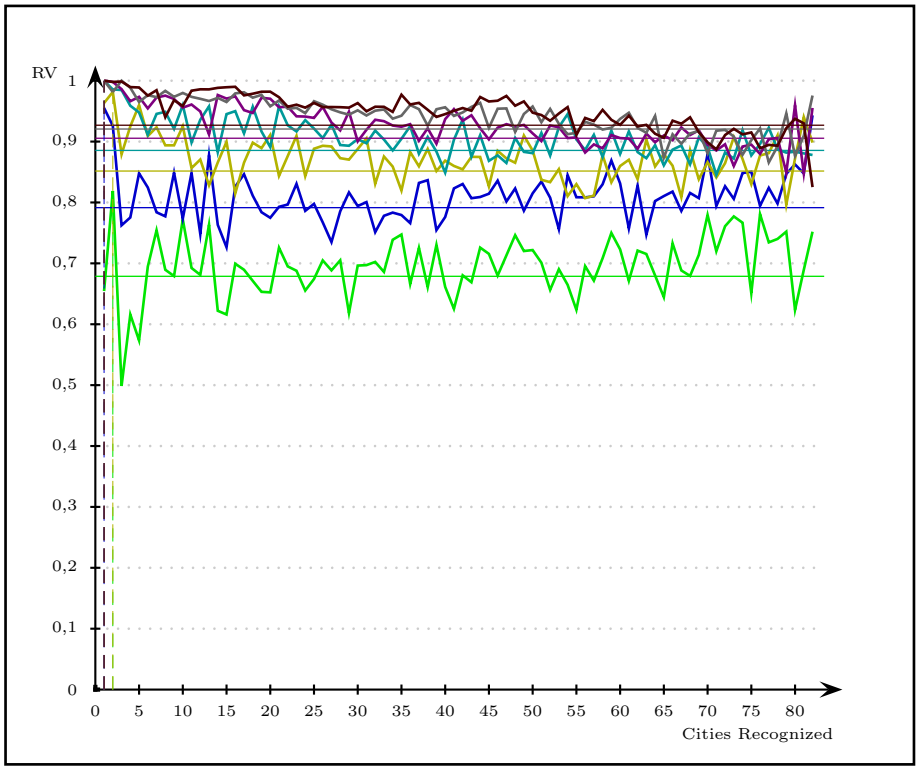


Abbildung C.60: Beispieldiagramm zur Manipulation der Rekognitionsvalidität: Die Differenz zwischen den Linienverläufen zeigt den Erfolg der Manipulation der Rekognitionsvalidität durch verschiedene Parametersetzungen in einem Roulette-Wheel-Selektionsverfahren mit Power-Transformation zur Bestimmung der Informationsbasis.

Analyse lexikografischer Paarvergleiche

Eine Bandbreite von Verfahren ist zur Analyse lexikografischer Paarvergleiche und insbesondere der TTB-Heuristik umgesetzt worden. Dabei sind unter anderem einige Ideen von Martignon und Hoffrage

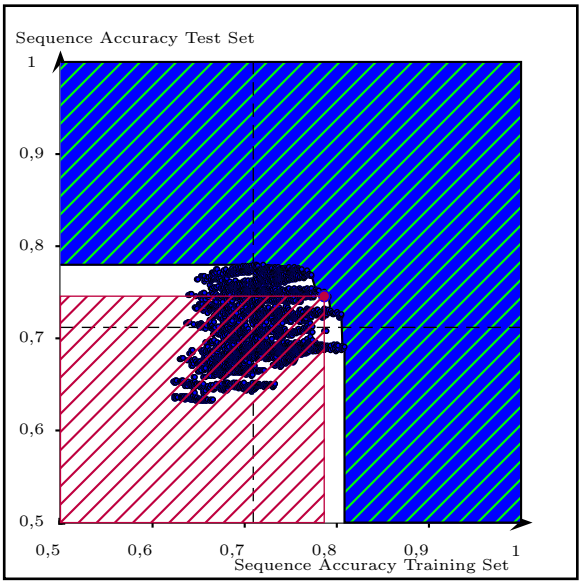


Abbildung C.61: Beispiel für ein TTB-Pareto-Diagramm: Jeder Punkt im Diagramm repräsentiert das Fitting- und Generalisierungsergebnis einer lexikografischen Strategie mit einer bestimmten Cuereihenfolge, das Ergebnis von TTB ist als roter Punkt markiert. Der nicht blau schattierte Bereich entspricht dem durch die paretooptimale Teilmenge der Strategien dominierten Bereich, der rot schattierte Bereich entspricht dem durch das TTB-Ergebnis dominierten Bereich. Die Ergebnisse sind dabei die Ergebnisse bei einer einmaligen Zufallsaufteilung des Datensatzes in Trainings- und Teststichprobe.

(2002) übernommen und nachprogrammiert worden. Zur Untersuchung der Leistungsfähigkeit des TTB-Algorithms in der Suche nach einer geeigneten Cue-Reihenfolge für lexikografische Paarvergleiche kann auf verschiedene Arten demonstriert werden. In Abbildung C.61 sind die *Fitting*- und *Prediction*-Leistungen aller möglichen Cueordnungen für ein Beispiel mit sieben Cues gegeneinander aufgetragen worden. Gleichzeitig ist die pareto-optimale Teilmenge markiert worden, die alle Sequenzen einschließt, die hinsichtlich der *Fitting*- oder der *Prediction*-Leistung von keiner anderen Sequenz auf beiden Dimensionen übertroffen werden. Das Ergebnis von TTB ist farbig markiert und der von der durch TTB ermittelten Sequenz dominierte Bereich an Ergebnissen ist hervorgehoben. In diesem Fall basieren die Ergebnisse auf einem einzigen Trial, in einer erweiterten Version können die Ergebnisse von Sequenzen über beliebig viele Trials gemittelt werden und in zahlreichen unterschiedlichen Diagrammformen visualisiert werden.

Einen Hinweis auf die Auswirkung von Reihenfolgeabweichungen gibt das von Martignon und Hoffrage (2002) entwickelte Sequenzdiagramm, bei dem für Gruppen von Sequenzen, die sich jeweils durch die gleiche Anzahl von notwendigen Verschiebungen von der Sequenz der TTB-Strategie ausgehend auszeichnen, Mittelwert, Minimum, Maximum und Streuung zu erkennen sind. Diese Diagrammform ist ebenfalls umgesetzt worden und Abbildung C.62 zeigt ein Beispiel: Bei wenigen Abweichungen bleibt hier das Ergebnis relativ stabil. Dies impliziert auch, dass sich geringfügige Fehler im Lernen von Cuereihenfolgen nicht drastisch auswirken.

Ebenfalls von Martignon und Hoffrage (2002) entwickelt ist die Grundidee für ein Histogramm der Ergebnisse aller möglichen Sequenzen, dass den Vergleich zwischen der TTB-Lösung und der im *Fitting* optimalen Lösung erlaubt. In der umgesetzten Variante, wird dabei das Diagramm für das *Fitting* eingezeichnet, ermittelt an einer Trainingsstichprobe, die Ergebnisse der Sequenzen werden aber ebenfalls an der verbleibenden Teststichprobe erhoben. So kann die Stabilität der Reihenfolge der Sequenzergebnisse zwischen *Fitting* und *Prediction* abgeschätzt werden und das Ergebnis von TTB an den Ergebnissen aller Sequenzen relativiert werden. Martignon und Hoffrage (2002) halten in ihrer Analyse fest:

> [...] the search principle of Take The Best achieves a satisfying performance, in that improving this performance involves paying too high a price in terms of computational complexity; to determine the optimal ordering for the German city environment takes a UNIX machine two days! (Martignon & Hoffrage, 2002, S. 40)

Hier kann erfreulicherweise festgestellt werden, dass es mittlerweile auch auf handelsüblichen PCs

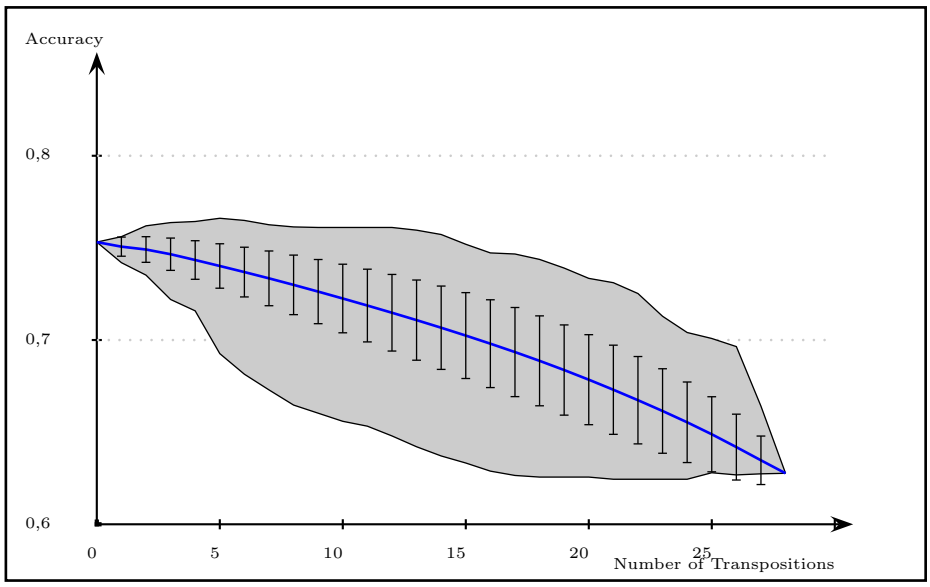


Abbildung C.62: Beispiel für ein TTB-Sequenzinversionsdiagramm: Die Ergebnisse aller Permutationen sind jeweils für Gruppen von Permutationen mit der gleichen Anzahl benötigter Reihenfolgeverschiebungen in Relation zum TTB-Ergebnis zusammengefasst. Dabei sind Minimum, Maximum, Mittelwert und Streuung der Ergebnisse markiert.

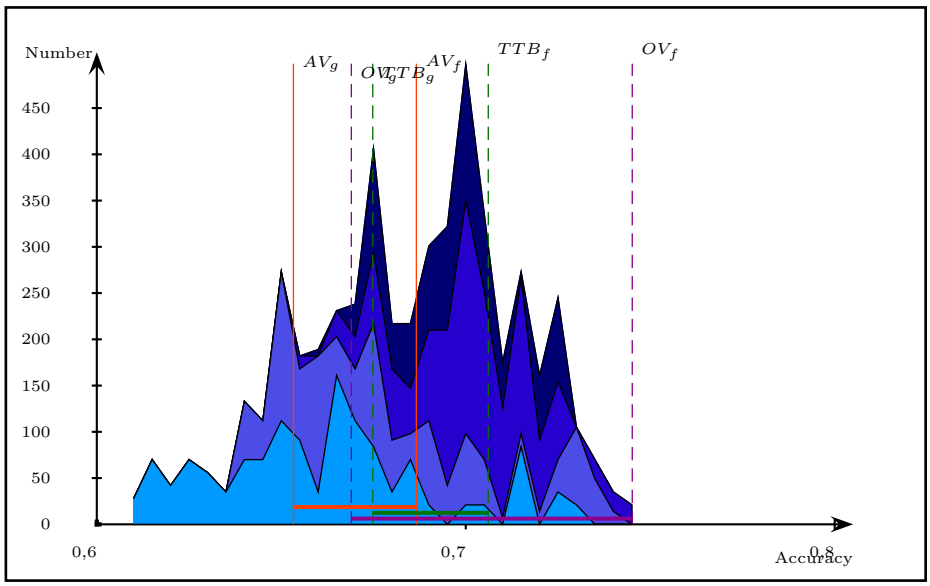


Abbildung C.63: Beispiel für ein Stabilitästhistogramm für lexikografische Paarvergleiche: Die Hauptlinie repräsentiert die Verteilung der Ergebnisse der lexikografischen Paarvergleich mit allen möglichen Cuepermutationen als Reihenfolge für eine zufällig ausgewählte Trainingsstichprobe. Dabei ist das durchschnittliche Ergebnis als AV_f, das Ergebnis von TTB als TTB_f und das beste Ergebnis als OV_f repräsentiert. Die vier Farben entsprechen dabei den jeweiligen Anteilen von Cueordnungen, deren Leistungen im Testdatensatz im untersten, zweituntersten, zweithöchsten und höchsten Viertel aller Ergebnisse liegen. Auch für den Testdatensatz sind das durchschnittliche Ergebnis als AV_g, das Ergebnis der von TTB im Trainingsdatensatz ermittelten Reihenfolge als TTB_g und das Ergebnis der im Trainingsdatensatz optimalen Cueordnung als OV_g repräsentiert, die den Ergebnissen gleicher Cueordnungen in Trainings- und Testdatensatz entsptrechenden vertikalen Linien sind durch horizonale Balken verbunden.

möglich ist, diese Analyse in deutlich weniger als einer Stunde Laufzeit durchzuführen. Damit ist der Überblick über die im Rahmen der Dissertation angefertigten Softwarepakete abgeschlossen.

Except during the nine months before he draws his first breath, no man manages his affairs as well as a tree does.

George Bernard Shaw, „Man and Superman“ (1903)

D

Vollständige Ergebnisse der Fast and Frugal Tree-Simulation

D.1 Vergleichsalgorithmen

Im Folgenden sind die ausführlichen Ergebnisse für Fitting und Prediction aufgelistet für die Algorithmen:

- CART (s. Tabelle D.1 und Tabelle D.2)
- Logistische Regression (s. Tabelle D.3 und Tabelle D.4)
- Context Model ($s = 0,3$, s. Tabelle D.5 und Tabelle D.6)
- Context Model ($s = 0,4$, s. Tabelle D.7 und Tabelle D.8)
- Context Model ($s = 0,5$, s. Tabelle D.9 und Tabelle D.10)

Tabelle D.1: Fitting-Ergebnisse für CART (33 Datensätze): Die Prozentsätze geben den Anteil der Trainingsstichprobe an der Gesamtstichprobe an. Die Zahlen entsprechen der mit 1000 multiplizierten gerundeten durchschnittlichen Klassifikationsgenauigkeit aus 1000 Trials, wobei 1,0 durch „000“ repräsentiert wird.

CART			Fitting																			
Datensatz	n	Cues	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	Ø
Bankruptcy	50	5	-	-	622	963	952	949	956	961	961	959	951	951	946	946	943	945	949	952	965	934
Solvency	52	8	-	-	-	950	936	934	940	945	942	941	939	938	938	937	939	942	944	950	956	942
Echocardiogram	62	7	-	-	704	903	887	888	899	905	908	912	911	912	912	914	911	914	915	917	930	897
Post-Operative	86	8	-	716	812	814	816	815	816	814	813	813	810	808	809	807	807	806	806	805	802	805
Broadway	89	5	-	693	876	873	875	877	879	879	885	884	883	884	882	881	880	879	876	875	876	869
Credit Score	100	6	-	992	991	991	991	993	994	996	997	997	996	995	994	993	993	992	992	991	990	993
Zoo	101	16	-	-	-	000	000	000	000	000	000	000	000	000	000	000	000	000	000	000	000	000
Iris	150	4	673	869	953	972	982	984	983	984	984	983	982	982	980	980	979	979	979	979	980	957
Hepatitis	153	7	755	854	864	877	883	886	887	887	889	891	890	891	891	890	889	888	887	885	883	877
Strikes	163	4	-	865	870	880	886	889	892	893	896	898	902	904	903	902	902	900	898	897	895	893
Wine	178	13	-	982	995	993	991	990	989	987	987	985	984	984	984	983	982	981	979	977	977	985
Baseball CYY	189	9	-	938	934	933	935	936	938	939	940	942	942	943	945	947	947	949	949	949	950	942
Forensic Glass	214	9	937	938	940	943	947	947	948	948	947	947	946	945	947	947	946	947	948	947	946	945
Horse Colic	218	6	-	873	887	886	886	888	886	888	891	892	892	894	895	897	899	898	898	897	896	891
SPECT	267	22	-	890	911	914	913	909	905	903	900	897	895	894	893	892	892	891	891	891	891	898
Heart (Hungaria)	294	10	916	921	915	914	914	914	912	912	911	911	912	912	911	913	912	912	911	912	913	913
Heart (Cleveland)	297	13	910	923	912	912	915	921	923	923	924	923	924	923	923	923	921	921	920	919	915	920
Alcohol	345	5	901	901	903	906	908	908	907	908	908	909	909	908	908	907	907	907	907	906	909	907
Ionosphere	351	33	-	980	974	973	973	971	971	972	973	974	976	977	978	979	980	980	981	981	983	977
Tourism	365	4	879	884	886	887	886	887	886	885	885	884	883	883	883	883	882	882	882	881	881	884
Voting	402	16	974	978	973	970	971	971	974	975	976	976	976	977	977	976	977	977	977	977	976	975
Education	435	4	-	877	878	877	877	875	875	873	873	871	870	869	867	867	866	865	864	864	863	871
Cardiac	558	5	914	917	918	918	919	919	920	920	921	921	921	921	921	922	922	922	923	924	926	920
Football	672	4	835	830	826	822	817	814	810	808	806	804	802	799	797	796	794	792	790	788	788	806
Breast Cancer	683	9	989	983	980	979	980	979	980	981	981	981	981	981	981	981	981	981	981	981	981	981
Diabetes	786	6	913	913	916	919	919	920	921	921	921	922	922	922	922	922	923	922	923	923	923	920
Authorship	841	69	-	994	995	995	994	994	993	993	993	992	992	992	992	992	992	993	993	993	993	993
Baseball HoF	1320	15	980	981	981	982	982	982	983	983	983	983	983	984	983	983	983	983	984	983	983	983
Yeast	1484	8	908	910	910	908	909	908	909	909	909	909	909	909	909	909	909	910	910	910	909	909
Supreme Court	4052	7	966	968	967	967	967	967	967	967	967	967	967	967	967	967	967	967	967	967	967	967
Hochschule	191	14	-	934	922	917	914	912	914	915	915	916	916	916	915	917	917	918	918	917	913	917
Beratung	437	16	-	-	922	924	923	925	926	925	925	925	925	925	924	925	924	925	923	924	923	924
Schule	1993	15	906	907	907	906	907	907	906	906	906	906	906	906	906	906	906	906	906	906	905	906
Ø			897	908	905	926	926	926	927	927	928	928	927	927	927	927	927	926	926	926	927	924

Tabelle D.2: Prediction-Ergebnisse für CART (33 Datensätze): Die Prozentsätze geben den Anteil der Trainingsstichprobe an der Gesamtstichprobe an. Die Zahlen entsprechen der mit 1000 multiplizierten gerundeten durchschnittlichen Klassifikationsgenauigkeit aus 1000 Trials, wobei 1,0 durch „000“ repräsentiert wird.

CART			Prediction																			
Datensatz	n	Cues	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	Ø
Bankruptcy	50	5	-	-	477	792	806	813	809	810	806	805	810	805	811	809	809	811	803	795	773	785
Solvency	52	8	-	-	-	727	738	742	741	730	734	738	737	739	749	740	754	757	760	776	789	747
Echocardiogram	62	7	-	-	679	632	633	645	653	664	675	677	689	682	691	690	689	688	690	683	644	671
Post-Operative	86	8	-	689	611	612	605	613	605	613	610	613	613	612	615	615	619	631	631	648	664	623
Broadway	89	5	-	654	661	671	678	685	686	686	683	687	690	694	692	693	697	698	705	710	706	688
Credit Score	100	6	-	910	955	969	974	974	974	974	974	975	977	976	977	978	978	978	978	981	980	971
Zoo	101	16	-	-	-	968	971	978	981	985	987	992	994	996	997	998	998	999	000	000	000	990
Iris	150	4	634	689	832	877	902	914	922	927	931	937	941	939	943	944	945	945	944	946	939	897
Hepatitis	153	7	741	719	719	724	726	730	731	734	735	736	739	739	737	743	741	745	746	753	757	737
Strikes	163	4	-	681	689	699	701	712	710	710	716	720	717	718	718	718	715	709	710	698	698	708
Wine	178	13	-	845	873	885	892	894	900	902	904	905	906	908	908	908	906	907	907	903	896	897
Baseball CYY	189	9	-	759	776	793	799	806	812	816	816	820	823	822	823	825	829	827	823	828	814	812
Forensic Glass	214	9	728	751	770	778	778	782	786	788	790	792	798	799	800	802	807	812	812	818	824	790
Horse Colic	218	6	-	608	622	642	647	651	652	655	660	663	666	667	673	672	673	670	672	661	657	656
SPECT	267	22	-	736	764	777	788	795	798	800	801	799	799	799	798	799	803	803	804	808	812	793
Heart (Hungaria)	294	10	717	735	743	747	750	754	753	756	758	758	757	759	756	755	758	757	760	759	759	752
Heart (Cleveland)	297	13	680	695	708	716	727	733	737	742	746	745	746	746	747	744	745	743	738	726	707	730
Alcohol	345	5	680	697	698	702	707	709	713	712	711	711	711	711	712	715	712	715	710	710	713	708
Ionosphere	351	33	-	818	844	858	866	867	869	869	871	873	873	876	878	879	881	881	882	882	880	869
Tourism	365	4	677	681	686	687	690	692	692	699	697	698	701	700	701	702	700	701	702	704	703	695
Voting	402	16	928	936	942	947	950	951	952	952	953	954	954	953	954	954	955	955	953	953	951	950
Education	435	4	-	705	710	713	713	712	711	712	711	712	710	710	710	712	711	711	711	710	716	711
Cardiac	558	5	743	740	739	737	736	737	737	734	734	736	738	737	737	738	740	741	742	739	744	738
Football	672	4	585	587	586	588	588	587	587	588	587	586	587	584	584	583	580	579	575	571	557	583
Breast Cancer	683	9	914	928	932	936	939	941	942	943	944	944	946	946	946	948	948	948	949	948	946	941
Diabetes	786	6	679	688	693	697	699	703	704	706	708	708	710	710	712	713	712	714	715	717	718	706
Authorship	841	69	-	911	928	934	941	943	947	948	949	950	951	953	953	953	954	954	955	953	953	946
Baseball HoF	1320	15	914	920	921	921	923	924	924	925	926	926	926	927	927	928	929	929	929	928	929	925
Yeast	1484	8	653	669	674	678	682	685	686	687	690	691	693	693	694	696	696	698	699	698	698	687
Supreme Court	4052	7	954	957	958	959	960	961	961	961	962	962	962	962	962	962	962	963	963	963	963	961
Hochschule	191	14	-	629	638	651	655	660	660	655	660	661	662	659	664	668	664	669	670	672	667	659
Beratung	437	16	-	-	659	660	663	661	662	663	660	661	661	660	660	657	655	652	649	646	642	657
Schule	1993	15	577	581	583	584	585	586	586	587	587	588	589	589	589	590	590	590	591	590	591	587
Ø			738	747	744	766	770	774	775	777	778	780	781	781	782	783	783	784	784	784	781	775

Tabelle D.3: Fitting-Ergebnisse für die logistische Regression (33 Datensätze): Die Prozentsätze geben den Anteil der Trainingsstichprobe an der Gesamtstichprobe an. Die Zahlen entsprechen der mit 1000 multiplizierten gerundeten durchschnittlichen Klassifikationsgenauigkeit aus 1000 Trials, wobei 1,0 durch „000“ repräsentiert wird.

Logistic Regression			Fitting																			
Datensatz	n	Cues	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	Ø
Bankruptcy	50	5	-	-	1000	999	995	994	989	986	980	978	972	970	967	965	962	962	960	958	958	976
Solvency	52	8	-	-	-	1000	1000	1000	999	998	997	994	989	985	980	976	967	959	954	947	944	981
Echocardiogram	62	7	-	-	961	975	954	933	907	891	871	859	849	841	838	836	830	828	827	826	828	874
Post-Operative	86	8	-	808	898	873	825	799	785	775	766	761	755	748	747	746	745	744	744	743	740	778
Broadway	89	5	-	972	936	889	860	837	826	813	811	803	797	797	793	788	790	790	790	788	790	826
Credit Score	100	6	-	531	654	757	801	864	876	922	937	938	948	949	951	950	950	948	949	950	954	879
Zoo	101	16	-	-	-	845	928	969	983	991	997	997	996	997	999	998	999	997	997	998	997	980
Iris	150	4	973	879	827	799	785	782	777	767	762	760	755	753	749	746	743	739	739	737	735	779
Hepatitis	153	7	670	879	885	863	855	844	838	835	833	831	828	829	830	828	827	827	828	827	828	831
Strikes	163	4	-	846	798	782	769	760	749	745	743	737	736	735	733	733	733	730	731	730	727	751
Wine	178	13	-	000	000	000	000	000	000	000	000	000	000	000	000	000	000	000	000	000	000	000
Baseball CYY	189	9	-	993	976	948	922	904	898	892	886	883	881	877	876	874	871	870	868	866	863	897
Forensic Glass	214	9	914	982	968	935	908	893	883	876	872	869	865	863	861	860	859	858	858	859	858	886
Horse Colic	218	6	-	822	767	741	727	726	717	713	712	706	703	703	699	696	694	692	690	690	689	716
SPECT	267	22	-	898	964	960	952	941	928	913	902	892	884	879	876	874	873	873	873	874	876	902
Heart (Hungaria)	294	10	915	954	915	891	879	873	866	862	861	857	855	853	852	851	849	850	849	849	849	870
Heart (Cleveland)	297	13	955	987	958	920	897	882	877	871	868	865	864	861	859	859	857	856	855	853	851	884
Alcohol	345	5	913	836	811	802	799	797	795	794	791	790	790	790	787	786	786	785	786	785	784	800
Ionosphere	351	33	-	993	999	1000	1000	999	998	995	992	987	982	975	970	962	957	952	947	943	941	977
Tourism	365	4	846	795	782	777	773	774	771	767	767	767	768	768	767	768	767	768	767	767	767	775
Voting	402	16	976	000	000	000	000	000	999	999	998	996	995	994	992	990	988	987	986	985	984	993
Education	435	4	-	803	788	784	781	778	776	775	775	774	773	772	772	772	772	772	772	772	773	777
Cardiac	558	5	912	858	847	845	843	841	841	841	842	842	841	841	841	841	841	841	840	841	840	846
Football	672	4	729	704	695	691	688	687	684	683	682	681	681	680	680	680	680	679	678	678	677	686
Breast Cancer	683	9	999	997	993	987	985	980	978	977	975	975	974	973	973	972	972	971	971	970	970	979
Diabetes	786	6	834	802	789	787	785	783	780	780	778	778	777	776	776	776	775	775	775	775	774	783
Authorship	841	69	-	000	000	000	000	000	000	000	000	000	000	000	000	000	000	000	000	000	000	000
Baseball HoF	1320	15	999	985	970	963	959	957	956	955	954	954	954	953	953	952	952	952	952	951	951	959
Yeast	1484	8	703	706	701	697	694	691	690	690	687	687	686	685	684	684	683	682	682	682	681	689
Supreme Court	4052	7	958	959	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958
Hochschule	191	14	-	999	970	909	853	827	808	798	791	784	780	774	771	767	765	761	758	756	754	813
Beratung	437	16	-	-	851	821	810	797	791	786	782	778	776	774	771	771	768	767	767	765	765	785
Schule	1993	15	751	714	703	695	691	688	686	684	683	681	680	679	678	677	676	676	674	674	673	687
Ø			878	882	883	876	869	865	861	859	856	853	851	849	848	847	845	844	843	842	842	857

Tabelle D.4: Prediction-Ergebnisse für die logistische Regression (33 Datensätze): Die Prozentsätze geben den Anteil der Trainingsstichprobe an der Gesamtstichprobe an. Die Zahlen entsprechen der mit 1000 multiplizierten gerundeten durchschnittlichen Klassifikationsgenauigkeit aus 1000 Trials.

Logistic Regression			Prediction																			
Datensatz	n	Cues	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	Ø
Bankruptcy	50	5	-	-	687	720	759	779	799	815	830	834	849	852	858	861	867	870	876	874	870	824
Solvency	52	8	-	-	-	642	698	747	753	776	789	802	804	810	820	813	820	819	822	823	829	785
Echocardiogram	62	7	-	-	601	616	638	646	664	672	681	696	701	710	718	716	727	727	732	734	711	688
Post-Operative	86	8	-	642	542	559	567	581	593	605	620	632	642	657	661	675	683	688	692	702	714	636
Broadway	89	5	-	626	649	680	699	713	722	729	728	735	739	738	742	750	745	749	759	756	752	723
Credit Score	100	6	-	433	538	632	690	755	780	826	846	858	871	882	888	892	895	902	904	912	914	801
Zoo	101	16	-	-	-	768	866	928	956	967	977	982	980	985	989	987	990	987	990	991	988	958
Iris	150	4	630	679	701	710	717	719	723	725	725	727	725	724	724	721	722	727	715	710	713	713
Hepatitis	153	7	514	649	713	742	753	764	772	776	779	784	787	787	787	793	797	795	798	808	808	758
Strikes	163	4	-	662	683	694	699	703	705	706	704	704	705	702	701	703	702	698	700	695	691	698
Wine	178	13	-	798	885	912	925	934	941	945	949	952	953	954	959	958	959	959	958	960	961	937
Baseball CYY	189	9	-	712	746	773	795	808	814	820	824	825	827	830	829	829	832	832	831	835	833	811
Forensic Glass	214	9	621	741	773	791	801	808	814	820	824	825	829	833	833	836	836	839	839	838	846	808
Horse Colic	218	6	-	600	618	633	639	645	648	654	656	658	662	660	663	667	662	665	663	659	663	651
SPECT	267	22	-	548	671	717	741	755	768	779	790	795	800	805	809	810	813	814	818	818	819	770
Heart (Hungaria)	294	10	668	730	762	787	797	803	810	813	813	818	820	823	823	823	825	822	828	830	830	801
Heart (Cleveland)	297	13	592	702	741	767	784	798	805	813	816	819	821	823	828	828	831	827	828	831	829	794
Alcohol	345	5	684	733	750	759	761	765	766	768	770	771	770	771	772	775	773	775	771	771	778	762
Ionosphere	351	33	-	570	762	791	807	819	829	837	847	853	860	862	866	869	871	873	876	883	886	831
Tourism	365	4	685	727	746	752	757	759	760	763	764	765	764	764	765	764	765	765	766	767	765	756
Voting	402	16	816	915	931	940	943	945	947	948	949	951	953	953	954	956	956	958	957	959	960	942
Education	435	4	-	744	751	754	756	758	760	761	760	761	761	763	761	762	764	763	765	766	766	760
Cardiac	558	5	713	785	813	823	830	834	836	837	837	839	840	841	840	840	839	841	843	841	843	827
Football	672	4	623	647	656	660	663	665	667	668	670	671	671	672	672	673	672	674	674	677	672	666
Breast Cancer	683	9	904	933	942	948	953	957	959	960	962	963	964	964	965	966	966	966	968	967	967	957
Diabetes	786	6	723	749	757	762	764	765	767	767	768	768	768	769	768	768	769	769	767	768	769	763
Authorship	841	69	-	875	958	971	976	978	980	982	982	982	983	984	984	984	984	984	984	985	985	975
Baseball HoF	1320	15	893	909	924	931	935	938	940	941	942	943	943	944	944	945	945	945	945	947	946	937
Yeast	1484	8	671	680	680	680	680	680	680	679	678	679	678	679	679	678	678	679	679	677	678	678
Supreme Court	4052	7	957	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958
Hochschule	191	14	-	571	595	617	639	656	668	676	685	693	697	700	701	706	706	709	711	709	700	674
Beratung	437	16	-	-	684	701	711	717	722	726	728	730	732	734	736	736	739	742	739	742	743	727
Schule	1993	15	633	652	658	662	664	665	666	667	666	667	668	667	667	667	667	666	666	665	668	663
Ø			708	713	738	753	769	780	787	793	798	801	804	806	808	809	811	812	813	814	814	789

Tabelle D.5: Fitting-Ergebnisse für das Context Model Ex0.3 (33 Datensätze): Die Prozentsätze geben den Anteil der Trainingsstichprobe an der Gesamtstichprobe an. Die Zahlen entsprechen der mit 1000 multiplizierten gerundeten durchschnittlichen Klassifikationsgenauigkeit aus 1000 Trials.

Context Model (Ex0.3)			Fitting																			
Datensatz	n	Cues	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	Ø
Bankruptcy	50	5	-	700	693	680	682	688	682	683	691	696	685	684	688	681	691	685	686	680	686	651
Solvency	52	8	804	759	726	744	778	767	769	782	787	785	791	793	789	791	796	791	790	798	798	781
Echocardiogram	62	7	752	680	684	691	686	684	685	688	693	706	703	709	703	708	711	714	718	716	722	703
Post-Operative	86	8	718	689	696	693	699	697	709	708	712	715	714	714	718	716	718	719	720	721	721	710
Broadway	89	5	755	690	690	681	683	692	693	700	694	701	701	706	706	711	709	710	717	716	721	704
Credit Score	100	6	695	678	681	675	676	675	672	680	682	683	677	686	686	696	694	699	699	705	709	687
Zoo	101	16	602	570	566	561	560	545	557	559	566	552	554	560	561	547	555	546	554	550	549	559
Iris	150	4	696	648	638	629	624	620	628	615	616	612	611	606	604	597	599	605	598	596	590	617
Hepatitis	153	7	764	768	771	779	780	785	786	787	785	788	790	789	790	790	791	790	791	791	791	785
Strikes	163	4	720	707	710	717	712	713	715	717	715	717	716	720	719	723	723	724	726	728	726	718
Wine	178	13	641	628	617	617	612	619	606	612	613	604	607	604	608	608	610	604	605	607	604	612
Baseball CYY	189	9	-	727	736	738	741	743	740	748	747	745	746	747	748	749	748	749	749	750	751	705
Forensic Glass	214	9	608	591	597	586	594	592	590	593	584	590	588	593	587	582	592	585	583	587	578	589
Horse Colic	218	6	607	588	586	578	584	580	575	578	579	575	580	581	579	577	578	581	581	584	584	582
SPECT	267	22	742	728	727	735	735	738	734	738	740	743	744	745	745	746	747	748	749	749	748	741
Heart (Hungaria)	294	10	642	612	602	592	592	587	581	581	579	578	585	581	581	575	576	575	576	573	573	586
Heart (Cleveland)	297	13	628	620	606	606	604	600	600	591	599	599	597	600	599	602	596	599	605	604	604	603
Alcohol	345	5	761	766	764	763	763	761	761	759	756	756	755	755	753	752	752	750	748	748	746	756
Ionosphere	351	33	608	608	599	594	591	590	592	588	587	591	588	584	584	583	583	585	588	585	586	590
Tourism	365	4	756	766	768	765	767	770	766	767	768	768	767	766	767	768	767	768	767	767	767	767
Voting	402	16	516	500	492	495	501	480	499	490	489	483	485	483	474	489	492	495	481	492	494	491
Education	435	4	575	566	561	560	562	561	560	560	561	558	561	561	560	558	563	561	557	562	560	561
Cardiac	558	5	828	837	839	841	840	841	840	841	841	841	840	841	840	841	840	841	841	841	841	840
Football	672	4	659	666	669	671	671	670	670	672	670	670	671	671	671	672	671	671	672	671	671	670
Breast Cancer	683	9	694	598	573	562	535	537	531	537	531	532	522	535	534	523	531	524	515	499	504	543
Diabetes	786	6	681	681	684	685	685	688	687	687	688	687	688	687	688	688	687	688	687	687	687	686
Authorship	841	69	712	716	721	721	717	722	720	721	723	722	721	724	724	724	725	723	727	725	727	722
Baseball HoF	1320	15	898	900	904	905	905	905	905	905	906	906	906	906	906	906	906	906	906	906	906	905
Yeast	1484	8	673	678	684	686	687	687	687	688	689	688	688	688	688	688	688	688	688	688	688	686
Supreme Court	4052	7	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958
Hochschule	191	14	671	666	664	670	672	678	683	681	687	691	690	695	695	696	698	699	702	703	704	687
Beratung	437	16	676	681	682	685	690	692	694	696	698	701	700	701	703	704	706	705	706	707	709	697
Schule	1993	15	592	601	606	608	610	613	613	614	616	616	618	618	619	620	620	621	621	621	622	614
Ø			698	684	682	681	682	681	682	683	683	684	683	685	684	684	685	685	685	685	686	682

Tabelle D.6: Prediction-Ergebnisse für das Context Model Ex0.3 (33 Datensätze): Die Prozentsätze geben den Anteil der Trainingsstichprobe an der Gesamtstichprobe an. Die Zahlen entsprechen der mit 1000 multiplizierten gerundeten durchschnittlichen Klassifikationsgenauigkeit aus 1000 Trials.

Context Model (Ex0.3)			Prediction																			
Datensatz	n	Cues	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	Ø
Bankruptcy	50	5	-	558	593	611	623	641	651	653	660	664	663	663	664	667	674	661	669	662	681	648
Solvency	52	8	532	589	646	680	726	733	746	767	768	773	782	783	785	787	789	784	787	791	784	738
Echocardiogram	62	7	603	618	639	649	657	664	666	673	680	685	686	691	698	695	702	702	700	710	698	674
Post-Operative	86	8	610	634	654	671	681	684	694	702	712	709	715	715	717	720	721	724	725	719	721	696
Broadway	89	5	604	621	634	644	658	668	672	675	687	687	686	693	698	693	702	709	709	705	700	676
Credit Score	100	6	605	637	647	648	650	655	653	658	667	663	661	672	674	686	682	697	682	701	686	665
Zoo	101	16	522	532	542	546	553	539	543	551	555	548	551	555	555	538	551	547	549	551	541	546
Iris	150	4	580	578	585	588	585	593	601	592	600	594	594	591	593	579	586	591	585	578	572	588
Hepatitis	153	7	727	753	767	773	779	780	783	785	787	788	787	788	790	789	788	791	791	789	791	780
Strikes	163	4	612	653	675	690	697	706	708	711	711	713	712	717	718	719	718	719	722	724	723	703
Wine	178	13	555	582	589	589	594	599	592	597	602	595	601	599	601	607	600	597	603	609	601	595
Baseball CYY	189	9	-	714	723	729	732	736	737	742	741	743	743	742	746	747	743	746	746	750	745	739
Forensic Glass	214	9	565	577	580	580	584	585	583	588	579	587	585	586	581	578	584	583	578	588	576	581
Horse Colic	218	6	532	543	554	556	565	563	563	570	566	570	572	572	575	574	570	575	578	576	585	566
SPECT	267	22	688	695	708	716	723	725	728	733	735	736	739	743	742	743	744	743	744	743	745	730
Heart (Hungaria)	294	10	598	591	583	581	577	577	571	570	571	569	577	571	570	568	570	567	570	564	563	574
Heart (Cleveland)	297	13	597	608	596	594	598	595	595	587	598	595	590	596	594	594	593	596	602	600	605	596
Alcohol	345	5	726	742	745	749	749	749	747	748	746	747	746	744	744	743	742	744	742	744	743	744
Ionosphere	351	33	559	572	579	580	580	578	582	579	580	584	579	579	577	580	579	581	584	579	580	578
Tourism	365	4	742	763	766	768	767	766	768	767	766	767	767	768	768	765	766	764	768	769	767	765
Voting	402	16	494	490	484	487	497	478	496	488	486	481	483	483	471	487	491	494	483	492	487	487
Education	435	4	522	542	538	545	551	554	555	553	556	554	556	558	557	554	560	557	556	557	558	552
Cardiac	558	5	820	836	839	840	840	840	841	840	840	840	841	840	842	840	841	840	839	839	840	839
Football	672	4	642	660	666	668	670	671	671	671	672	672	671	671	671	670	671	671	669	670	671	668
Breast Cancer	683	9	691	593	568	556	531	534	528	536	529	529	521	535	532	522	533	523	514	500	511	541
Diabetes	786	6	660	670	675	678	681	681	683	684	685	685	685	685	687	687	687	685	688	687	689	682
Authorship	841	69	699	708	717	718	715	719	719	719	721	722	720	723	724	723	723	721	725	724	727	719
Baseball HoF	1320	15	890	899	902	904	905	905	906	906	906	906	906	906	906	906	906	906	906	906	907	904
Yeast	1484	8	669	679	682	685	686	686	687	687	687	688	688	688	688	688	687	687	689	689	686	686
Supreme Court	4052	7	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	959	958
Hochschule	191	14	621	637	646	656	663	670	674	676	680	683	687	689	689	693	697	696	700	701	701	677
Beratung	437	16	641	661	670	677	681	686	689	692	695	696	697	699	701	702	702	704	709	704	708	690
Schule	1993	15	582	595	602	605	608	610	611	613	614	615	615	618	619	620	619	620	620	622	618	612
Ø			640	651	659	664	669	671	673	675	677	677	678	679	680	679	681	681	681	682	681	673

Tabelle D.7: Fitting-Ergebnisse für das Context Model Ex0.4 (33 Datensätze): Die Prozentsätze geben den Anteil der Trainingsstichprobe an der Gesamtstichprobe an. Die Zahlen entsprechen der mit 1000 multiplizierten gerundeten durchschnittlichen Klassifikationsgenauigkeit aus 1000 Trials.

Context Model (Ex0.4)			Fitting																			
Datensatz	n	Cues	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	Ø
Bankruptcy	50	5	-	708	692	672	680	691	692	692	692	690	692	688	684	687	690	693	681	675	686	688
Solvency	52	8	800	764	730	744	779	772	778	785	794	791	795	799	795	798	802	798	795	801	803	785
Echocardiogram	62	7	740	695	693	705	701	699	700	706	704	715	712	718	713	718	717	720	724	720	727	712
Post-Operative	86	8	728	702	712	710	712	713	718	718	717	722	719	720	721	719	720	720	720	721	721	717
Broadway	89	5	761	696	690	696	696	703	705	710	700	712	711	715	714	719	716	717	721	721	727	712
Credit Score	100	6	714	699	701	699	701	704	703	709	705	717	716	719	722	722	724	725	729	728	731	714
Zoo	101	16	605	574	573	569	559	556	565	549	550	555	546	548	557	558	562	552	555	550	557	560
Iris	150	4	712	670	655	641	652	649	649	648	649	647	651	644	650	647	652	652	649	649	652	654
Hepatitis	153	7	777	778	777	787	786	789	790	790	788	789	792	791	790	791	791	791	791	791	791	788
Strikes	163	4	729	710	709	715	717	714	719	717	719	717	716	719	720	723	725	727	725	726	726	720
Wine	178	13	644	630	620	614	619	615	622	622	619	615	619	612	611	613	615	620	612	617	616	619
Baseball CYY	189	9	-	745	750	756	753	758	757	762	760	761	762	763	764	765	764	766	767	768	769	760
Forensic Glass	214	9	625	616	620	609	613	602	619	611	607	612	615	612	608	613	609	613	609	609	603	612
Horse Colic	218	6	622	597	589	584	588	587	583	585	586	585	582	582	584	582	582	585	585	581	584	587
SPECT	267	22	759	749	749	757	757	758	759	762	762	764	766	766	765	766	767	767	769	768	769	762
Heart (Hungaria)	294	10	651	623	622	611	599	597	597	597	594	592	593	594	595	589	587	589	590	585	588	600
Heart (Cleveland)	297	13	636	635	612	610	614	602	607	601	606	605	600	599	597	607	602	603	605	600	607	608
Alcohol	345	5	769	763	760	757	756	753	752	750	748	748	747	747	746	745	746	745	745	745	745	751
Ionosphere	351	33	615	615	609	604	599	601	608	597	600	602	601	601	599	599	598	595	598	599	599	602
Tourism	365	4	761	767	768	765	767	770	766	767	768	768	767	766	767	768	767	768	767	767	767	767
Voting	402	16	509	496	491	496	486	502	493	492	487	485	482	482	499	497	497	490	480	486	485	491
Education	435	4	584	570	574	564	563	560	559	560	560	560	563	556	560	557	559	557	565	563	559	563
Cardiac	558	5	838	839	840	841	840	841	840	841	841	841	840	841	840	841	840	841	841	841	841	840
Football	672	4	665	669	671	671	671	670	670	672	670	670	671	671	671	672	671	671	672	671	671	671
Breast Cancer	683	9	717	619	600	566	572	566	550	549	539	528	546	530	535	531	533	532	521	516	518	556
Diabetes	786	6	689	688	688	687	687	689	689	690	689	690	689	690	689	689	689	690	689	689	689	689
Authorship	841	69	718	724	724	728	730	729	728	731	728	728	729	733	733	730	734	734	733	736	734	730
Baseball HoF	1320	15	904	904	906	906	906	906	905	906	906	906	906	906	906	906	906	906	906	906	906	906
Yeast	1484	8	682	685	688	688	688	688	688	688	689	688	688	688	688	688	688	688	688	688	688	688
Supreme Court	4052	7	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958
Hochschule	191	14	695	681	689	692	695	698	703	703	708	710	711	713	714	715	715	716	717	718	719	706
Beratung	437	16	689	697	703	705	708	712	714	715	716	718	717	719	719	720	721	721	721	721	722	714
Schule	1993	15	605	613	619	618	620	622	621	621	623	624	624	624	625	624	625	625	625	626	626	622
Ø			706	693	690	689	690	690	691	691	690	691	692	691	692	693	693	693	692	692	693	692

D.2 Ursprüngliche Baumalgorithmen

Im Folgenden sind die ausführlichen Ergebnisse für Fitting und Prediction aufgelistet für die Algorithmen:

- FaF MaxSens (s. Tabelle D.11 und Tabelle D.12)
- FaF ZigSens (s. Tabelle D.13 und Tabelle D.14)
- FaF MaxVal (s. Tabelle D.15 und Tabelle D.16)
- FaF ZigVal (s. Tabelle D.17 und Tabelle D.18)
- FaF MaxAcc (s. Tabelle D.19 und Tabelle D.20)

Tabelle D.8: Prediction-Ergebnisse für das Context Model Ex0.4 (33 Datensätze): Die Prozentsätze geben den Anteil der Trainingsstichprobe an der Gesamtstichprobe an. Die Zahlen entsprechen der mit 1000 multiplizierten gerundeten durchschnittlichen Klassifikationsgenauigkeit aus 1000 Trials.

Context Model (Ex0.4)			Prediction																			
Datensatz	n	Cues	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	Ø
Bankruptcy	50	5	-	558	593	603	615	634	652	662	649	655	659	660	657	664	666	663	654	662	664	643
Solvency	52	8	534	585	647	682	722	735	751	765	770	776	787	789	787	790	797	791	790	791	793	741
Echocardiogram	62	7	606	615	648	658	670	679	680	688	692	692	696	698	707	708	709	705	711	712	707	683
Post-Operative	86	8	614	648	674	687	697	705	709	714	719	718	720	721	720	725	724	725	726	719	721	705
Broadway	89	5	605	626	644	655	671	678	684	686	695	695	701	701	707	703	712	709	715	706	700	684
Credit Score	100	6	613	654	665	678	678	685	687	695	693	704	707	717	717	717	722	727	722	727	711	696
Zoo	101	16	519	536	548	546	547	543	551	542	547	550	542	539	552	554	560	551	550	549	549	546
Iris	150	4	601	596	602	604	618	623	624	628	637	632	638	630	639	636	644	639	642	633	638	626
Hepatitis	153	7	739	766	778	784	787	787	788	790	791	791	789	790	791	790	789	791	791	790	791	785
Strikes	163	4	601	646	670	685	696	700	710	709	715	712	711	714	717	714	717	718	717	716	718	699
Wine	178	13	567	583	589	594	605	599	611	604	608	608	610	605	605	611	608	612	607	616	609	603
Baseball CYY	189	9	-	730	738	744	748	751	755	757	757	759	761	758	761	763	763	763	767	767	764	756
Forensic Glass	214	9	582	594	600	601	602	595	610	602	607	608	609	608	602	609	603	608	604	604	598	602
Horse Colic	218	6	534	547	555	560	564	570	569	572	572	575	574	574	575	574	578	578	578	581	582	569
SPECT	267	22	697	711	724	734	741	743	749	754	754	755	759	762	759	762	760	762	765	762	765	748
Heart (Hungaria)	294	10	602	598	594	593	584	583	585	583	584	582	584	582	584	581	575	579	581	572	578	584
Heart (Cleveland)	297	13	599	610	597	598	603	596	599	594	602	601	597	593	593	602	600	601	600	592	612	599
Alcohol	345	5	734	745	746	747	747	747	745	746	744	745	745	743	745	744	743	746	744	747	748	745
Ionosphere	351	33	567	581	586	589	587	586	594	590	593	597	597	594	593	590	592	588	595	600	591	590
Tourism	365	4	751	766	767	768	767	766	768	767	766	767	767	768	768	765	766	764	768	769	767	766
Voting	402	16	481	482	483	488	481	495	492	491	482	482	479	481	497	497	494	490	478	487	487	487
Education	435	4	522	533	546	542	550	547	549	552	552	554	556	553	556	554	554	552	564	558	558	550
Cardiac	558	5	834	840	841	840	841	840	841	840	840	840	841	840	842	840	841	840	839	839	840	840
Football	672	4	652	667	669	670	671	671	671	671	672	672	671	671	671	670	671	671	669	670	671	670
Breast Cancer	683	9	704	609	592	557	565	561	548	544	536	527	547	529	532	532	531	528	519	515	517	552
Diabetes	786	6	665	673	677	679	683	682	683	685	685	685	686	687	687	687	687	685	690	689	692	684
Authorship	841	69	705	719	722	725	727	727	726	729	727	727	728	732	733	730	732	734	732	735	732	727
Baseball HoF	1320	15	897	903	905	905	906	906	906	906	906	906	906	906	906	906	906	906	906	906	907	905
Yeast	1484	8	678	686	687	688	688	688	688	688	687	688	688	688	688	688	687	687	689	689	686	687
Supreme Court	4052	7	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	959	958
Hochschule	191	14	633	653	669	678	685	693	696	701	703	705	707	709	711	714	716	714	718	717	716	697
Beratung	437	16	656	680	692	699	703	708	709	712	713	714	715	717	718	718	718	721	723	717	721	708
Schule	1993	15	596	608	613	615	618	619	620	621	621	622	622	623	624	624	624	624	625	626	624	619
Ø			647	658	667	671	676	679	682	683	684	685	687	686	688	689	689	689	689	688	688	680

Tabelle D.9: Fitting-Ergebnisse für das Context Model Ex0.5 (33 Datensätze): Die Prozentsätze geben den Anteil der Trainingsstichprobe an der Gesamtstichprobe an. Die Zahlen entsprechen der mit 1000 multiplizierten gerundeten durchschnittlichen Klassifikationsgenauigkeit aus 1000 Trials.

Context Model (Ex0.5)			Fitting																			
Datensatz	n	Cues	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	Ø
Bankruptcy	50	5	-	709	684	686	679	695	685	684	691	700	692	689	689	688	689	682	676	684	677	688
Solvency	52	8	804	772	735	748	786	776	781	795	798	795	798	804	800	798	806	801	802	807	807	790
Echocardiogram	62	7	741	701	710	715	708	711	706	709	710	717	717	721	714	717	716	718	720	716	720	715
Post-Operative	86	8	740	712	717	716	717	718	722	719	719	723	720	720	721	719	720	720	720	721	721	720
Broadway	89	5	751	713	711	711	706	714	712	717	708	717	714	715	713	719	715	716	718	718	719	716
Credit Score	100	6	725	712	723	713	713	720	712	722	721	727	726	727	729	729	730	729	731	730	731	724
Zoo	101	16	621	596	566	574	559	556	566	550	551	554	553	551	549	553	557	553	547	548	554	561
Iris	150	4	717	679	669	661	662	663	665	662	663	663	663	662	666	666	666	667	666	668	667	668
Hepatitis	153	7	776	783	782	791	788	791	792	791	789	790	792	791	790	791	791	791	791	791	791	789
Strikes	163	4	728	708	706	713	716	712	717	714	718	718	720	722	722	723	724	722	723	725	723	719
Wine	178	13	667	634	627	630	633	627	624	626	619	623	627	620	624	619	626	624	622	622	621	627
Baseball CYY	189	9	-	755	762	766	763	767	766	770	769	770	769	772	772	771	772	772	772	772	773	768
Forensic Glass	214	9	641	625	636	631	638	632	634	636	634	631	639	641	636	638	644	636	634	636	633	635
Horse Colic	218	6	626	604	595	592	590	591	586	588	590	586	586	587	586	584	588	587	586	588	588	591
SPECT	267	22	784	777	777	782	785	787	788	790	791	796	797	796	797	801	801	799	801	801	802	792
Heart (Hungaria)	294	10	653	637	637	624	616	612	610	610	611	614	611	614	607	602	605	606	604	603	605	615
Heart (Cleveland)	297	13	652	637	616	615	626	608	612	604	609	607	601	603	605	604	600	605	607	604	608	612
Alcohol	345	5	767	757	753	750	749	746	747	745	746	745	745	746	745	745	745	745	745	745	745	748
Ionosphere	351	33	623	627	621	615	617	614	616	617	617	610	613	610	613	612	610	611	610	611	609	614
Tourism	365	4	762	767	768	765	767	770	766	767	768	768	767	766	767	768	767	768	767	767	767	767
Voting	402	16	529	512	507	500	489	496	490	500	500	488	487	497	494	495	491	485	489	492	492	496
Education	435	4	584	572	569	566	562	558	559	559	562	561	563	565	564	561	566	565	565	561	563	564
Cardiac	558	5	841	839	840	841	840	841	840	841	841	841	840	841	840	841	840	841	841	841	841	840
Football	672	4	668	669	671	672	672	670	670	672	670	670	671	671	671	672	671	671	672	671	671	671
Breast Cancer	683	9	733	650	633	613	598	603	601	589	593	591	579	571	580	567	563	572	549	540	524	592
Diabetes	786	6	690	684	683	683	681	683	683	681	682	681	682	682	683	684	683	684	684	684	684	683
Authorship	841	69	726	733	733	739	738	740	738	742	741	741	743	744	743	745	742	745	744	745	746	740
Baseball HoF	1320	15	907	905	906	906	906	906	905	906	906	906	906	906	906	906	906	906	906	906	906	906
Yeast	1484	8	686	686	689	689	688	688	688	688	689	688	688	688	688	688	688	688	688	688	688	688
Supreme Court	4052	7	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958
Hochschule	191	14	707	698	707	710	709	714	717	717	719	721	720	721	722	721	721	722	722	722	722	716
Beratung	437	16	708	716	719	721	724	723	725	724	725	727	726	726	726	727	727	726	726	727	727	724
Schule	1993	15	617	621	623	622	622	623	624	623	623	623	624	623	623	623	623	624	623	623	623	623
Ø			714	701	698	697	697	697	697	697	698	698	698	699	698	698	698	698	697	697	697	699

Tabelle D.10: Prediction-Ergebnisse für das Context Model Ex0.5 (33 Datensätze): Die Prozentsätze geben den Anteil der Trainingsstichprobe an der Gesamtstichprobe an. Die Zahlen entsprechen der mit 1000 multiplizierten gerundeten durchschnittlichen Klassifikationsgenauigkeit aus 1000 Trials.

Context Model (Ex0.5)			Prediction																			
Datensatz	n	Cues	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	Ø
Bankruptcy	50	5	-	551	577	598	602	623	633	633	642	654	648	653	656	656	657	655	644	664	637	632
Solvency	52	8	531	585	641	675	726	735	746	769	773	775	787	788	791	789	799	796	797	792	792	741
Echocardiogram	62	7	605	627	659	668	680	689	690	697	702	699	701	699	710	711	709	705	706	712	705	688
Post-Operative	86	8	622	662	684	699	709	710	714	719	722	719	722	722	721	725	724	725	726	719	721	709
Broadway	89	5	602	634	658	662	676	691	693	694	704	703	705	706	712	702	710	710	712	705	700	688
Credit Score	100	6	626	667	686	693	696	710	703	714	716	721	723	728	728	729	730	734	725	729	711	709
Zoo	101	16	519	539	538	545	540	542	551	542	540	546	550	543	539	551	550	547	541	551	546	543
Iris	150	4	609	610	623	631	634	645	649	649	658	656	656	659	663	663	667	663	668	660	661	649
Hepatitis	153	7	751	774	785	787	790	790	790	791	792	792	789	790	792	790	789	791	791	790	791	787
Strikes	163	4	591	633	655	677	688	693	703	702	709	710	712	714	714	715	713	712	716	716	717	694
Wine	178	13	563	587	597	606	606	609	608	609	607	611	620	612	614	611	617	617	618	614	620	608
Baseball CYY	189	9	-	745	753	758	758	763	765	766	767	769	770	768	769	771	772	771	774	774	767	765
Forensic Glass	214	9	595	604	617	621	630	625	627	632	632	627	635	634	629	636	636	632	633	634	635	627
Horse Colic	218	6	531	549	557	563	566	572	571	577	574	578	577	577	578	578	581	577	577	585	581	571
SPECT	267	22	717	733	748	755	764	766	773	777	780	782	785	786	788	793	793	792	797	792	801	775
Heart (Hungaria)	294	10	602	607	606	599	597	597	595	596	600	601	603	604	599	597	598	602	600	598	600	600
Heart (Cleveland)	297	13	599	611	600	597	608	596	602	597	604	597	596	598	597	597	596	600	604	596	603	600
Alcohol	345	5	739	746	747	745	745	745	744	745	745	745	745	744	745	744	743	746	744	747	748	745
Ionosphere	351	33	573	590	596	597	603	604	602	606	609	605	604	603	607	608	605	606	603	609	599	602
Tourism	365	4	756	767	767	768	767	766	768	767	766	767	767	768	768	765	766	764	768	769	767	766
Voting	402	16	504	498	495	495	482	490	485	496	495	487	484	495	493	492	486	485	488	489	489	491
Education	435	4	515	528	534	537	543	542	545	544	552	552	553	559	556	554	564	559	557	549	565	548
Cardiac	558	5	839	841	841	840	841	840	841	840	840	840	841	840	842	840	841	840	839	839	840	840
Football	672	4	657	668	670	671	671	671	672	671	672	672	671	671	671	670	671	671	669	670	671	670
Breast Cancer	683	9	716	632	613	597	584	590	591	578	584	585	573	565	576	564	562	571	548	539	522	584
Diabetes	786	6	663	670	671	673	675	674	675	674	676	676	678	679	678	680	680	679	682	682	683	676
Authorship	841	69	715	727	731	736	735	738	737	742	739	741	741	742	742	744	742	744	744	745	747	739
Baseball HoF	1320	15	902	905	906	906	906	906	906	906	906	906	906	906	906	906	906	906	906	906	907	906
Yeast	1484	8	684	688	688	688	688	688	688	688	687	688	688	688	688	688	687	687	689	689	686	688
Supreme Court	4052	7	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	959	958
Hochschule	191	14	646	673	688	699	705	711	713	715	717	717	719	719	719	722	723	720	723	722	720	709
Beratung	437	16	674	701	711	715	718	720	722	723	724	724	725	725	726	726	725	728	729	725	727	719
Schule	1993	15	606	616	618	619	621	620	621	622	621	621	622	621	622	622	622	623	623	624	622	620
Ø			652	664	673	678	682	686	687	689	691	692	693	693	694	694	695	694	694	694	692	686

Tabelle D.11: Fitting-Ergebnisse für das FaF MaxSens (33 Datensätze): Die Prozentsätze geben den Anteil der Trainingsstichprobe an der Gesamtstichprobe an. Die Zahlen entsprechen der mit 1000 multiplizierten gerundeten durchschnittlichen Klassifikationsgenauigkeit aus 1000 Trials.

FaF MaxSens			Fitting																			
Datensatz	n	Cues	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	Ø
Bankruptcy	50	5	-	681	614	628	620	602	619	618	611	616	621	618	621	610	622	625	626	597	608	620
Solvency	52	8	752	720	587	604	612	580	582	587	590	584	591	591	581	579	586	573	568	563	556	599
Echocardiogram	62	7	793	670	594	598	625	633	656	661	669	679	680	693	692	694	702	706	713	709	712	678
Post-Operative	86	8	769	718	711	706	715	716	721	719	718	722	720	720	721	719	720	720	720	721	721	721
Broadway	89	5	798	690	595	600	632	631	657	665	673	672	680	684	693	699	703	707	710	708	712	679
Credit Score	100	6	761	658	680	710	715	737	733	747	738	750	742	753	747	756	750	758	752	759	753	737
Zoo	101	16	675	585	534	496	489	480	457	452	444	441	435	424	423	419	413	410	408	407	406	463
Iris	150	4	609	496	493	487	472	445	448	447	449	431	439	440	425	411	435	421	431	428	433	455
Hepatitis	153	7	629	606	628	657	641	645	630	638	622	601	599	587	583	571	563	547	535	532	530	597
Strikes	163	4	658	696	701	708	707	708	711	711	709	712	712	711	710	712	712	712	712	712	712	707
Wine	178	13	511	502	492	479	481	475	469	471	458	468	458	458	445	455	444	437	443	435	441	464
Baseball CYY	189	9	-	422	361	331	336	317	317	314	314	314	310	312	315	309	308	311	310	303	297	322
Forensic Glass	214	9	604	669	710	738	754	758	767	771	773	775	775	776	779	777	778	778	777	777	776	753
Horse Colic	218	6	578	531	524	509	502	508	498	500	494	500	485	489	484	484	475	473	472	459	453	496
SPECT	267	22	410	360	340	364	363	339	357	356	347	345	350	342	339	354	340	344	353	348	358	353
Heart (Hungaria)	294	10	650	654	656	652	654	654	660	662	663	663	663	666	665	665	666	666	666	666	666	661
Heart (Cleveland)	297	13	620	585	575	563	562	553	552	549	550	547	544	545	545	545	543	544	544	544	546	556
Alcohol	345	5	619	517	484	505	465	456	468	448	461	434	425	445	417	416	418	399	401	376	374	449
Ionosphere	351	33	659	735	746	748	747	750	750	751	750	748	750	749	749	749	749	750	750	749	749	744
Tourism	365	4	513	461	416	401	390	390	387	368	358	352	339	327	319	320	315	313	308	300	282	361
Voting	402	16	674	683	647	639	635	618	621	611	609	610	611	611	607	608	609	615	622	625	632	625
Education	435	4	656	640	619	608	597	594	593	586	583	573	568	559	554	550	540	536	525	518	515	575
Cardiac	558	5	331	332	355	368	402	427	435	456	458	474	480	488	496	493	500	502	519	519	538	451
Football	672	4	589	597	609	616	619	624	623	628	631	630	633	631	631	632	634	634	633	632	633	624
Breast Cancer	683	9	798	824	827	831	834	837	839	839	838	838	839	838	840	839	837	838	836	833	825	833
Diabetes	786	6	570	570	556	534	534	524	529	524	525	527	527	532	533	532	535	536	537	538	538	537
Authorship	841	69	208	205	207	206	205	206	206	206	206	206	206	205	206	206	206	206	206	206	206	206
Baseball HoF	1320	15	171	159	157	156	156	157	157	158	158	157	157	157	157	157	157	157	158	158	160	158
Yeast	1484	8	688	686	689	689	688	688	688	688	689	688	688	688	688	688	688	688	688	688	688	688
Supreme Court	4052	7	934	955	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	957
Hochschule	191	14	419	437	434	423	440	458	465	465	481	479	489	505	501	519	523	529	541	548	559	485
Beratung	437	16	373	351	340	352	350	347	348	348	348	344	346	342	342	341	338	335	337	333	329	344
Schule	1993	15	556	558	561	562	561	565	566	570	581	577	585	587	593	600	603	605	610	613	614	582
Ø			599	574	557	558	560	557	560	560	559	558	558	558	556	557	557	556	557	553	554	560

Tabelle D.12: Prediction-Ergebnisse für FaF MaxSens (33 Datensätze): Die Prozentsätze geben den Anteil der Trainingsstichprobe an der Gesamtstichprobe an. Die Zahlen entsprechen der mit 1000 multiplizierten gerundeten durchschnittlichen Klassifikationsgenauigkeit aus 1000 Trials.

FaF MaxSens			Prediction																			
Datensatz	n	Cues	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	Ø
Bankruptcy	50	5		525	595	617	620	612	635	629	632	618	644	644	646	621	640	650	643	575	584	618
Solvency	52	8	487	562	596	607	617	610	616	613	611	612	611	606	616	613	600	595	593	561	569	594
Echocardiogram	62	7	614	550	512	550	584	575	610	620	643	640	653	652	671	664	680	678	678	686	703	630
Post-Operative	86	8	675	651	661	677	704	712	717	719	720	718	721	722	721	725	724	725	726	719	721	708
Broadway	89	5	615	549	514	542	585	591	614	624	637	628	654	653	668	669	680	682	680	683	698	630
Credit Score	100	6	655	615	680	706	728	738	744	744	749	746	753	752	760	756	753	760	771	759	762	733
Zoo	101	16	539	519	502	479	471	470	455	445	440	437	430	424	424	418	413	414	412	402	416	448
Iris	150	4	476	438	450	445	440	422	425	422	423	412	417	418	408	399	414	406	410	421	406	424
Hepatitis	153	7	522	534	583	616	602	608	594	601	585	572	568	557	549	543	542	534	529	523	525	562
Strikes	163	4	632	673	687	692	698	702	702	705	708	709	709	711	713	712	710	712	709	713	713	700
Wine	178	13	489	493	495	476	478	473	464	469	454	464	453	455	448	453	442	441	438	434	435	461
Baseball CYY	189	9		400	348	324	330	315	310	306	306	307	302	310	309	296	298	303	299	289	283	313
Forensic Glass	214	9	580	653	702	728	746	755	764	768	773	776	777	777	776	777	776	777	778	775	781	749
Horse Colic	218	6	497	486	491	484	485	495	487	485	482	487	477	479	474	473	463	463	465	445	449	477
SPECT	267	22	332	319	308	333	335	320	331	338	328	329	338	330	327	345	333	333	343	336	359	333
Heart (Hungaria)	294	10	606	631	641	646	651	654	657	659	661	663	664	662	665	667	665	666	667	665	666	656
Heart (Cleveland)	297	13	548	535	524	519	526	521	525	524	531	533	527	537	535	539	541	543	544	550	553	535
Alcohol	345	5	562	497	468	494	460	452	465	445	461	434	423	447	416	415	418	399	404	375	374	443
Ionosphere	351	33	651	730	746	749	750	749	749	748	749	750	748	749	749	750	752	748	748	749	749	743
Tourism	365	4	422	397	378	368	364	367	362	349	342	335	324	312	304	309	303	303	295	289	276	337
Voting	402	16	669	679	644	637	632	617	616	608	604	606	608	608	606	606	608	614	618	625	627	623
Education	435	4	606	606	583	574	568	568	568	561	561	549	543	533	529	528	520	521	510	512	516	550
Cardiac	558	5	285	296	324	344	380	406	416	437	440	457	463	472	481	481	485	490	504	505	516	431
Football	672	4	533	555	579	593	603	609	611	615	621	620	625	621	623	622	624	627	624	628	623	608
Breast Cancer	683	9	799	817	822	825	830	835	837	836	835	837	837	836	838	836	835	835	834	830	824	830
Diabetes	786	6	558	561	548	528	527	522	525	524	523	524	525	530	530	531	534	535	538	535	535	533
Authorship	841	69	206	206	206	206	206	206	206	206	206	205	206	206	206	206	206	206	206	206	207	206
Baseball HoF	1320	15	167	156	156	155	155	155	155	156	157	157	156	156	156	156	156	158	157	160	162	157
Yeast	1484	8	688	688	688	688	688	688	688	688	687	688	688	688	688	688	687	687	689	689	686	688
Supreme Court	4052	7	934	955	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	959	957
Hochschule	191	14	376	382	387	382	398	417	427	430	449	448	461	480	474	493	503	510	521	539	551	454
Beratung	437	16	339	331	328	340	341	339	343	342	345	342	344	340	339	340	338	335	335	334	328	338
Schule	1993	15	542	553	555	560	560	562	564	569	581	575	583	586	593	600	603	605	611	614	613	580
Ø			536	532	535	541	546	546	550	550	552	550	551	552	552	551	552	552	553	548	551	547

Tabelle D.13: Fitting-Ergebnisse für das FaF ZigSens (33 Datensätze): Die Prozentsätze geben den Anteil der Trainingsstichprobe an der Gesamtstichprobe an. Die Zahlen entsprechen der mit 1000 multiplizierten gerundeten durchschnittlichen Klassifikationsgenauigkeit aus 1000 Trials.

FaF ZigSens			Fitting																			
Datensatz	n	Cues	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	Ø
Bankruptcy	50	5	-	747	801	817	818	809	835	843	839	840	844	845	846	844	839	839	838	838	826	828
Solvency	52	8	752	756	827	830	815	841	844	838	841	841	839	840	839	836	839	837	835	843	841	828
Echocardiogram	62	7	817	773	692	695	682	661	679	662	689	675	690	685	705	696	709	711	715	715	720	704
Post-Operative	86	8	814	687	649	624	607	594	584	580	569	560	561	552	551	547	541	540	538	533	516	587
Broadway	89	5	817	776	690	695	687	658	683	665	686	673	692	684	700	697	709	714	714	715	719	704
Credit Score	100	6	819	801	788	778	777	772	769	767	765	767	763	761	762	761	760	759	761	760	761	771
Zoo	101	16	800	777	764	751	749	742	730	727	719	720	709	685	706	697	678	673	683	669	676	719
Iris	150	4	663	601	612	612	598	592	591	594	595	605	607	606	616	624	622	637	646	662	659	618
Hepatitis	153	7	816	777	762	755	749	744	741	737	732	731	729	726	724	721	719	715	714	713	713	738
Strikes	163	4	766	749	744	738	734	732	732	732	728	729	728	728	726	725	726	725	726	724	720	732
Wine	178	13	728	778	768	783	788	781	786	780	784	780	783	779	779	777	779	776	774	776	773	777
Baseball CYY	189	9	-	629	626	615	619	612	614	609	610	603	605	600	598	594	587	589	582	562	553	600
Forensic Glass	214	9	827	795	792	789	790	790	790	790	788	788	788	787	788	787	786	785	783	782	781	790
Horse Colic	218	6	708	697	673	681	679	659	671	669	678	669	681	681	685	692	706	711	713	732	743	691
SPECT	267	22	787	757	746	744	742	740	736	736	736	736	737	735	735	736	735	736	739	740	744	742
Heart (Hungaria)	294	10	688	668	646	625	614	598	595	577	571	567	564	560	557	559	557	558	558	558	558	588
Heart (Cleveland)	297	13	668	667	639	620	630	611	619	613	615	617	610	616	613	620	617	621	621	623	624	624
Alcohol	345	5	664	620	612	603	595	592	592	587	587	583	580	580	576	572	570	566	564	559	557	587
Ionosphere	351	33	743	747	740	737	735	736	737	735	734	731	732	731	730	729	728	729	727	726	725	733
Tourism	365	4	539	498	479	468	462	460	457	453	448	446	442	439	435	434	431	429	426	422	414	452
Voting	402	16	943	955	958	957	961	961	963	963	963	963	963	963	963	963	964	964	964	965	965	961
Education	435	4	707	698	689	680	685	687	691	691	699	693	700	697	703	712	719	731	735	747	751	706
Cardiac	558	5	751	729	721	717	715	714	713	714	716	715	714	716	714	715	715	714	714	713	710	717
Football	672	4	668	648	642	641	637	636	634	635	634	633	633	632	632	631	631	630	630	629	629	636
Breast Cancer	683	9	916	899	893	888	882	881	876	874	871	868	867	864	865	863	863	862	860	857	853	874
Diabetes	786	6	679	663	657	648	647	643	644	640	639	639	638	638	638	637	637	637	637	637	637	644
Authorship	841	69	619	613	621	616	622	620	617	617	613	616	611	606	605	604	603	599	604	602	602	611
Baseball HoF	1320	15	561	562	563	564	566	568	569	569	569	569	569	570	570	569	569	569	568	568	567	567
Yeast	1484	8	601	587	579	575	571	571	569	568	567	566	567	566	566	565	565	566	566	566	566	571
Supreme Court	4052	7	547	541	538	538	538	537	537	538	537	537	537	537	537	537	537	537	537	537	537	538
Hochschule	191	14	814	760	728	709	696	686	681	673	670	669	664	662	662	658	655	655	653	651	652	684
Beratung	437	16	746	716	694	687	680	678	676	671	672	672	670	670	670	667	667	665	664	662	657	678
Schule	1993	15	623	615	613	612	611	612	610	611	612	611	612	613	613	613	613	613	613	613	614	613
Ø			729	705	695	691	687	682	684	681	681	679	680	678	679	678	678	679	679	679	678	685

Tabelle D.14: Prediction-Ergebnisse für FaF ZigSens (33 Datensätze): Die Prozentsätze geben den Anteil der Trainingsstichprobe an der Gesamtstichprobe an. Die Zahlen entsprechen der mit 1000 multiplizierten gerundeten durchschnittlichen Klassifikationsgenauigkeit aus 1000 Trials.

FaF ZigSens			Prediction																			
Datensatz	n	Cues	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	Ø
Bankruptcy	50	5	-	623	704	740	762	783	802	810	816	822	826	823	822	826	822	827	821	813	813	792
Solvency	52	8	536	556	662	700	687	751	764	759	764	774	777	776	776	772	777	777	771	786	781	734
Echocardiogram	62	7	542	534	503	542	556	543	575	569	610	596	632	618	657	645	677	704	693	722	726	613
Post-Operative	86	8	556	489	478	475	472	472	472	468	465	466	460	455	453	448	436	431	428	408	341	456
Broadway	89	5	547	538	503	540	559	548	587	577	606	590	633	614	656	649	686	702	699	709	726	614
Credit Score	100	6	722	759	766	770	770	768	769	765	765	763	763	761	761	759	760	763	757	758	745	760
Zoo	101	16	661	707	718	715	722	721	714	708	705	706	696	673	697	684	667	669	671	660	678	693
Iris	150	4	507	519	540	547	550	552	556	558	562	576	576	580	591	605	596	615	623	650	645	576
Hepatitis	153	7	727	722	717	716	712	711	708	710	709	705	704	704	706	705	707	710	709	710	706	710
Strikes	163	4	683	713	720	720	722	724	721	722	724	721	721	721	721	720	716	718	714	715	714	717
Wine	178	13	616	707	738	754	766	770	772	771	774	774	773	775	773	774	773	771	768	771	766	757
Baseball CYY	189	9	-	570	585	587	589	597	599	596	596	590	589	591	588	579	575	574	562	541	534	580
Forensic Glass	214	9	763	781	786	788	788	787	788	787	788	788	787	788	784	784	783	783	784	779	786	784
Horse Colic	218	6	559	585	589	607	614	605	621	624	633	627	648	646	653	661	678	683	685	713	730	640
SPECT	267	22	674	698	706	710	714	715	716	718	717	719	719	719	720	721	725	726	729	731	739	717
Heart (Hungaria)	294	10	569	590	602	592	590	579	580	567	568	562	560	560	559	557	559	557	556	560	562	570
Heart (Cleveland)	297	13	555	580	577	571	586	572	590	585	595	596	587	600	605	611	611	621	626	620	628	596
Alcohol	345	5	543	558	563	569	568	570	571	567	570	567	564	565	562	558	558	553	557	552	556	562
Ionosphere	351	33	713	724	726	726	726	725	725	724	724	725	723	724	723	722	725	720	721	720	719	723
Tourism	365	4	406	406	412	410	411	414	413	413	411	410	405	402	399	399	394	390	387	381	387	403
Voting	402	16	905	934	945	948	954	957	957	959	959	961	961	961	961	962	962	961	962	966	964	955
Education	435	4	686	684	679	667	675	675	679	682	692	683	691	691	695	704	712	727	729	744	755	697
Cardiac	558	5	659	674	683	691	695	698	700	701	705	705	706	707	709	709	709	712	709	709	708	699
Football	672	4	613	615	621	624	626	627	627	627	628	628	628	627	627	627	626	627	627	632	626	626
Breast Cancer	683	9	890	884	882	879	876	875	872	870	866	865	863	861	862	859	859	857	856	854	850	867
Diabetes	786	6	619	631	636	633	634	635	636	636	635	635	635	635	634	634	634	633	633	631	630	633
Authorship	841	69	569	591	608	605	616	615	612	615	610	614	607	601	601	602	599	595	599	592	592	602
Baseball HoF	1320	15	554	555	558	561	563	565	566	565	567	568	567	567	568	568	566	566	566	564	568	564
Yeast	1484	8	542	549	551	551	552	554	553	554	554	554	554	556	556	558	560	561	561	563	565	555
Supreme Court	4052	7	539	538	537	537	538	537	537	537	537	537	538	537	537	537	537	537	537	537	536	537
Hochschule	191	14	624	625	627	623	627	632	632	635	638	637	641	640	641	645	649	646	649	655	659	638
Beratung	437	16	626	631	633	638	642	643	646	647	649	650	654	652	655	654	653	657	655	652	653	647
Schule	1993	15	598	605	605	608	607	608	608	609	610	610	611	611	612	613	614	613	614	615	613	610
Ø			623	633	641	647	651	652	657	656	659	658	661	659	663	662	664	666	665	667	667	655

Tabelle D.15: Fitting-Ergebnisse für das FaF MaxVal (33 Datensätze): Die Prozentsätze geben den Anteil der Trainingsstichprobe an der Gesamtstichprobe an. Die Zahlen entsprechen der mit 1000 multiplizierten gerundeten durchschnittlichen Klassifikationsgenauigkeit aus 1000 Trials, wobei 1,0 durch „000“ repräsentiert wird.

FaF MaxVal			Fitting																			
Datensatz	n	Cues	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	Ø
Bankruptcy	50	5	-	862	773	747	718	735	683	670	668	692	646	648	638	675	629	625	625	679	613	685
Solvency	52	8	752	864	806	760	740	693	678	660	655	641	638	633	618	613	612	595	592	592	585	670
Echocardiogram	62	7	824	863	817	771	744	739	731	725	725	727	728	726	722	722	722	725	723	724	724	746
Post-Operative	86	8	848	773	754	739	728	724	725	722	719	723	720	720	721	719	720	720	720	721	721	734
Broadway	89	5	823	875	812	769	742	740	732	733	724	726	725	723	722	724	723	723	722	725	724	747
Credit Score	100	6	867	823	809	792	783	771	765	763	751	751	745	742	736	735	733	731	731	730	731	763
Zoo	101	16	969	993	996	999	999	999	000	000	000	000	000	000	000	000	000	000	000	000	000	998
Iris	150	4	760	750	714	696	694	706	688	684	676	689	671	670	673	685	671	677	679	693	692	693
Hepatitis	153	7	831	800	788	792	789	791	792	791	789	790	792	791	790	791	791	791	791	791	791	793
Strikes	163	4	786	741	733	735	732	732	734	735	738	739	742	744	747	749	757	763	766	768	766	748
Wine	178	13	821	664	653	628	618	615	608	607	606	605	602	602	604	602	602	602	601	601	601	623
Baseball CYY	189	9	-	794	793	796	792	792	790	790	788	788	787	787	784	785	784	782	781	781	779	787
Forensic Glass	214	9	812	769	761	755	750	750	747	746	742	743	740	739	741	738	737	735	733	733	728	747
Horse Colic	218	6	736	674	648	641	632	623	621	615	616	610	612	610	610	608	610	610	609	611	610	627
SPECT	267	22	804	793	793	795	795	795	793	795	794	794	795	794	794	795	794	794	794	794	794	795
Heart (Hungaria)	294	10	816	771	753	738	727	720	714	707	701	696	690	680	677	674	667	663	658	651	645	702
Heart (Cleveland)	297	13	787	734	717	711	704	709	709	714	716	716	718	723	725	732	734	736	742	746	751	728
Alcohol	345	5	789	786	788	786	788	788	791	789	788	789	789	791	790	790	790	789	790	789	789	789
Ionosphere	351	33	773	762	751	748	748	750	750	751	750	748	750	749	749	749	749	750	750	749	749	751
Tourism	365	4	772	764	761	756	757	756	752	751	752	751	750	749	750	750	749	749	748	747	749	753
Voting	402	16	807	763	725	706	686	676	657	656	648	643	643	640	640	639	639	639	639	639	639	670
Education	435	4	718	693	685	682	671	670	663	663	654	654	648	639	638	635	634	626	626	615	610	654
Cardiac	558	5	845	840	839	839	838	838	838	839	838	838	838	838	838	839	839	839	840	840	840	839
Football	672	4	684	674	673	672	670	669	669	669	668	668	669	669	670	670	670	670	670	670	670	671
Breast Cancer	683	9	911	875	869	865	861	860	856	856	855	852	851	849	850	852	850	854	855	860	866	860
Diabetes	786	6	681	677	677	677	677	679	679	680	679	679	679	678	679	679	679	679	678	678	678	679
Authorship	841	69	793	795	793	795	795	794	794	794	794	794	794	795	794	794	794	794	794	794	794	794
Baseball HoF	1320	15	905	902	902	901	900	899	898	898	898	898	898	897	897	897	897	897	897	896	895	898
Yeast	1484	8	690	687	689	689	689	688	688	688	689	688	688	688	688	688	688	688	688	688	688	688
Supreme Court	4052	7	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958
Hochschule	191	14	780	737	727	725	721	722	723	722	723	724	723	724	724	723	722	723	723	723	723	727
Beratung	437	16	732	730	728	728	728	727	728	727	728	728	727	728	728	728	728	727	727	728	728	728
Schule	1993	15	616	613	614	613	613	614	614	613	614	613	614	614	614	614	614	614	614	614	614	614
Ø			796	782	767	758	751	749	744	743	741	741	739	738	737	738	736	735	735	737	735	747

Tabelle D.16: Prediction-Ergebnisse für FaF MaxVal (33 Datensätze): Die Prozentsätze geben den Anteil der Trainingsstichprobe an der Gesamtstichprobe an. Die Zahlen entsprechen der mit 1000 multiplizierten gerundeten durchschnittlichen Klassifikationsgenauigkeit aus 1000 Trials.

FaF MaxVal			Prediction																			
Datensatz	n	Cues	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	Ø
Bankruptcy	50	5	-	575	539	547	538	579	535	530	525	561	517	516	512	548	500	480	463	521	408	522
Solvency	52	8	487	582	588	574	576	548	545	545	547	529	532	533	514	516	522	500	495	522	534	536
Echocardiogram	62	7	611	615	644	678	695	705	706	714	717	713	712	713	721	721	720	713	718	721	719	698
Post-Operative	86	8	623	648	682	694	707	709	713	717	721	719	722	721	721	725	724	725	726	719	721	707
Broadway	89	5	613	619	641	674	693	698	709	709	718	715	717	719	720	715	720	721	720	717	711	697
Credit Score	100	6	710	742	751	747	748	743	746	741	738	734	735	733	730	729	730	733	725	728	711	734
Zoo	101	16	743	882	927	944	952	961	966	972	975	977	979	982	981	984	988	987	990	991	993	957
Iris	150	4	612	663	655	654	658	672	663	659	655	670	652	649	655	666	660	660	668	679	676	659
Hepatitis	153	7	762	780	788	789	791	790	790	790	792	792	789	790	791	790	789	791	791	790	791	788
Strikes	163	4	632	653	665	678	686	691	700	702	712	710	714	720	723	724	733	742	744	748	742	706
Wine	178	13	631	593	608	600	596	603	603	602	603	603	604	604	599	603	602	598	604	599	604	603
Baseball CYY	189	9	-	771	774	777	779	780	780	778	779	779	778	774	775	774	773	771	771	768	758	774
Forensic Glass	214	9	698	719	724	729	726	730	726	727	728	729	726	724	726	725	722	725	722	726	727	724
Horse Colic	218	6	572	591	599	602	601	600	601	601	600	600	600	599	601	601	603	601	598	601	603	599
SPECT	267	22	784	794	794	793	794	794	795	793	794	794	793	793	794	792	795	795	794	794	793	793
Heart (Hungaria)	294	10	672	695	696	698	696	697	691	687	686	681	677	666	666	666	660	656	652	648	642	675
Heart (Cleveland)	297	13	602	611	618	629	628	643	646	656	662	663	664	673	674	685	688	688	694	701	720	660
Alcohol	345	5	754	774	779	782	783	785	786	788	789	789	789	789	789	788	789	790	790	790	792	785
Ionosphere	351	33	695	740	745	749	750	749	749	748	749	750	748	749	749	750	752	748	748	749	749	746
Tourism	365	4	730	735	736	736	736	736	737	737	736	736	739	739	738	735	737	732	737	740	742	736
Voting	402	16	740	728	702	689	674	665	651	650	644	642	639	638	638	637	636	635	638	637	636	659
Education	435	4	639	646	644	645	640	642	637	637	629	630	622	614	613	611	612	605	600	586	581	623
Cardiac	558	5	835	836	837	837	837	837	838	837	837	837	838	837	839	838	838	838	838	837	839	837
Football	672	4	655	662	664	665	666	667	667	667	668	668	668	668	668	667	669	669	668	670	670	667
Breast Cancer	683	9	862	850	850	848	848	847	845	846	845	844	842	840	841	842	841	846	846	849	855	847
Diabetes	786	6	663	672	675	678	678	678	679	678	679	679	679	680	679	679	679	678	679	680	681	677
Authorship	841	69	794	794	794	794	794	794	794	794	794	795	794	794	794	794	794	794	794	794	793	794
Baseball HoF	1320	15	898	897	897	896	896	896	896	895	895	895	894	895	895	895	894	894	894	895	894	895
Yeast	1484	8	687	687	687	687	688	688	688	688	687	688	688	688	688	688	687	687	689	689	686	688
Supreme Court	4052	7	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	959	958
Hochschule	191	14	690	714	721	721	723	723	722	723	722	721	723	722	721	723	724	721	724	723	720	720
Beratung	437	16	723	727	728	728	728	728	728	728	728	727	728	728	728	727	726	729	730	726	728	727
Schule	1993	15	613	614	614	614	614	614	614	614	613	614	613	613	614	614	614	614	614	615	613	614
Ø			700	714	719	722	723	726	724	725	725	725	723	723	723	725	724	722	722	724	721	722

D.3 Neue Baumalgorithmen

Im Folgenden sind die ausführlichen Ergebnisse für Fitting und Prediction aufgelistet für die Algorithmen:

- FaF ZigSens Plus (s. Tabelle D.21 und Tabelle D.22)
- FaF ZigSens Majority (s. Tabelle D.23 und Tabelle D.24)
- FaF ZigSens Fold (s. Tabelle D.25 und Tabelle D.26)
- FaF ZigVal Plus (s. Tabelle D.27 und Tabelle D.28)
- FaF ZigVal Majority (s. Tabelle D.29 und Tabelle D.30)
- FaF ZigVal Fold (s. Tabelle D.31 und Tabelle D.32)

Tabelle D.17: Fitting-Ergebnisse für das FaF ZigVal (33 Datensätze): Die Prozentsätze geben den Anteil der Trainingsstichprobe an der Gesamtstichprobe an. Die Zahlen entsprechen der mit 1000 multiplizierten gerundeten durchschnittlichen Klassifikationsgenauigkeit aus 1000 Trials.

FaF ZigVal			Fitting																			
Datensatz	n	Cues	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	Ø
Bankruptcy	50	5	-	804	809	821	820	809	838	841	841	842	844	845	846	844	840	839	838	837	826	832
Solvency	52	8	752	825	836	834	841	841	843	837	841	842	838	840	839	837	840	836	835	843	840	834
Echocardiogram	62	7	781	806	785	730	710	702	699	691	689	686	681	681	682	677	676	673	672	669	668	703
Post-Operative	86	8	745	710	691	669	659	646	640	628	622	617	609	604	604	599	600	598	604	606	615	635
Broadway	89	5	780	818	771	733	714	701	696	688	691	682	683	681	682	677	679	671	671	668	667	703
Credit Score	100	6	885	873	869	867	861	862	856	859	850	851	844	845	837	838	834	833	832	831	830	850
Zoo	101	16	965	973	959	954	940	938	929	926	919	919	911	906	906	903	898	892	891	887	886	921
Iris	150	4	666	599	611	595	573	545	546	536	533	516	531	525	522	503	527	528	532	521	534	550
Hepatitis	153	7	838	800	769	752	739	726	721	717	707	702	700	695	696	694	688	686	684	685	686	720
Strikes	163	4	800	773	772	768	765	763	764	763	762	764	763	762	760	760	760	760	756	757	751	764
Wine	178	13	783	783	768	784	786	776	784	776	781	776	780	776	775	771	774	772	770	771	769	777
Baseball CYY	189	9	-	689	653	637	651	642	636	635	639	641	642	642	643	645	645	646	646	646	646	646
Forensic Glass	214	9	874	850	844	839	832	830	827	828	827	820	823	821	819	818	818	816	813	813	812	828
Horse Colic	218	6	781	749	745	745	747	743	745	747	749	749	748	748	748	748	747	745	743	741	735	747
SPECT	267	22	787	731	696	696	690	682	679	676	674	673	670	667	663	664	663	661	660	659	659	682
Heart (Hungaria)	294	10	860	845	839	832	831	829	825	825	823	821	819	817	816	814	814	812	812	812	812	824
Heart (Cleveland)	297	13	869	837	821	815	814	811	808	808	807	806	804	804	803	803	802	800	799	798	794	811
Alcohol	345	5	691	627	612	603	591	589	587	583	581	576	574	575	570	567	566	560	557	554	553	585
Ionosphere	351	33	872	826	801	794	785	780	778	771	770	769	762	761	761	755	750	749	746	742	735	774
Tourism	365	4	539	491	473	463	455	453	452	448	446	444	441	440	440	440	439	441	443	445	451	455
Voting	402	16	945	941	941	942	945	947	949	950	951	951	951	952	952	953	953	954	954	954	954	949
Education	435	4	743	728	723	722	719	719	718	716	717	717	718	717	718	719	719	721	721	722	719	721
Cardiac	558	5	724	693	676	660	660	649	645	642	636	632	628	625	622	618	612	610	604	599	599	639
Football	672	4	686	663	654	653	647	647	645	645	644	645	645	644	644	644	644	644	643	642	642	648
Breast Cancer	683	9	964	956	952	952	951	951	950	949	950	949	949	948	949	948	948	948	947	947	945	950
Diabetes	786	6	687	669	660	653	651	648	647	644	643	642	642	642	641	639	639	639	639	639	639	647
Authorship	841	69	699	666	655	646	640	636	635	633	630	630	630	625	626	625	622	622	618	615	611	635
Baseball HoF	1320	15	542	536	533	535	537	537	537	535	534	532	532	531	530	528	528	528	528	528	528	533
Yeast	1484	8	592	584	579	579	572	570	566	562	559	557	551	551	547	546	541	539	536	534	535	558
Supreme Court	4052	7	587	597	598	581	570	573	569	562	559	553	545	537	535	528	522	519	519	520	523	552
Hochschule	191	14	838	794	752	733	719	701	691	684	679	672	665	660	660	657	653	652	651	650	649	693
Beratung	437	16	779	731	708	695	686	681	679	675	673	674	675	674	674	674	676	677	677	680	683	688
Schule	1993	15	674	654	647	643	640	637	636	634	634	633	633	632	631	632	631	631	631	631	632	638
Ø			765	746	733	725	719	714	713	710	708	706	704	702	701	699	698	697	696	695	695	712

Tabelle D.18: Prediction-Ergebnisse für FaF ZigVal (33 Datensätze): Die Prozentsätze geben den Anteil der Trainingsstichprobe an der Gesamtstichprobe an. Die Zahlen entsprechen der mit 1000 multiplizierten gerundeten durchschnittlichen Klassifikationsgenauigkeit aus 1000 Trials.

FaF ZigVal			Prediction																			
Datensatz	n	Cues	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	Ø
Bankruptcy	50	5	-	607	706	746	766	791	805	810	815	820	824	823	822	826	822	827	820	814	812	792
Solvency	52	8	535	630	731	746	764	769	774	775	778	774	778	777	774	777	781	775	777	783	776	751
Echocardiogram	62	7	548	581	561	576	581	573	591	596	603	605	616	618	622	629	624	635	629	642	648	604
Post-Operative	86	8	534	504	504	503	501	498	500	496	494	499	494	489	489	487	483	490	513	514	539	502
Broadway	89	5	550	585	561	576	584	586	596	591	610	600	614	610	621	626	629	632	631	634	645	604
Credit Score	100	6	689	736	779	790	809	807	814	813	815	812	816	810	812	809	812	810	805	806	791	797
Zoo	101	16	794	874	886	897	892	896	897	897	891	895	890	888	889	886	885	878	878	868	886	882
Iris	150	4	506	514	532	521	516	497	503	494	494	484	494	489	490	470	488	488	488	478	470	496
Hepatitis	153	7	668	680	690	692	691	689	687	687	682	682	683	681	677	677	684	681	682	680	684	683
Strikes	163	4	686	723	737	742	747	750	750	753	753	754	751	753	753	749	748	750	743	745	736	743
Wine	178	13	615	711	748	757	767	768	771	768	772	770	770	771	770	770	769	768	767	770	766	756
Baseball CYY	189	9	-	619	610	612	622	626	621	625	628	630	629	638	641	638	641	641	640	641	640	630
Forensic Glass	214	9	724	763	781	793	798	800	802	803	806	806	806	808	806	806	806	802	807	802	807	796
Horse Colic	218	6	592	649	674	696	710	717	725	730	732	736	738	738	739	741	735	739	736	727	738	715
SPECT	267	22	599	611	612	625	630	631	636	639	638	639	639	639	639	635	640	639	636	635	644	632
Heart (Hungaria)	294	10	738	780	793	799	803	803	805	805	806	807	806	808	808	808	806	808	809	809	811	801
Heart (Cleveland)	297	13	710	739	751	756	760	764	767	766	769	770	769	770	770	770	769	768	766	755	761	760
Alcohol	345	5	569	563	564	570	566	567	567	564	568	564	561	564	558	557	557	551	556	547	550	561
Ionosphere	351	33	695	707	709	713	719	720	722	720	722	725	718	721	723	717	717	716	712	714	703	715
Tourism	365	4	401	401	408	406	408	411	411	413	413	412	413	410	412	410	415	412	416	424	435	412
Voting	402	16	897	916	925	933	935	941	943	945	947	947	949	949	951	951	952	951	952	955	953	942
Education	435	4	691	698	699	697	699	700	699	701	702	701	703	703	703	704	708	705	702	704	706	701
Cardiac	558	5	604	610	610	607	614	609	612	610	607	606	603	601	600	598	593	593	592	590	590	603
Football	672	4	600	608	619	623	628	630	632	633	635	635	637	637	638	639	639	638	641	645	640	631
Breast Cancer	683	9	927	935	939	942	943	944	945	945	944	945	945	945	946	946	946	946	946	946	945	943
Diabetes	786	6	627	636	640	637	638	639	640	640	638	638	638	637	637	637	637	638	638	635	636	637
Authorship	841	69	633	628	627	626	626	623	624	623	620	621	622	618	621	619	617	619	614	607	604	621
Baseball HoF	1320	15	531	527	526	531	532	532	533	531	531	530	529	528	528	527	525	526	524	524	526	528
Yeast	1484	8	526	546	551	557	554	554	552	550	548	546	541	541	537	535	533	531	526	523	528	541
Supreme Court	4052	7	574	590	591	578	567	571	567	559	557	552	545	537	534	528	521	519	519	520	522	550
Hochschule	191	14	584	588	598	599	608	607	609	612	616	614	618	621	622	624	623	627	625	628	626	613
Beratung	437	16	611	629	636	642	645	647	651	650	654	655	656	659	659	663	663	668	671	669	671	653
Schule	1993	15	602	613	617	620	621	623	624	624	625	627	626	626	627	628	627	627	629	630	628	623
Ø			631	652	664	670	674	675	678	678	679	679	679	679	679	678	679	679	678	678	679	673

Tabelle D.19: Fitting-Ergebnisse für das FaF MaxAcc (33 Datensätze): Die Prozentsätze geben den Anteil der Trainingsstichprobe an der Gesamtstichprobe an. Die Zahlen entsprechen der mit 1000 multiplizierten gerundeten durchschnittlichen Klassifikationsgenauigkeit aus 1000 Trials.

FaF MaxAcc			Fitting																			
Datensatz	n	Cues	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	Ø
Bankruptcy	50	5	-	670	737	707	702	684	670	663	656	655	640	639	634	625	623	619	619	616	609	654
Solvency	52	8	752	663	734	719	672	687	678	656	647	650	638	634	632	626	619	606	606	595	581	652
Echocardiogram	62	7	787	752	728	738	730	724	722	719	719	724	725	725	721	722	722	725	723	724	724	729
Post-Operative	86	8	755	724	728	723	721	719	723	720	719	723	720	720	721	719	720	720	720	721	721	723
Broadway	89	5	793	757	730	738	726	725	725	725	718	722	722	722	721	723	723	723	722	725	724	730
Credit Score	100	6	759	757	753	742	739	737	732	733	731	733	731	730	730	730	730	729	731	730	731	736
Zoo	101	16	764	771	703	709	673	675	656	644	636	630	615	617	607	610	603	606	601	601	596	648
Iris	150	4	725	699	707	694	689	699	687	681	673	686	670	669	673	685	671	677	679	693	692	687
Hepatitis	153	7	804	783	777	786	781	788	789	790	787	789	791	791	790	791	791	791	791	791	791	789
Strikes	163	4	757	725	716	699	698	700	698	697	701	702	701	699	701	701	702	705	707	708	709	707
Wine	178	13	697	700	693	655	648	649	635	628	619	617	609	607	606	604	602	603	601	601	601	630
Baseball CYY	189	9	-	792	793	796	792	792	790	790	788	788	787	787	784	785	784	782	781	781	779	787
Forensic Glass	214	9	736	727	730	722	726	718	718	724	719	720	728	725	727	731	734	737	738	736	733	728
Horse Colic	218	6	653	633	610	600	590	586	576	571	565	564	553	553	547	543	532	527	525	509	502	565
SPECT	267	22	834	795	786	788	787	788	790	793	794	793	795	794	794	795	794	794	794	794	794	795
Heart (Hungaria)	294	10	666	644	647	642	641	637	639	640	639	639	639	641	640	639	639	639	639	639	640	642
Heart (Cleveland)	297	13	621	586	571	565	559	554	551	548	546	542	541	540	540	540	539	539	539	538	538	552
Alcohol	345	5	784	786	788	786	788	788	791	789	788	789	789	791	790	790	790	789	790	789	789	789
Ionosphere	351	33	698	678	667	662	656	655	653	650	649	646	646	645	645	644	643	644	644	645	645	653
Tourism	365	4	772	764	761	756	757	756	752	751	752	751	750	749	750	750	749	749	748	747	749	753
Voting	402	16	832	841	843	848	860	864	859	862	867	868	868	871	873	878	877	878	877	876	870	864
Education	435	4	637	615	607	604	595	594	593	587	582	582	577	573	567	560	549	540	531	520	515	575
Cardiac	558	5	845	840	839	839	838	838	838	839	838	838	838	838	838	839	839	839	840	840	840	839
Football	672	4	670	663	665	664	664	664	666	668	667	668	669	669	670	670	670	670	670	670	670	668
Breast Cancer	683	9	731	735	734	738	738	737	738	738	739	740	740	741	740	740	740	740	740	740	741	738
Diabetes	786	6	679	677	677	677	677	679	679	680	679	679	679	678	679	679	679	679	678	678	678	678
Authorship	841	69	793	795	793	795	795	794	794	794	794	794	794	795	794	794	794	794	794	794	794	794
Baseball HoF	1320	15	905	902	902	901	900	899	898	898	898	898	898	897	897	897	897	897	897	896	895	898
Yeast	1484	8	688	686	689	689	688	688	688	688	689	688	688	688	688	688	688	688	688	688	688	688
Supreme Court	4052	7	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958
Hochschule	191	14	785	736	729	725	721	721	722	722	723	724	723	724	724	723	722	723	723	723	723	727
Beratung	437	16	743	732	727	728	727	727	728	727	728	728	727	728	728	728	728	727	727	728	728	729
Schule	1993	15	605	596	595	598	596	600	603	603	607	608	611	609	610	611	612	613	614	614	614	606
Ø			749	733	731	727	722	722	719	718	716	716	714	714	713	713	711	711	710	709	708	719

Tabelle D.20: Prediction-Ergebnisse für FaF MaxAcc (33 Datensätze): Die Prozentsätze geben den Anteil der Trainingsstichprobe an der Gesamtstichprobe an. Die Zahlen entsprechen der mit 1000 multiplizierten gerundeten durchschnittlichen Klassifikationsgenauigkeit aus 1000 Trials.

FaF MaxAcc			Prediction																			
Datensatz	n	Cues	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	Ø
Bankruptcy	50	5	-	484	542	558	558	582	552	552	541	570	539	535	531	543	523	520	510	559	582	543
Solvency	52	8	487	531	553	553	553	546	548	546	544	530	528	526	512	508	492	469	461	410	336	507
Echocardiogram	62	7	615	685	709	707	711	717	716	720	721	716	714	714	721	721	721	713	718	721	719	709
Post-Operative	86	8	691	710	714	719	720	722	720	721	723	719	722	722	721	725	724	725	726	719	721	719
Broadway	89	5	618	685	705	705	712	715	715	715	721	717	718	720	721	715	720	721	720	717	711	709
Credit Score	100	6	589	684	710	724	725	728	730	729	730	727	729	731	730	730	730	734	725	729	711	717
Zoo	101	16	562	639	621	645	629	636	624	626	617	617	610	608	601	604	600	598	594	602	584	611
Iris	150	4	551	584	618	622	630	648	647	644	647	661	645	645	654	665	660	660	668	679	676	642
Hepatitis	153	7	714	734	757	771	772	782	784	788	789	790	788	790	791	790	789	791	791	790	791	778
Strikes	163	4	610	640	654	649	657	663	668	666	673	674	677	679	678	678	676	682	678	682	688	667
Wine	178	13	582	578	605	597	597	607	613	607	605	607	606	605	600	604	602	598	604	599	604	601
Baseball CYY	189	9	-	773	775	777	779	780	780	778	779	779	778	774	775	774	773	771	771	768	758	774
Forensic Glass	214	9	624	666	682	687	693	690	692	699	695	700	706	702	705	710	713	717	719	720	719	697
Horse Colic	218	6	510	538	540	545	543	547	544	539	534	540	533	531	530	525	515	514	518	498	501	529
SPECT	267	22	749	753	752	765	772	775	786	790	792	791	793	793	794	792	795	795	794	794	793	782
Heart (Hungaria)	294	10	640	641	640	639	639	641	640	639	640	640	641	638	639	641	640	641	642	640	638	640
Heart (Cleveland)	297	13	555	558	553	554	549	550	546	544	541	541	542	541	538	537	540	537	538	544	545	545
Alcohol	345	5	765	775	779	782	783	785	786	788	789	789	789	789	789	788	789	790	790	790	792	786
Ionosphere	351	33	606	621	634	642	643	642	641	642	643	645	642	643	644	643	646	641	644	644	641	639
Tourism	365	4	720	733	736	736	736	736	737	737	736	736	739	739	738	735	737	732	737	740	742	736
Voting	402	16	731	781	798	812	829	838	834	839	845	847	849	853	857	862	862	864	866	867	859	836
Education	435	4	554	563	569	572	570	572	571	567	565	565	562	560	555	550	542	535	528	517	517	554
Cardiac	558	5	835	836	837	837	837	837	838	837	837	837	838	837	839	838	838	838	838	837	839	837
Football	672	4	615	635	644	650	654	658	662	665	666	668	668	668	668	667	669	669	668	670	670	660
Breast Cancer	683	9	735	737	738	738	738	739	739	740	740	740	740	739	739	739	740	741	742	740	740	739
Diabetes	786	6	668	673	675	678	678	678	679	678	679	679	679	680	679	679	679	678	679	680	681	678
Authorship	841	69	794	794	794	794	794	794	794	794	794	795	794	794	794	794	794	794	794	794	793	794
Baseball HoF	1320	15	898	897	897	896	896	896	896	895	895	895	894	895	895	895	894	894	894	895	894	895
Yeast	1484	8	688	688	688	688	688	688	688	688	687	688	688	688	688	688	687	687	689	689	686	688
Supreme Court	4052	7	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	959	958
Hochschule	191	14	659	682	702	709	713	719	719	721	721	721	723	722	721	723	724	721	724	723	720	714
Beratung	437	16	697	718	722	725	726	726	728	727	728	727	728	728	728	727	726	729	730	726	728	725
Schule	1993	15	554	564	570	579	582	587	593	594	601	602	606	606	608	610	611	612	614	615	613	596
Ø			664	683	693	697	699	702	702	702	702	703	702	702	701	702	700	699	699	699	696	697

Tabelle D.21: Fitting-Ergebnisse für das FaF ZigSens Plus (33 Datensätze): Die Prozentsätze geben den Anteil der Trainingsstichprobe an der Gesamtstichprobe an. Die Zahlen entsprechen der mit 1000 multiplizierten gerundeten durchschnittlichen Klassifikationsgenauigkeit aus 1000 Trials.

FaF ZigSens Plus			Fitting																			
Datensatz	n	Cues	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	Ø
Bankruptcy	50	5	-	747	772	791	806	811	835	836	843	844	850	847	847	846	846	843	844	840	824	826
Solvency	52	8	742	751	769	789	790	817	824	824	828	830	832	830	838	834	831	837	834	829	828	813
Echocardiogram	62	7	798	791	727	710	705	700	713	705	719	712	719	717	726	722	728	727	732	726	727	727
Post-Operative	86	8	769	725	713	709	714	717	721	719	719	718	722	721	720	719	721	721	722	720	721	722
Broadway	89	5	816	773	774	785	793	796	799	802	800	802	805	808	805	808	809	810	808	808	809	801
Credit Score	100	6	805	723	717	724	716	730	719	734	727	738	728	738	731	737	732	738	733	739	734	734
Zoo	101	16	791	768	761	759	752	744	735	732	723	717	722	702	710	690	696	681	689	684	675	723
Iris	150	4	676	608	620	600	599	598	592	592	597	601	594	614	619	630	623	633	642	663	661	619
Hepatitis	153	7	729	684	665	657	652	640	621	622	613	599	587	576	572	567	553	543	535	530	529	604
Strikes	163	4	725	732	728	729	727	728	727	726	724	727	728	727	727	727	727	727	724	724	721	727
Wine	178	13	723	770	765	783	786	779	786	783	783	780	783	779	778	777	778	776	774	777	772	775
Baseball CYY	189	9	-	637	623	620	616	616	614	616	613	609	603	608	599	593	589	590	584	567	551	603
Forensic Glass	214	9	719	723	747	760	764	776	780	781	784	786	788	787	786	785	785	784	784	782	781	773
Horse Colic	218	6	649	626	617	609	606	609	604	603	601	602	600	599	598	599	596	596	595	592	589	605
SPECT	267	22	639	573	550	560	547	537	536	536	529	527	529	523	516	521	512	509	512	509	512	536
Heart (Hungaria)	294	10	710	689	684	672	667	670	664	666	664	664	664	663	664	663	663	663	663	663	663	670
Heart (Cleveland)	297	13	668	617	583	573	570	563	558	554	553	550	549	548	546	546	546	546	546	546	547	564
Alcohol	345	5	-	642	634	634	617	620	611	607	610	601	603	595	589	590	580	581	570	563	560	600
Ionosphere	351	33	720	734	740	739	737	734	733	734	733	732	731	733	730	729	729	729	727	726	725	731
Tourism	365	4	567	525	502	492	480	478	471	470	460	454	448	442	438	435	431	429	426	422	414	462
Voting	402	16	840	857	871	887	888	899	898	905	909	909	912	916	917	922	925	926	927	928	928	903
Education	435	4	671	647	625	612	601	598	587	583	576	565	561	553	549	548	539	535	527	519	515	574
Cardiac	558	5	542	523	528	532	539	543	544	549	550	563	564	567	568	570	569	573	574	571	571	555
Football	672	4	601	597	598	601	607	610	610	613	616	619	618	620	619	621	621	624	623	623	628	614
Breast Cancer	683	9	883	876	870	868	866	865	865	866	863	864	863	864	863	863	861	861	860	858	854	865
Diabetes	786	6	684	667	660	651	646	644	645	643	640	638	639	639	638	637	637	637	637	636	637	645
Authorship	841	69	616	615	616	622	621	621	617	618	613	613	609	607	608	609	607	597	604	606	602	612
Baseball HoF	1320	15	562	558	561	565	566	568	568	570	569	570	570	570	570	570	569	569	568	568	567	567
Yeast	1484	8	689	687	686	688	688	688	688	687	687	688	687	687	688	688	688	688	689	689	689	688
Supreme Court	4052	7	939	955	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	957
Hochschule	191	14	650	618	598	590	585	581	581	579	577	575	574	576	570	571	572	572	569	570	572	583
Beratung	437	16	605	578	572	562	562	560	559	558	561	561	562	561	562	564	564	565	566	567	569	566
Schule	1993	15	603	599	595	597	596	599	598	600	602	603	605	605	607	609	611	612	613	613	614	604
Ø			704	685	680	680	678	679	678	678	677	676	676	675	675	674	673	672	671	670	668	677

Tabelle D.22: Prediction-Ergebnisse für FaF ZigSens Plus (33 Datensätze): Die Prozentsätze geben den Anteil der Trainingsstichprobe an der Gesamtstichprobe an. Die Zahlen entsprechen der mit 1000 multiplizierten gerundeten durchschnittlichen Klassifikationsgenauigkeit aus 1000 Trials.

FaF ZigSens Plus			Prediction																			
Datensatz	n	Cues	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	Ø
Bankruptcy	50	5	-	597	703	741	770	787	810	821	825	830	832	835	840	834	839	838	837	817	832	799
Solvency	52	8	519	587	707	726	752	775	777	784	789	793	797	801	795	797	803	794	787	800	796	757
Echocardiogram	62	7	626	603	571	605	631	618	656	646	679	663	688	682	696	688	709	703	697	697	713	662
Post-Operative	86	8	676	645	666	682	704	712	711	718	721	721	720	721	722	724	720	722	717	728	722	708
Broadway	89	5	697	678	716	752	764	774	789	787	790	794	799	800	804	802	804	803	803	809	818	778
Credit Score	100	6	640	654	697	711	728	726	728	732	736	740	735	736	734	742	740	731	741	745	747	723
Zoo	101	16	670	712	725	731	729	727	719	717	711	703	712	692	705	680	686	673	678	674	665	700
Iris	150	4	518	525	546	541	553	553	555	555	565	572	564	586	593	605	598	608	616	640	646	576
Hepatitis	153	7	587	581	591	599	603	592	577	577	577	565	556	547	544	538	532	527	528	530	537	563
Strikes	163	4	655	694	707	710	712	712	715	716	719	717	718	718	718	718	717	717	718	717	718	711
Wine	178	13	620	708	731	759	765	769	773	772	771	773	772	773	775	773	772	774	773	767	764	757
Baseball CYY	189	9	-	574	586	594	589	596	597	600	600	595	595	597	587	582	582	577	576	552	532	584
Forensic Glass	214	9	639	694	733	748	757	766	776	778	780	782	784	785	784	786	785	786	781	781	776	763
Horse Colic	218	6	534	554	563	572	577	580	579	582	584	584	585	588	584	589	588	588	592	586	583	579
SPECT	267	22	474	472	469	486	486	482	484	487	482	484	490	483	482	486	479	486	493	487	498	484
Heart (Hungaria)	294	10	641	655	661	660	661	663	660	661	663	661	662	663	662	662	662	664	662	663	661	660
Heart (Cleveland)	297	13	573	554	528	528	535	530	537	529	536	536	535	539	540	544	543	543	545	544	547	540
Alcohol	345	5	-	596	595	602	588	600	591	587	594	584	588	582	573	575	567	571	562	558	560	582
Ionosphere	351	33	667	717	725	725	726	726	726	724	724	725	724	721	723	723	723	721	720	720	720	720
Tourism	365	4	446	442	439	439	429	430	426	429	420	415	411	404	404	398	395	391	384	380	380	414
Voting	402	16	821	846	864	880	884	895	893	900	905	905	908	914	915	918	923	926	926	929	925	899
Education	435	4	619	605	587	578	566	567	559	558	550	540	531	528	524	523	520	516	511	511	514	548
Cardiac	558	5	439	453	475	490	506	510	516	524	528	542	547	552	552	556	555	563	561	557	554	525
Football	672	4	534	551	566	579	589	592	596	601	606	611	611	613	613	613	615	618	621	621	619	598
Breast Cancer	683	9	870	867	864	864	863	863	860	861	859	860	860	862	859	858	858	857	858	856	852	861
Diabetes	786	6	635	638	637	636	635	635	635	634	634	635	635	634	634	633	634	632	634	638	631	635
Authorship	841	69	571	591	602	611	614	615	612	615	608	608	606	604	604	607	603	591	596	595	593	602
Baseball HoF	1320	15	556	553	557	562	563	566	565	568	568	566	567	567	566	567	568	566	565	566	568	564
Yeast	1484	8	685	686	686	686	686	686	686	686	687	686	687	687	687	687	688	687	687	688	689	687
Supreme Court	4052	7	938	955	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	957
Hochschule	191	14	470	482	487	494	503	505	516	519	524	533	535	541	544	547	545	551	552	552	557	524
Beratung	437	16	480	497	509	515	524	527	535	536	539	541	543	545	549	549	551	551	551	553	555	534
Schule	1993	15	575	584	586	590	592	594	593	597	599	601	604	604	606	608	611	610	613	616	614	600
Ø			613	623	637	647	653	656	658	659	662	661	662	662	663	663	663	662	662	662	662	654

Tabelle D.23: Fitting-Ergebnisse für das FaF ZigSens Majority (33 Datensätze): Die Prozentsätze geben den Anteil der Trainingsstichprobe an der Gesamtstichprobe an. Die Zahlen entsprechen der mit 1000 multiplizierten gerundeten durchschnittlichen Klassifikationsgenauigkeit aus 1000 Trials, wobei 1,0 durch „000“ repräsentiert wird.

FaF ZigSens Majority			Fitting																			
Datensatz	n	Cues	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	Ø
Bankruptcy	50	5	-	809	839	853	863	846	873	869	872	864	869	864	863	859	863	858	861	851	849	857
Solvency	52	8	742	818	849	852	834	859	860	852	859	853	852	849	851	846	846	847	846	843	839	842
Echocardiogram	62	7	798	854	827	787	766	759	751	745	743	741	739	737	739	737	735	732	734	732	732	757
Post-Operative	86	8	770	737	731	721	719	720	725	719	719	719	722	721	720	719	721	721	722	720	721	725
Broadway	89	5	831	819	819	817	814	813	807	810	807	808	808	810	806	809	809	810	809	808	809	812
Credit Score	100	6	831	796	779	782	774	771	772	767	767	766	765	761	761	761	760	761	759	760	760	771
Zoo	101	16	852	880	895	922	933	951	957	979	979	989	993	995	999	999	000	000	000	000	000	964
Iris	150	4	727	716	704	702	692	697	695	699	707	712	711	712	716	732	734	739	745	773	772	720
Hepatitis	153	7	808	781	763	756	748	742	738	738	735	731	728	725	723	721	719	716	714	713	712	737
Strikes	163	4	794	766	750	744	737	740	737	736	733	733	733	730	731	730	730	729	726	723	720	738
Wine	178	13	834	869	855	894	896	897	903	901	907	904	909	908	912	912	917	918	918	922	924	900
Baseball CYY	189	9	-	779	743	723	722	714	709	705	708	703	703	711	710	705	711	710	715	703	691	715
Forensic Glass	214	9	833	805	789	789	790	791	790	789	789	789	788	788	788	785	785	784	784	782	782	790
Horse Colic	218	6	711	691	671	671	668	660	671	665	671	672	677	682	693	687	702	710	710	730	744	689
SPECT	267	22	787	758	747	744	742	738	735	737	734	735	736	734	734	736	734	736	740	739	744	742
Heart (Hungaria)	294	10	738	703	698	686	677	675	670	669	666	666	665	663	664	664	663	663	663	663	663	675
Heart (Cleveland)	297	13	656	615	584	575	575	564	560	556	556	551	550	549	546	546	546	546	546	546	547	564
Alcohol	345	5	-	761	757	760	754	753	751	750	748	746	743	741	739	736	732	730	725	720	710	742
Ionosphere	351	33	770	747	741	739	737	734	733	734	733	732	732	733	730	729	729	729	727	726	725	735
Tourism	365	4	715	686	670	664	652	648	642	640	636	633	630	627	627	623	619	619	617	615	612	641
Voting	402	16	928	951	958	960	961	962	963	963	963	963	963	963	963	964	964	964	964	965	965	960
Education	435	4	663	639	625	606	599	600	589	585	579	568	564	556	552	550	540	537	529	519	515	574
Cardiac	558	5	747	728	721	717	715	714	715	716	713	714	714	715	715	715	715	714	713	713	711	717
Football	672	4	666	649	641	638	638	638	636	635	635	633	633	633	631	632	631	630	629	629	629	636
Breast Cancer	683	9	873	851	840	836	833	826	829	824	824	815	814	812	807	805	804	800	798	796	796	820
Diabetes	786	6	755	734	728	720	719	718	716	717	717	716	716	716	716	716	717	717	717	717	717	721
Authorship	841	69	827	808	803	806	804	804	811	816	821	823	832	837	844	851	857	858	862	864	864	831
Baseball HoF	1320	15	737	728	729	731	732	734	733	736	736	737	738	739	739	739	739	740	740	740	740	736
Yeast	1484	8	689	687	686	688	688	688	688	687	687	688	687	687	688	688	688	688	689	689	689	688
Supreme Court	4052	7	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958
Hochschule	191	14	812	758	728	708	694	686	682	677	673	667	665	664	659	659	655	654	652	652	652	684
Beratung	437	16	751	714	696	686	680	677	675	672	672	671	670	669	668	668	667	665	664	660	656	678
Schule	1993	15	625	616	613	613	611	612	611	611	612	613	612	612	613	613	613	614	613	613	614	613
Ø			774	764	756	753	749	748	748	747	747	746	746	745	746	745	746	745	745	745	744	749

Tabelle D.24: Prediction-Ergebnisse für FaF ZigSens Majority (33 Datensätze): Die Prozentsätze geben den Anteil der Trainingsstichprobe an der Gesamtstichprobe an. Die Zahlen entsprechen der mit 1000 multiplizierten gerundeten durchschnittlichen Klassifikationsgenauigkeit aus 1000 Trials.

FaF ZigSens Majority			Prediction																			
Datensatz	n	Cues	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	Ø
Bankruptcy	50	5	-	589	647	709	742	749	814	826	828	830	829	832	834	826	825	827	817	800	812	785
Solvency	52	8	516	584	682	708	704	761	772	766	771	784	779	778	777	777	777	775	770	775	759	738
Echocardiogram	62	7	625	676	666	700	710	707	721	722	727	727	732	732	729	730	730	733	729	716	720	712
Post-Operative	86	8	677	681	702	711	720	720	719	722	722	723	720	721	722	724	720	722	717	728	722	715
Broadway	89	5	712	724	757	785	791	794	799	800	802	803	804	801	808	804	804	803	803	809	818	791
Credit Score	100	6	687	739	765	768	772	768	766	766	765	762	761	762	763	761	762	755	764	760	754	758
Zoo	101	16	701	804	847	890	904	932	938	966	968	980	985	989	994	995	996	996	996	996	996	941
Iris	150	4	560	587	603	619	621	631	637	644	662	669	668	672	678	699	693	701	707	746	731	659
Hepatitis	153	7	707	722	719	714	713	711	709	708	707	707	706	705	706	706	705	709	707	706	712	709
Strikes	163	4	655	692	703	708	712	713	711	712	715	712	713	711	711	710	710	711	716	715	718	708
Wine	178	13	623	748	772	849	861	868	882	885	893	894	901	901	906	906	911	914	916	922	919	867
Baseball CYY	189	9	-	695	683	679	679	679	676	678	680	679	679	686	680	676	681	676	672	656	633	676
Forensic Glass	214	9	748	775	783	786	787	787	787	788	786	786	785	786	786	786	785	785	781	781	776	782
Horse Colic	218	6	549	571	578	596	604	600	617	618	632	632	642	645	665	656	676	687	685	714	733	637
SPECT	267	22	673	698	705	712	713	715	718	719	719	719	721	722	721	722	724	726	726	731	737	717
Heart (Hungaria)	294	10	661	670	681	675	673	669	666	665	665	664	663	663	662	663	663	664	662	663	661	666
Heart (Cleveland)	297	13	536	544	524	527	539	530	537	531	538	537	535	539	540	543	543	543	545	544	547	538
Alcohol	345	5	-	722	728	736	735	741	739	738	738	736	734	733	729	727	726	725	722	722	716	730
Ionosphere	351	33	693	721	725	725	725	726	726	724	725	725	724	721	723	723	723	721	720	720	720	722
Tourism	365	4	604	600	601	600	599	597	596	596	593	592	592	589	588	583	583	578	575	569	574	590
Voting	402	16	885	929	943	949	953	956	958	959	960	960	961	961	962	961	962	963	963	964	963	953
Education	435	4	584	580	575	564	556	564	556	557	549	540	531	528	525	524	520	516	510	511	514	542
Cardiac	558	5	660	674	684	691	696	697	699	702	703	704	706	709	708	711	710	713	711	707	710	700
Football	672	4	608	616	622	625	626	626	626	627	627	628	628	627	628	626	627	626	630	631	623	625
Breast Cancer	683	9	853	841	834	832	827	824	825	820	819	814	814	809	805	804	803	800	800	799	797	817
Diabetes	786	6	689	697	701	702	704	705	706	706	705	708	707	708	710	710	712	712	714	718	716	707
Authorship	841	69	764	774	783	793	795	796	804	811	815	819	827	833	841	848	856	856	860	860	861	821
Baseball HoF	1320	15	724	722	726	729	731	733	733	735	736	736	737	738	738	739	740	738	738	738	741	734
Yeast	1484	8	685	686	686	686	686	686	686	686	686	686	687	687	687	687	688	687	687	688	689	687
Supreme Court	4052	7	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958
Hochschule	191	14	600	622	623	624	627	630	634	633	635	638	642	639	645	642	646	646	649	652	652	636
Beratung	437	16	623	631	636	638	641	643	648	648	650	651	653	654	655	655	657	658	651	654	650	647
Schule	1993	15	599	604	606	608	609	608	609	610	610	610	612	611	611	612	614	612	614	616	614	610
Ø			672	693	704	715	719	722	727	728	730	731	731	732	733	733	734	734	734	735	735	724

Tabelle D.25: Fitting-Ergebnisse für das FaF ZigSens Fold (33 Datensätze): Die Prozentsätze geben den Anteil der Trainingsstichprobe an der Gesamtstichprobe an. Die Zahlen entsprechen der mit 1000 multiplizierten gerundeten durchschnittlichen Klassifikationsgenauigkeit aus 1000 Trials, wobei 1,0 durch „000“ repräsentiert wird.

FaF ZigSens Fold			Fitting																			
Datensatz	n	Cues	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	Ø
Bankruptcy	50	5	-	804	845	851	855	848	867	864	866	863	863	857	857	858	854	850	849	851	828	852
Solvency	52	8	742	812	861	860	835	861	863	853	858	851	853	849	849	845	845	842	841	845	841	843
Echocardiogram	62	7	798	868	825	796	777	768	760	755	756	752	753	747	755	750	755	759	757	768	761	772
Post-Operative	86	8	818	737	729	720	720	720	724	719	719	719	722	721	720	719	721	721	722	720	721	727
Broadway	89	5	856	816	821	818	814	814	806	811	808	807	807	807	804	805	805	802	801	798	790	810
Credit Score	100	6	814	783	762	749	750	744	742	740	738	735	736	732	729	732	730	731	729	730	730	744
Zoo	101	16	832	873	874	896	931	928	948	964	979	982	986	994	997	998	999	000	000	000	000	957
Iris	150	4	743	728	717	708	696	707	702	696	705	712	708	704	706	723	719	722	720	746	752	717
Hepatitis	153	7	819	793	781	782	780	778	777	778	781	779	780	779	778	778	779	777	778	777	777	782
Strikes	163	4	790	763	751	744	738	741	737	735	733	732	733	731	730	729	729	729	726	724	720	738
Wine	178	13	834	874	862	898	900	900	904	903	908	906	910	909	912	912	917	918	918	922	924	902
Baseball CYY	189	9	-	811	776	756	754	744	737	731	732	726	723	730	727	720	722	723	725	714	702	736
Forensic Glass	214	9	831	803	783	785	784	785	784	784	785	787	788	787	787	789	790	793	795	797	801	791
Horse Colic	218	6	708	699	672	676	673	662	672	666	672	673	677	682	693	687	701	710	711	729	745	690
SPECT	267	22	798	785	780	781	784	782	778	780	779	782	781	778	780	779	777	777	776	774	774	780
Heart (Hungaria)	294	10	726	696	691	683	674	676	670	668	667	666	665	664	664	664	664	663	663	663	663	673
Heart (Cleveland)	297	13	656	629	585	576	581	564	565	557	559	554	550	551	546	548	546	546	546	546	547	566
Alcohol	345	5	-	797	791	791	790	788	786	786	786	787	786	785	786	786	786	786	786	788	788	788
Ionosphere	351	33	751	739	735	735	733	730	729	731	729	730	730	731	729	728	728	728	727	726	725	731
Tourism	365	4	759	741	731	725	715	713	707	704	705	699	698	695	694	691	688	689	688	688	688	706
Voting	402	16	930	953	955	957	959	960	962	962	962	962	963	963	964	964	964	964	964	965	965	960
Education	435	4	687	658	622	605	599	612	604	600	593	582	564	556	552	557	552	547	534	526	515	582
Cardiac	558	5	826	815	814	810	811	811	814	814	813	814	814	814	815	817	815	816	818	820	822	815
Football	672	4	677	659	653	649	648	647	646	645	645	643	644	643	642	643	643	641	641	642	644	647
Breast Cancer	683	9	865	835	828	822	821	818	820	817	818	811	810	809	805	804	802	799	797	796	796	815
Diabetes	786	6	762	737	731	723	721	720	718	718	718	717	716	716	716	716	716	717	717	717	717	722
Authorship	841	69	844	830	822	814	810	805	802	801	796	796	795	795	794	795	794	793	794	794	793	804
Baseball HoF	1320	15	849	838	840	837	839	841	839	842	841	842	842	843	844	844	844	847	845	844	841	842
Yeast	1484	8	689	687	686	688	688	688	688	688	688	688	688	688	688	688	688	688	688	688	688	688
Supreme Court	4052	7	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958
Hochschule	191	14	829	776	753	738	732	720	716	713	711	707	705	707	703	706	704	705	703	704	709	723
Beratung	437	16	771	747	729	718	715	712	712	709	708	709	707	708	706	706	705	705	705	705	707	715
Schule	1993	15	625	617	613	613	611	613	611	612	612	613	612	612	613	613	613	614	613	613	614	613
Ø			786	778	769	766	764	762	762	761	761	760	760	759	759	759	759	759	759	760	759	763

Tabelle D.26: Prediction-Ergebnisse für FaF ZigSens Fold (33 Datensätze): Die Prozentsätze geben den Anteil der Trainingsstichprobe an der Gesamtstichprobe an. Die Zahlen entsprechen der mit 1000 multiplizierten gerundeten durchschnittlichen Klassifikationsgenauigkeit aus 1000 Trials.

FaF ZigSens Fold			Prediction																			
Datensatz	n	Cues	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	Ø
Bankruptcy	50	5	-	578	647	701	729	748	808	811	822	830	823	826	829	826	821	822	812	801	807	780
Solvency	52	8	511	583	666	697	707	756	768	766	772	779	779	780	770	776	777	775	767	774	761	735
Echocardiogram	62	7	629	643	676	692	706	713	722	726	731	726	734	734	733	729	739	747	732	744	731	715
Post-Operative	86	8	624	684	702	712	714	720	719	722	722	723	720	721	722	724	720	722	717	728	722	712
Broadway	89	5	640	720	754	771	781	788	794	793	795	796	794	795	799	793	793	792	791	792	792	777
Credit Score	100	6	689	745	747	740	745	738	734	736	733	731	731	732	734	729	732	726	733	729	725	732
Zoo	101	16	667	776	817	849	901	900	926	944	969	970	976	986	992	993	995	996	996	997	997	929
Iris	150	4	559	593	607	621	620	637	637	637	652	661	660	657	661	679	667	672	667	700	684	646
Hepatitis	153	7	756	762	770	769	774	776	778	777	777	776	776	776	778	776	775	778	773	771	772	773
Strikes	163	4	657	697	705	707	712	713	712	711	715	715	715	714	714	713	713	713	716	715	717	709
Wine	178	13	621	742	769	842	859	866	880	885	894	895	901	901	905	907	912	914	916	922	919	866
Baseball CYY	189	9	-	724	712	710	708	706	702	702	704	700	697	707	700	691	694	691	687	669	647	697
Forensic Glass	214	9	735	763	771	777	774	776	776	778	779	778	782	781	781	786	786	791	791	795	794	779
Horse Colic	218	6	549	580	578	602	610	601	619	618	633	633	642	645	664	655	676	687	686	714	735	638
SPECT	267	22	723	757	767	769	773	775	775	777	777	778	775	778	778	775	778	776	771	774	771	771
Heart (Hungaria)	294	10	649	664	671	672	670	668	666	664	665	663	663	664	662	663	663	664	662	663	661	664
Heart (Cleveland)	297	13	536	542	524	529	536	529	536	531	535	533	535	538	540	541	542	543	545	544	547	537
Alcohol	345	5	-	760	768	776	777	779	781	782	780	782	784	783	783	783	784	786	787	786	791	781
Ionosphere	351	33	694	709	715	718	719	720	721	719	720	721	722	719	721	721	722	720	720	720	720	718
Tourism	365	4	663	674	678	677	676	677	675	674	673	672	672	669	668	666	668	666	663	662	675	671
Voting	402	16	888	929	940	945	950	954	956	958	959	960	960	961	962	962	962	963	963	964	962	953
Education	435	4	607	598	572	563	556	574	569	570	561	551	531	528	525	529	528	521	512	512	514	548
Cardiac	558	5	771	786	793	796	799	802	805	804	806	808	810	811	810	813	810	815	816	815	824	805
Football	672	4	622	628	633	635	635	635	636	637	636	637	635	635	636	635	635	634	635	638	636	634
Breast Cancer	683	9	843	824	820	816	814	815	815	812	812	809	809	805	803	803	801	799	799	799	797	810
Diabetes	786	6	691	699	703	703	704	705	706	706	704	707	706	707	709	709	711	712	714	718	716	707
Authorship	841	69	804	810	809	807	804	800	799	798	794	793	792	793	794	793	793	792	793	792	794	798
Baseball HoF	1320	15	830	827	831	830	835	836	836	839	839	839	838	841	842	843	844	845	842	842	841	838
Yeast	1484	8	687	687	688	687	688	688	688	688	688	688	688	689	687	688	688	687	687	687	688	688
Supreme Court	4052	7	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958
Hochschule	191	14	644	659	670	669	673	677	680	680	684	688	692	691	693	691	698	699	699	701	704	684
Beratung	437	16	659	675	678	681	684	688	691	691	693	694	697	696	697	698	698	700	697	702	704	691
Schule	1993	15	600	604	606	608	609	609	609	610	610	610	612	612	611	612	614	612	614	616	614	610
Ø			684	709	720	728	733	737	742	742	745	746	746	746	747	747	748	749	747	750	749	738

Tabelle D.27: Fitting-Ergebnisse für das FaF ZigVal Plus (33 Datensätze): Die Prozentsätze geben den Anteil der Trainingsstichprobe an der Gesamtstichprobe an. Die Zahlen entsprechen der mit 1000 multiplizierten gerundeten durchschnittlichen Klassifikationsgenauigkeit aus 1000 Trials.

FaF ZigVal Plus			Fitting																			
Datensatz	n	Cues	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	Ø
Bankruptcy	50	5	-	838	879	876	876	873	874	870	871	867	870	864	863	864	863	858	860	861	850	866
Solvency	52	8	742	830	851	857	843	858	863	851	856	850	852	846	848	843	844	843	843	842	838	842
Echocardiogram	62	7	804	904	866	834	813	801	803	795	796	791	792	789	794	792	795	796	792	794	789	807
Post-Operative	86	8	824	787	762	742	731	723	719	712	704	704	703	701	699	697	697	697	697	696	691	720
Broadway	89	5	858	839	820	800	796	795	786	781	781	776	781	780	774	776	777	778	771	768	761	789
Credit Score	100	6	896	871	871	872	867	865	864	861	855	856	849	849	840	841	835	835	831	831	830	854
Zoo	101	16	991	992	983	981	975	974	971	969	967	967	964	962	964	960	960	957	960	958	961	969
Iris	150	4	752	745	735	732	722	737	738	729	738	744	739	727	730	741	730	730	724	731	715	734
Hepatitis	153	7	837	804	771	755	733	726	717	718	710	704	698	696	695	692	689	686	685	685	686	720
Strikes	163	4	800	770	758	756	752	755	752	753	750	751	753	751	751	750	752	752	751	749	748	756
Wine	178	13	890	901	889	901	900	902	904	903	908	906	910	909	914	913	918	918	919	922	924	908
Baseball CYY	189	9	-	783	749	727	728	715	700	692	687	684	673	674	671	666	662	662	660	658	656	691
Forensic Glass	214	9	874	848	835	829	830	828	826	825	826	822	821	820	819	818	817	816	815	813	813	826
Horse Colic	218	6	777	750	738	740	742	744	746	745	747	749	748	748	748	746	746	744	742	740	735	746
SPECT	267	22	793	727	703	696	692	681	681	678	674	671	667	667	666	664	662	660	658	659	659	682
Heart (Hungaria)	294	10	883	857	848	842	837	835	832	830	829	829	828	827	827	827	826	827	827	827	827	835
Heart (Cleveland)	297	13	879	848	829	825	823	816	813	808	810	807	808	806	804	804	802	802	800	798	795	815
Alcohol	345	5	-	760	753	749	744	743	738	739	739	736	737	739	739	742	743	744	747	747	750	744
Ionosphere	351	33	869	815	786	786	778	768	769	767	763	765	759	756	755	753	750	746	740	735	728	768
Tourism	365	4	738	703	681	676	666	664	660	657	656	654	651	649	650	648	647	646	646	643	637	662
Voting	402	16	942	938	939	941	945	947	949	948	951	951	952	952	953	953	953	953	954	954	954	949
Education	435	4	769	766	764	759	758	756	757	754	756	754	755	752	752	752	750	750	749	746	745	755
Cardiac	558	5	723	690	674	666	658	653	647	643	638	633	629	625	620	616	611	609	601	599	598	639
Football	672	4	682	663	652	650	648	648	646	647	645	644	645	645	644	644	644	643	642	642	642	648
Breast Cancer	683	9	972	967	963	962	960	960	960	959	959	959	958	959	958	959	958	958	958	958	958	960
Diabetes	786	6	766	742	737	731	727	725	721	721	721	719	718	717	716	715	715	714	714	714	714	723
Authorship	841	69	848	825	818	813	806	803	800	800	799	797	798	800	799	800	800	799	800	800	797	805
Baseball HoF	1320	15	688	665	658	653	649	649	647	648	647	648	649	649	650	650	651	652	652	653	655	653
Yeast	1484	8	691	673	666	663	660	659	657	654	653	652	649	648	646	645	642	641	637	635	628	653
Supreme Court	4052	7	820	815	823	820	811	808	811	804	802	796	800	799	798	791	788	789	798	774	730	799
Hochschule	191	14	841	788	755	729	715	702	692	687	677	672	667	662	658	657	653	652	652	651	649	693
Beratung	437	16	777	730	709	694	684	680	678	673	673	675	674	674	674	675	676	677	678	680	683	688
Schule	1993	15	674	654	647	642	639	638	636	635	634	634	633	632	633	632	631	631	631	631	632	638
Ø			813	797	785	779	773	771	768	765	764	763	762	760	759	758	757	757	756	754	751	768

Tabelle D.28: Prediction-Ergebnisse für FaF ZigVal Plus (33 Datensätze): Die Prozentsätze geben den Anteil der Trainingsstichprobe an der Gesamtstichprobe an. Die Zahlen entsprechen der mit 1000 multiplizierten gerundeten durchschnittlichen Klassifikationsgenauigkeit aus 1000 Trials.

FaF ZigVal Plus			Prediction																			
Datensatz	n	Cues	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	Ø
Bankruptcy	50	5	-	640	733	771	799	819	826	831	832	839	830	833	835	838	826	827	819	822	815	807
Solvency	52	8	512	606	728	747	762	777	775	777	776	783	780	778	771	771	778	777	764	783	778	749
Echocardiogram	62	7	627	649	638	657	674	671	694	690	718	708	727	726	737	733	747	749	738	746	747	704
Post-Operative	86	8	645	603	601	609	622	628	635	631	629	635	635	635	639	642	638	643	645	654	642	632
Broadway	89	5	695	685	700	716	718	730	733	731	735	737	742	739	738	738	743	745	742	733	723	728
Credit Score	100	6	646	713	768	780	806	802	814	809	818	813	813	813	816	811	811	804	811	806	800	792
Zoo	101	16	851	913	928	936	941	942	940	946	944	946	945	943	949	945	943	947	947	947	944	937
Iris	150	4	590	635	652	667	665	687	694	689	704	713	710	695	700	712	692	696	685	697	674	682
Hepatitis	153	7	661	683	688	694	692	686	688	685	687	685	684	680	682	680	679	680	677	682	682	683
Strikes	163	4	666	703	718	728	731	733	731	732	734	734	735	736	735	735	736	735	735	734	732	728
Wine	178	13	695	813	840	865	870	879	885	888	895	895	900	900	906	907	911	913	915	921	919	880
Baseball CYY	189	9	-	700	692	686	686	682	671	670	666	665	662	663	658	654	657	653	657	650	650	668
Forensic Glass	214	9	706	745	770	782	791	793	797	800	801	803	805	807	806	805	806	805	805	805	795	791
Horse Colic	218	6	581	638	668	695	708	716	724	729	734	734	739	738	738	741	739	740	741	737	734	714
SPECT	267	22	600	607	614	625	632	631	635	637	634	639	639	639	639	641	639	640	639	635	639	632
Heart (Hungaria)	294	10	752	788	799	805	807	809	810	812	813	812	813	815	816	815	818	817	817	817	819	808
Heart (Cleveland)	297	13	708	738	746	752	760	760	764	764	765	765	766	766	769	768	767	763	766	761	760	758
Alcohol	345	5	-	707	709	715	711	720	718	719	724	720	724	724	724	727	731	735	739	738	745	724
Ionosphere	351	33	696	712	712	720	722	722	726	724	722	729	727	724	727	725	726	723	717	717	712	720
Tourism	365	4	603	607	611	612	614	614	615	617	616	617	618	618	616	615	618	613	612	608	608	613
Voting	402	16	889	910	920	928	935	941	941	944	947	947	948	950	950	951	952	953	953	953	951	940
Education	435	4	720	738	744	745	745	746	745	744	744	745	747	746	747	746	747	747	746	747	742	744
Cardiac	558	5	605	608	610	613	613	609	609	608	607	607	606	602	598	597	594	594	587	587	585	602
Football	672	4	595	610	619	625	629	630	632	633	635	636	637	637	639	638	640	639	643	643	635	631
Breast Cancer	683	9	935	945	949	951	952	954	954	954	955	956	956	956	956	956	957	958	957	956	958	953
Diabetes	786	6	697	709	713	714	715	715	716	715	713	714	714	713	714	713	713	714	713	718	715	713
Authorship	841	69	786	795	797	796	795	794	791	794	792	790	795	795	794	796	796	795	793	792	789	794
Baseball HoF	1320	15	676	655	650	646	646	646	642	644	645	644	647	646	647	648	651	648	650	651	655	649
Yeast	1484	8	624	637	641	643	644	645	644	643	643	642	641	640	638	640	636	633	633	627	620	638
Supreme Court	4052	7	807	808	818	816	808	805	810	802	800	794	798	797	797	789	786	789	798	772	729	796
Hochschule	191	14	571	592	594	600	603	609	608	609	614	618	617	621	625	624	622	625	625	621	626	612
Beratung	437	16	609	630	637	641	645	649	651	652	654	656	657	658	661	662	665	667	667	671	676	653
Schule	1993	15	603	612	618	620	621	622	624	624	625	626	625	627	627	627	627	628	628	627	627	623
Ø			678	701	716	724	729	732	735	735	737	738	739	738	739	739	739	739	738	738	734	730

Tabelle D.29: Fitting-Ergebnisse für das FaF ZigVal Majority (33 Datensätze): Die Prozentsätze geben den Anteil der Trainingsstichprobe an der Gesamtstichprobe an. Die Zahlen entsprechen der mit 1000 multiplizierten gerundeten durchschnittlichen Klassifikationsgenauigkeit aus 1000 Trials.

FaF ZigVal Majority			Fitting																			
Datensatz	n	Cues	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	Ø
Bankruptcy	50	5	-	832	879	877	878	866	874	871	872	862	870	865	864	859	864	858	861	851	850	864
Solvency	52	8	742	833	840	845	842	857	859	853	857	854	852	850	853	848	846	847	846	844	840	843
Echocardiogram	62	7	798	899	864	831	816	800	805	794	795	791	792	788	794	792	795	796	792	794	789	807
Post-Operative	86	8	810	778	756	741	728	723	720	712	705	704	703	701	699	698	696	697	697	697	691	719
Broadway	89	5	844	833	820	806	799	800	790	788	785	781	787	784	778	778	779	779	772	769	762	791
Credit Score	100	6	889	877	872	872	860	860	857	857	848	851	844	844	837	839	833	834	830	831	830	851
Zoo	101	16	990	990	980	979	972	974	971	970	964	968	965	963	962	960	962	959	961	960	963	969
Iris	150	4	745	741	735	732	719	737	738	728	738	744	740	727	730	741	730	731	724	731	715	733
Hepatitis	153	7	839	806	773	754	733	726	717	718	710	704	698	696	695	692	688	686	685	685	686	721
Strikes	163	4	799	772	762	759	753	759	755	755	754	754	757	754	755	756	757	757	757	755	752	759
Wine	178	13	883	901	888	900	900	902	905	902	908	906	911	909	914	913	918	918	919	922	924	908
Baseball CYY	189	9	-	779	747	726	728	715	699	692	687	684	673	674	670	666	662	662	660	658	656	691
Forensic Glass	214	9	879	853	841	836	833	830	829	827	826	823	822	820	820	819	817	816	815	812	813	828
Horse Colic	218	6	774	739	731	731	736	740	743	744	746	748	747	748	748	746	746	744	743	740	735	744
SPECT	267	22	796	729	703	695	692	680	681	678	673	671	667	667	665	664	662	660	659	659	659	682
Heart (Hungaria)	294	10	877	848	841	836	834	832	830	829	828	828	827	827	827	827	826	826	827	827	827	833
Heart (Cleveland)	297	13	866	835	822	815	811	807	803	800	803	800	800	799	798	799	798	798	798	796	795	807
Alcohol	345	5	-	760	752	749	744	743	738	739	739	736	737	739	739	741	743	744	747	748	750	744
Ionosphere	351	33	871	826	802	796	784	775	776	774	770	766	763	760	760	754	754	749	745	742	735	774
Tourism	365	4	737	703	681	676	666	664	660	657	656	654	651	649	650	649	647	646	646	643	637	662
Voting	402	16	942	938	940	944	945	947	949	949	951	951	951	952	953	953	953	953	954	954	954	949
Education	435	4	772	762	751	744	743	742	742	741	742	741	738	739	740	742	742	743	744	745	745	745
Cardiac	558	5	723	689	675	666	658	652	647	643	638	633	629	625	620	616	612	609	602	599	598	639
Football	672	4	682	663	652	650	648	648	646	647	645	644	645	645	644	644	644	643	642	642	642	648
Breast Cancer	683	9	967	965	962	962	960	960	960	959	959	959	958	959	958	959	958	958	958	958	958	960
Diabetes	786	6	765	743	737	731	727	725	721	721	721	718	718	717	716	715	715	714	714	714	714	723
Authorship	841	69	848	825	818	813	807	803	799	800	800	797	798	799	799	801	801	799	801	800	798	806
Baseball HoF	1320	15	685	665	656	653	649	649	648	648	646	649	649	649	649	650	651	652	652	654	655	653
Yeast	1484	8	694	676	670	666	662	659	657	655	653	652	649	648	646	645	642	641	637	635	628	653
Supreme Court	4052	7	820	815	823	820	811	808	811	804	802	796	800	799	798	791	788	789	798	775	730	799
Hochschule	191	14	836	788	755	729	715	702	693	686	678	672	667	662	658	657	653	652	653	652	649	692
Beratung	437	16	779	730	709	694	684	681	678	673	673	675	674	674	674	675	676	677	678	680	683	688
Schule	1993	15	674	654	647	642	639	638	636	635	634	634	633	632	633	632	631	631	631	631	632	638
Ø			811	795	784	778	772	770	768	765	764	762	761	760	759	758	757	757	756	755	751	767

Tabelle D.30: Prediction-Ergebnisse für FaF ZigVal Majority (33 Datensätze): Die Prozentsätze geben den Anteil der Trainingsstichprobe an der Gesamtstichprobe an. Die Zahlen entsprechen der mit 1000 multiplizierten gerundeten durchschnittlichen Klassifikationsgenauigkeit aus 1000 Trials.

FaF ZigVal Majority			Prediction																			
Datensatz	n	Cues	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	Ø
Bankruptcy	50	5	-	619	732	768	795	808	824	830	831	833	831	833	835	828	826	827	819	802	812	803
Solvency	52	8	514	596	723	747	760	780	777	781	778	784	781	780	775	778	778	773	767	777	764	748
Echocardiogram	62	7	627	664	649	669	683	679	701	697	721	711	730	726	737	734	749	747	739	745	745	708
Post-Operative	86	8	657	616	617	623	632	634	637	633	630	635	635	634	639	642	638	643	645	655	642	636
Broadway	89	5	699	701	717	732	734	744	749	745	746	748	755	749	748	744	750	747	744	734	725	737
Credit Score	100	6	667	727	778	791	810	806	816	811	817	814	812	813	815	811	812	805	812	806	802	796
Zoo	101	16	834	905	920	933	935	940	942	945	942	947	949	945	945	944	945	947	949	946	946	935
Iris	150	4	583	622	649	668	662	688	694	689	704	713	710	696	700	712	692	696	684	697	674	681
Hepatitis	153	7	663	687	691	694	691	687	688	684	687	685	683	680	681	680	679	680	678	682	682	683
Strikes	163	4	660	699	716	726	729	734	732	733	737	737	737	738	739	738	741	743	746	742	741	730
Wine	178	13	675	810	837	865	871	878	885	887	895	895	901	901	907	907	911	912	915	920	919	878
Baseball CYY	189	9	-	697	692	685	687	682	671	670	666	665	661	662	658	655	657	653	657	650	650	668
Forensic Glass	214	9	720	757	779	791	798	799	803	804	804	805	807	807	806	806	807	805	805	805	795	795
Horse Colic	218	6	574	622	656	681	698	709	718	728	732	733	738	738	738	741	739	740	741	737	734	710
SPECT	267	22	602	610	612	625	632	632	635	637	634	639	639	639	640	641	638	640	639	635	640	632
Heart (Hungaria)	294	10	756	790	802	807	810	811	812	814	815	814	815	816	817	816	818	817	817	817	819	810
Heart (Cleveland)	297	13	704	736	745	750	757	758	762	765	766	768	770	772	776	776	776	775	779	774	771	762
Alcohol	345	5	-	707	709	715	711	720	719	719	725	719	725	724	724	726	731	735	738	738	745	724
Ionosphere	351	33	691	708	710	718	717	719	720	722	721	721	721	717	722	718	719	714	712	710	702	715
Tourism	365	4	604	607	611	612	614	614	615	617	616	617	618	618	616	615	618	613	612	608	608	613
Voting	402	16	889	913	922	931	935	941	941	945	948	947	948	950	950	951	952	953	953	953	951	941
Education	435	4	705	713	713	714	715	720	721	724	721	722	720	725	728	728	734	735	737	745	741	724
Cardiac	558	5	603	608	611	613	612	609	609	608	606	607	606	602	598	597	594	594	587	587	585	602
Football	672	4	595	610	619	624	629	630	632	633	635	636	637	637	639	638	640	639	643	643	635	631
Breast Cancer	683	9	934	946	950	952	953	954	954	954	955	956	956	956	956	956	957	958	957	956	958	953
Diabetes	786	6	699	709	713	714	716	715	716	715	713	714	714	713	714	713	713	714	713	718	715	713
Authorship	841	69	786	794	797	796	796	794	792	794	793	790	794	794	795	797	797	795	794	792	788	794
Baseball HoF	1320	15	671	655	650	646	645	646	644	645	644	645	647	645	646	648	651	649	650	652	655	649
Yeast	1484	8	630	643	646	648	647	647	645	644	643	642	641	640	638	640	636	633	633	627	620	639
Supreme Court	4052	7	807	808	818	816	808	805	810	802	800	794	798	797	797	789	786	789	798	773	729	796
Hochschule	191	14	570	592	595	601	603	608	608	609	614	618	618	622	625	623	622	625	625	620	626	612
Beratung	437	16	608	629	637	641	645	649	651	653	654	657	657	659	661	662	665	667	667	671	676	653
Schule	1993	15	602	612	618	620	621	622	624	624	625	626	625	627	627	627	627	628	628	627	627	623
Ø			678	700	716	725	729	732	735	735	737	737	739	738	739	739	739	739	739	738	734	730

Tabelle D.31: Fitting-Ergebnisse für das FaF ZigVal Fold (33 Datensätze): Die Prozentsätze geben den Anteil der Trainingsstichprobe an der Gesamtstichprobe an. Die Zahlen entsprechen der mit 1000 multiplizierten gerundeten durchschnittlichen Klassifikationsgenauigkeit aus 1000 Trials.

FaF ZigVal Fold			Fitting																			
Datensatz	n	Cues	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	Ø
Bankruptcy	50	5	-	819	878	870	870	866	870	864	866	863	864	857	857	858	854	850	850	851	828	857
Solvency	52	8	742	825	869	866	847	861	863	855	859	851	853	850	848	844	847	843	842	844	838	845
Echocardiogram	62	7	798	915	878	857	836	823	823	815	816	810	808	805	808	804	805	801	794	796	778	820
Post-Operative	86	8	832	781	761	747	738	732	730	725	722	718	721	719	718	717	718	719	720	719	720	735
Broadway	89	5	875	838	828	816	810	809	800	802	798	797	798	801	795	799	799	802	802	806	806	810
Credit Score	100	6	872	900	863	848	837	827	816	817	801	800	795	789	780	779	772	771	766	764	761	808
Zoo	101	16	994	987	982	976	972	971	971	969	964	966	966	962	965	959	961	959	962	959	961	969
Iris	150	4	761	765	749	747	731	752	744	735	742	749	742	730	732	743	734	733	728	737	723	741
Hepatitis	153	7	888	842	819	805	796	791	786	785	783	777	774	769	768	765	761	755	751	740	737	784
Strikes	163	4	812	778	766	759	754	760	755	755	755	757	760	758	757	758	760	758	756	756	752	761
Wine	178	13	899	908	902	907	906	907	907	904	910	908	911	910	914	913	918	918	919	922	924	911
Baseball CYY	189	9	-	828	790	764	761	746	731	721	716	711	698	699	693	686	682	682	680	679	680	719
Forensic Glass	214	9	899	868	848	843	841	838	837	833	834	833	830	831	830	829	830	831	834	834	840	840
Horse Colic	218	6	783	754	736	739	741	742	745	744	747	748	748	748	748	746	746	744	742	740	735	746
SPECT	267	22	855	818	798	783	777	771	763	758	752	751	742	737	738	727	727	720	712	711	703	755
Heart (Hungaria)	294	10	878	844	837	831	831	830	828	827	826	827	826	826	826	826	826	826	827	827	827	831
Heart (Cleveland)	297	13	867	839	821	815	814	808	805	800	804	801	800	799	798	799	798	798	798	796	795	808
Alcohol	345	5	-	798	789	788	787	785	783	783	783	783	783	783	783	783	784	783	784	785	784	785
Ionosphere	351	33	871	826	802	796	784	775	776	774	770	766	763	760	760	754	754	749	745	742	735	774
Tourism	365	4	780	753	739	733	723	720	717	713	712	710	708	706	706	704	702	703	703	705	708	718
Voting	402	16	950	943	944	947	947	947	949	949	951	951	952	952	953	953	953	953	954	954	954	950
Education	435	4	771	759	751	744	743	741	741	740	740	740	738	739	740	741	742	742	744	745	745	745
Cardiac	558	5	830	806	800	792	790	786	791	790	789	788	785	784	783	781	780	779	777	777	776	789
Football	672	4	695	674	665	661	660	659	657	658	656	654	655	655	654	654	654	653	652	653	653	659
Breast Cancer	683	9	970	966	963	962	960	960	960	959	959	958	958	959	958	958	958	958	958	958	958	960
Diabetes	786	6	772	747	739	734	730	729	724	723	724	721	720	718	717	716	715	715	714	714	714	726
Authorship	841	69	934	918	907	897	894	889	887	884	882	883	879	882	878	881	880	875	876	872	868	888
Baseball HoF	1320	15	807	773	763	750	743	740	735	731	727	726	726	722	721	721	720	719	718	717	716	735
Yeast	1484	8	715	698	691	690	688	686	686	686	685	686	684	684	684	682	682	682	681	680	680	687
Supreme Court	4052	7	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958
Hochschule	191	14	888	832	799	772	758	743	735	729	720	715	709	707	700	700	696	691	689	687	678	734
Beratung	437	16	825	780	752	739	729	725	720	715	714	715	713	711	711	711	711	711	711	711	712	727
Schule	1993	15	679	656	648	642	639	638	636	635	634	634	633	632	633	632	631	631	631	631	632	638
Ø			840	824	813	805	800	797	795	792	791	789	788	786	785	784	784	782	781	781	778	794

Tabelle D.32: Prediction-Ergebnisse für FaF ZigVal Fold (33 Datensätze): Die Prozentsätze geben den Anteil der Trainingsstichprobe an der Gesamtstichprobe an. Die Zahlen entsprechen der mit 1000 multiplizierten gerundeten durchschnittlichen Klassifikationsgenauigkeit aus 1000 Trials.

FaF ZigVal Fold			Prediction																			
Datensatz	n	Cues	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	Ø
Bankruptcy	50	5	-	605	702	752	767	801	812	817	823	833	823	827	829	828	820	823	811	804	808	794
Solvency	52	8	509	596	726	744	763	775	775	779	777	780	782	779	771	774	778	775	770	774	760	747
Echocardiogram	62	7	625	652	658	680	694	694	711	708	725	718	732	730	732	725	740	735	717	738	710	707
Post-Operative	86	8	631	635	640	655	658	675	675	670	673	676	687	682	693	694	690	703	699	711	706	677
Broadway	89	5	691	711	727	739	744	757	761	762	761	767	765	772	770	774	776	787	784	798	815	761
Credit Score	100	6	684	747	772	769	781	772	770	771	770	766	764	766	767	759	764	755	762	756	757	761
Zoo	101	16	807	892	916	926	932	935	940	944	941	944	944	943	948	943	945	947	947	944	945	931
Iris	150	4	565	632	650	674	669	698	695	691	703	713	707	693	696	708	687	689	675	685	651	678
Hepatitis	153	7	722	736	742	744	746	748	748	747	748	746	746	742	741	735	732	728	719	715	708	736
Strikes	163	4	666	707	721	726	729	734	732	733	738	741	741	742	744	743	745	747	746	743	741	732
Wine	178	13	678	806	839	862	871	880	884	888	896	896	900	900	906	907	911	913	916	920	919	878
Baseball CYY	189	9	-	725	716	711	707	704	695	691	686	684	679	679	673	668	671	668	671	669	673	687
Forensic Glass	214	9	728	764	786	797	802	803	807	808	808	811	811	813	812	811	815	816	820	824	829	804
Horse Colic	218	6	576	634	655	685	702	708	720	727	732	733	738	738	738	741	739	740	741	737	734	712
SPECT	267	22	666	694	704	708	712	716	711	710	706	712	705	705	705	701	700	697	694	687	689	701
Heart (Hungaria)	294	10	750	783	795	801	806	807	809	811	812	812	813	814	815	815	817	817	817	817	819	807
Heart (Cleveland)	297	13	698	735	744	749	758	758	762	765	766	767	770	771	776	776	776	775	779	774	771	761
Alcohol	345	5	-	747	756	767	768	774	775	777	777	778	779	779	779	779	782	784	785	782	788	775
Ionosphere	351	33	687	707	707	715	716	718	719	720	717	719	721	716	721	718	718	712	711	711	702	713
Tourism	365	4	663	678	684	681	683	682	681	681	681	679	682	680	678	677	681	680	679	679	694	680
Voting	402	16	890	913	923	932	935	941	941	945	947	947	948	950	950	951	952	953	953	953	951	941
Education	435	4	698	709	713	714	715	718	718	721	718	720	720	725	728	726	731	732	734	742	741	722
Cardiac	558	5	732	748	757	757	761	760	766	765	768	770	769	770	766	768	766	769	767	763	768	763
Football	672	4	611	623	632	635	638	639	641	642	643	643	643	643	645	645	645	644	646	648	646	640
Breast Cancer	683	9	932	944	949	951	952	953	953	954	954	955	955	955	956	956	957	957	957	956	958	953
Diabetes	786	6	699	710	714	715	716	716	716	715	713	714	713	713	713	712	713	713	713	718	715	713
Authorship	841	69	863	882	883	879	880	878	877	876	874	874	873	876	873	875	875	870	869	864	857	874
Baseball HoF	1320	15	786	759	752	742	737	735	730	727	724	722	723	720	719	718	720	716	716	714	716	730
Yeast	1484	8	655	667	671	675	677	676	678	679	680	679	680	680	680	681	679	678	678	674	675	676
Supreme Court	4052	7	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958
Hochschule	191	14	613	631	642	648	648	650	650	650	656	656	656	660	660	657	657	654	654	648	651	650
Beratung	437	16	647	676	679	685	689	693	694	695	697	698	698	701	701	703	705	708	706	708	713	695
Schule	1993	15	605	613	618	620	621	622	624	624	625	626	625	627	627	627	627	628	628	627	627	623
Ø			701	728	743	751	756	760	762	762	764	765	765	765	766	765	766	766	764	765	763	757