

Evolutionäre Merkmalsselektion und Suchraumpartitionierung für die datenbasierte Fuzzy–Modellierung hochkomplexer Systeme

von der
Fakultät für Elektrotechnik und Informationstechnik
der Universität Dortmund
genehmigte Dissertation
zur Erlangung des akademischen Grades
Doktor der Ingenieurwissenschaften

von

Daniel Schauten

Dortmund 2008

Tag der mündlichen Prüfung: 18. 03. 2008
Hauptreferent: Prof. em. Dr. rer. nat. H. Kiendl
Korreferent: Prof. Dr. - Ing. habil. Georg Bretthauer

Vorwort

Die vorliegende Arbeit entstand während meiner Tätigkeit am Lehrstuhl für Elektrische Steuerung und Regelung der Universität Dortmund unter der Betreuung von Herrn Prof. em. Dr. rer. nat. H. Kiendl. Für die Übernahme des Korreferats danke ich herzlich Herrn Prof. Dr. - Ing. habil. Georg Bretthauer. Der Deutschen Forschungsgemeinschaft danke ich für die finanzielle Unterstützung während meiner Arbeit im Rahmen des Sonderforschungsbereichs 531 (*Computational Intelligence*) der Universität Dortmund.

Ganz besonders danke ich meiner Frau Britta und meinen Kindern Dustin und Daria für die stetige emotionale Unterstützung, Geduld und Rückhalt, die es mir ermöglichten den zeitlichen Aufwand zu bewältigen. Weiterhin gilt mein besonderer Dank der Fakultät für Elektrotechnik und Informationstechnik für das vorzeitig geschenkte Vertrauen für eine längerfristige Zusammenarbeit.

Bei Herrn Dr. rer. nat. F. Hoffmann bedanke ich mich für die erfolgreiche Zusammenarbeit auf dem Gebiet der iterativen strukturellen Evolution von Fuzzy-Reglern. Ferner verbindet mich mit meinen Kollegen Dr.-Ing. M. Sternke, Dr.-Ing. P. Krause und Dr.-Ing. T. Slawinski eine mehrjährige Zusammenarbeit auf dem Gebiet der datenbasierten Fuzzy-Modellierung. Für vielfältige Simulationen und Untersuchungen einzelner Fragestellungen danke ich zudem den Studenten, deren Diplom- und Studienarbeiten ich betreuen durfte.

Weiterhin bedanke ich mich bei allen Mitarbeiterinnen und Mitarbeitern des Lehrstuhls für Regelungssystemtechnik für die vielfältige Unterstützung und die gute Zusammenarbeit in gemeinsamen Projekten. Wertvolle Hinweise inhaltlicher und formaler Art für die endgültige Fassung des Manuskripts verdanke ich insbesondere Frau Besten (Kasimir) und Herrn Prof. em. Dr. rer. nat. H. Kiendl.

Schließlich bedanke ich mich bei Herrn Prof. Dr.-Ing. Prof. h.c. T. Bertram als Lehrstuhlinhaber für den eingeräumtem Freiraum zur Fertigstellung meiner Arbeit.

Dortmund, April 2008

Daniel Schauten

Inhaltsverzeichnis

1	Einleitung	1
2	Einordnung und Zielsetzung	5
2.1	Ausgangssituation	5
2.1.1	Quantitative Modellierung	6
2.1.2	Datenbasierte Fuzzy-Modellierung	11
2.1.3	Merkmalsselektion zur Komplexitätsreduktion	19
2.1.4	Einfluss der Fuzzy-Partitionierung des Merkmalsraumes	21
2.2	Ziele dieser Arbeit	23
3	Methodische Grundlagen	26
3.1	Fuzzy-Logik	26
3.2	Fuzzy-Systeme	27
3.2.1	Mamdani-Fuzzy-Systeme	28
3.2.2	Takagi-Sugeno-Fuzzy-Systeme	34
3.3	Datenbasierte Fuzzy-Modellierung	35
3.3.1	Grundidee des Fuzzy-ROSA-Verfahrens	36
3.3.2	Der Regelgenerierungsprozess	37
3.3.3	Ansatzpunkte für die vorliegende Arbeit	38
3.4	Selektion relevanter Merkmale	38
3.4.1	Bewertungsfunktion	39
3.4.2	Suche von relevanten Sätzen von Einflussgrößen	41
3.5	Evolutionsstrategien	42
3.5.1	Einführung	42
3.5.2	Ablauf einer Evolutionsstrategie	44
3.5.3	Abbruchkriterium	51

4	Merkmalsselektion und Fuzzy-Partitionierung	52
4.1	Problemstellung	52
4.2	Rekonstruktionsbasierte Einflussgrößenselektion	53
4.3	Normierung der Einflussgrößen	54
4.3.1	Kontinuierliche Einflussgrößen	54
4.3.2	Gemischt kontinuierliche und nominale Einflussgrößen . .	56
4.4	Abstands-basierte Ausgangsgrößen-wertschätzung	57
4.4.1	Abstands-basierte Zugehörigkeitsfunktionen	57
4.4.2	Rekonstruktionsvorschriften	58
4.4.3	Anwendungsgrenzen der abstands-basierten Rekonstruktion	59
4.5	Zellen-basierte Ausgangsgrößen-wertschätzung	59
4.5.1	Partitionierung des Merkmals-raumes	60
4.5.2	Partialfunktionen	60
4.6	Relevanzbewertung eines Merkmalssatzes	62
4.7	Optimierung der Partitionierung des Merkmals-raumes	64
4.7.1	Kodierung der Partitionierung	64
4.7.2	Zusätzliche Zielfunktionen	67
4.8	Selektionsstrategien	68
4.8.1	Iterativer Selektionsprozess	68
4.8.2	Global operierender Selektionsprozess	69
4.9	Generierung effizienter Fuzzy-Zugehörigkeitsfunktionen	71
4.9.1	Eingangsseitige Zugehörigkeitsfunktionen	71
4.9.2	Ausgangsseitige Zugehörigkeitsfunktionen	73
5	Diskretisierung der Evolutionsstrategie	77
5.1	Motivation diskretisierter Suchräume	77
5.1.1	Vorzeitige Konvergenz in lokalen Optima	77
5.1.2	Bedeutung der Diversität einer Population	78
5.1.3	Motivation der Diskretisierung von Suchräumen	79
5.2	Statische Diskretisierung des Suchraums	80
5.2.1	Evolutionäre Suche in statisch diskretisierten Suchräumen	80

5.2.2	Diversifizierung der Population	82
5.2.3	Rekombination bei diskreten Suchräumen	83
5.2.4	Mutation bei diskreten Suchräumen	83
5.2.5	Selektion bei diskreten Suchräumen	84
5.3	Adaptive Diskretisierung des Suchraums	86
5.3.1	Motivation adaptiv diskretisierter Suchräume	86
5.3.2	Adaption der Suchraumdiskretisierung durch Gitterebenen	87
5.3.3	Evolutionäre Suche in adaptiv diskretisierten Suchräumen	89
5.3.4	Steuerung der adaptiven Suchraumdiskretisierung	91
5.4	Effizienzvergleich anhand analytischer Testfunktionen	92
5.4.1	Testfunktionen	92
5.4.2	Bewertungskriterien	98
5.4.3	Ergebnisse	100
6	Komplexe Anwendungen	104
6.1	Benchmarkprobleme	104
6.1.1	Klassifikation von Schwertlilien (IRIS)	104
6.1.2	Bestimmung der Herkunft von Weinen (WINE)	105
6.1.3	Klassifikation von Intron–Exon–Übergängen (DNA)	106
6.1.4	Klassifikation der Erdoberfläche in Satellitenbildern (SAT)	108
6.1.5	Vorhersage der Mackey–Glass–Zeitfolge (MACKEY)	108
6.1.6	Modellierung von Wohnungspreisen in Boston (BOSTON)	109
6.2	Industrielle Anwendungen	111
6.2.1	Kurzfristige Lastprognose für die Kraftwerksführung	111
6.2.2	Adaption von Strategieparametern der Kraftwerksführung	114
6.2.3	Qualitätskontrolle von Automatikgetrieben	119
6.2.4	Strömungskraftschätzung eines Hydraulikventils	121
6.2.5	Evolution von TSK–Fuzzy–Reglern	128
7	Zusammenfassung und Ausblick	135

A	Eigenschaften einer Fuzzy-Menge	139
B	Statistische Tests	141
C	Least-Squares-Algorithmus	144
	Literaturverzeichnis	146

Kapitel 1

Einleitung

Eine dominierende Aufgabenstellung in den ingenieurwissenschaftlichen und wirtschaftswissenschaftlichen Disziplinen besteht in der Analyse, Überwachung und gezielten Beeinflussung von Systemen. Dabei bestimmt häufig das Vorhandensein eines geeigneten Systemmodells den Erfolg. Das Modell eines realen Systems kann einerseits passiv zur Systembeobachtung eingesetzt werden, wie beispielsweise im Rahmen einer Qualitätssicherung, der Prognose zukünftiger Systemzustände oder zur Schätzung nichtmessbarer Systemzustände. Demgegenüber kann ein Modell auch aktiv am realen System oder in einer Computersimulation eingesetzt werden. So können gezielte Systemeingriffe, wie unter anderem Steuerungs-, Regelungs- oder Parameteradaptionstrategien, simulativ getestet und optimiert oder real durchgeführt werden.

Die zunehmende Vernetzung technischer Systeme einerseits sowie steigende Genauigkeitsanforderungen andererseits steigern zunehmend die Komplexität der Modellierungsprobleme, so dass an die Modellierungsverfahren immer höhere Anforderungen gestellt werden. Von besonderem Interesse sind dabei robuste Modellierungsverfahren, mit denen qualitatives Wissen verarbeitet werden kann und die auch bei unscharfer oder unvollständiger Information in hochdimensionalen Variablenräumen gute Resultate in akzeptabler Zeit erzielen können. Aufgrund dieser hohen Ansprüche ist in einer Vielzahl von Anwendungsfällen eine rigorose wissensbasierte physikalisch-mathematische Modellierung nicht mehr möglich.

Zur Begegnung dieser Problematik wurden zahlreiche Verfahren zur *datenbasierten Modellierung* entwickelt. Die Zielsetzung der Modellierung ist in den einzelnen Anwendungsfällen sehr unterschiedlich. In den Bereichen der Wirtschafts- und Sozialwissenschaften steht häufig allein der Aspekt der Beschreibung von Wirkungszusammenhängen, die sogenannte *qualitative Modellbildung* und damit die *Transparenz* der Modelle, im Vordergrund. In den ingenieurwissenschaftlichen oder technischen Anwendungsgebieten ist zudem die

quantitative Modellierung im Sinne der erzielbaren *Genauigkeit* und der mit der Modellerstellung verbundene rechentechnische *Aufwand* von sehr großer Bedeutung. Für die aktiven Modelle in Echtzeitsimulationen oder mit einer realen Systemanbindung spielt zudem der Rechenaufwand für die Modellauswertung eine übergeordnete Rolle.

Diesbezüglich haben sich in der jüngeren Vergangenheit Modellierungsansätze aus dem Bereich der *Computational Intelligence* (CI) hervorgehoben. Die derzeit wichtigsten Methoden der CI sind die Fuzzy-Logik (FL), die künstlichen neuronalen Netze (KNN) und die evolutionären Algorithmen (EA). All diese Verfahren sind in Anlehnung an die Leistungsfähigkeit der Natur entwickelt worden. So bildet zum Beispiel die Anwendung der Fuzzy-Logik das menschliche Problemlösungsverhalten nach und künstliche neuronale Netze sind in Anlehnung an die neuronale Struktur des Gehirns höherer Lebewesen konzipiert worden. Die für den Anwender wichtigen Gesichtspunkte *Transparenz* und *Genauigkeit* sind in der Praxis häufig gegenläufige Ziele. Quantitative nichtlineare Modelle wie Neuronale Netze gelten allgemein als leistungsfähige Universalapproximatoren, die einmal antrainierte Lerndatenpunkte sehr genau und mit kurzen Auswertezeiten wiedergeben können. Demgegenüber sind diese Modelle nicht transparent und somit einer direkten Interpretation nur sehr selten zugänglich. Das intransparente Verhalten Neuronaler Netze wird dadurch offenbart, indem sie trotz sehr genauer Wiedergabe gelernter Datenpunkte häufig ein unbefriedigendes interpolierendes Verhalten und in der Regel ein inakzeptables extrapolierendes Verhalten bei Eingabe unbekannter Datenpunkte zeigen. Eine weitere Schwierigkeit liegt zudem in der geeigneten Wahl der zum Training verwendeten Struktur.

Einen Ansatz zur Generierung qualitativer Modelle bilden regelbasierte Modelle, die Wirkungszusammenhänge in Form von interpretierbaren WENN-DANN-Regeln beschreiben, was in den Ingenieurwissenschaften großen Anklang gefunden hat. Der Vorteil dieser Modellart ist die einfache Nachvollziehbarkeit von Wirkungszusammenhängen eines Systems aufgrund der dem Menschen nahestehenden Wissensrepräsentation des Modells. Dadurch ist es zudem möglich, ein gegebenes Modell einfach zu erweitern und zu modifizieren, um es an veränderte oder neue Situationen anzupassen. Basierend auf der von ZADEH 1965 eingeführten Fuzzy-Logik ist die Erweiterung der regelbasierten Modelle zu Fuzzy-Modellen vollzogen geworden. Im Gegensatz zur klassischen booleschen Logik, die nur die Wahrheitswerte *wahr* oder *falsch* kennt, können mit der Fuzzy-Logik auch Wahrheitswerte *zwischen* wahr und falsch verarbeitet werden, wodurch die Modellierungsgüte gegenüber den regelbasierten Modellen deutlich verbessert werden konnte. Durch die eindrucksvolle Verknüpfung von

Interpretierbarkeit mit einer meist ausreichend hohen Abbildungsgenauigkeit werden Fuzzy-Modelle in der Praxis häufig bevorzugt.

Datenbasierte Fuzzy-Modellierungsverfahren basieren auf der Auswertung von akkumulierten Beobachtungsdaten zur Bestimmung der Zugehörigkeitsfunktionen und der Fuzzy-Regeln. Durch den Einsatz immer leistungsfähigerer Computer und ausgefeilterer Methoden können mittlerweile sehr genaue Modelle generiert werden, die sehr gut mit dem zu modellierenden System übereinstimmen. Ein etabliertes effizientes Verfahren hierfür ist das von KIENDL, KRABS und KRONE entwickelte Fuzzy-ROSA¹-Verfahren [KK89, Kra94, KK95, Kro99], welches auch im Rahmen dieser Arbeit verwendet wird. Zur Beherrschung immer komplexerer Systeme ist der Bedarf, den Anwendungsbereich der datenbasierten Fuzzy-Modellierung in Richtung höherer Komplexität auszudehnen, ungebrochen. Bisher sind hierzu in das Fuzzy-ROSA-Verfahren von SLAWINSKI [Sla01] effiziente Analyse- und Regelsuchalgorithmen für hochdimensionale Suchräume und von KRAUSE [Kra01] Verfahren zur Erhöhung der Genauigkeit unter Beibehaltung der Interpretierbarkeit implementiert worden.

Das wesentliche Ziel der vorliegenden Arbeit ist die Weiterentwicklung der datenbasierten Fuzzy-Modellierung für hochkomplexe Systeme, die insbesondere durch eine hohe Anzahl von potenziellen Einflussgrößen unbekannter Relevanz gekennzeichnet sind. Der Schwerpunkt liegt dabei in der Entwicklung und dem Einsatz effizienter Verfahren für eine der eigentlichen Modellierung vorgeschalteten Merkmalsselektion. Dadurch soll einerseits im Rahmen der Modellgenerierung unter Erhalt der Transparenz die Modellgenauigkeit möglichst gesteigert und der rechentechnische Aufwand zur Modellgenerierung und insbesondere auch zur Modellauswertung deutlich verringert werden.

Die Arbeit ist folgendermaßen strukturiert: In Kapitel 2 werden die dieser Arbeit zugrunde liegende Problemstellung zur datenbasierten Modellierung herausgearbeitet und die daraus resultierenden Ziele präzisiert.

Zur Einlösung dieser Ziele werden zunächst in Kapitel 3 die im Folgenden benötigten methodischen Grundlagen hinsichtlich der datenbasierten Fuzzy-Modellierung, der Selektion relevanter Merkmalsätze und von Evolutionsstrategien zusammengestellt.

Den zentralen Ansatz dieser Arbeit beinhalten die Kapitel 4 und 5. Kapitel 4 greift die aus Kapitel 3 herausgearbeitete Problematik auf, dass es bei einer großen Zahl von gegebenen primären potenziellen Einflussgrößen nicht sinnvoll ist, alle diese Größen für die datenbasierte Modellierung heranzuziehen. Im Allgemeinen führt eine Beschränkung auf einen kleineren Satz von möglichst re-

¹Regel-Orientierte Statistische Analyse

levanten und nichtredundanten Einflussgrößen schneller und mit einem besseren Ergebnis zum Ziel. Die hierfür bekannten Auswahlverfahren werden analysiert. Basierend auf den dabei festgestellten Vor- und Nachteilen wird dann ein neues Verfahren vorgeschlagen, das im Sinne der oben genannten Ziele besonders leistungsfähig ist.

In Kapitel 5 wird davon ausgegangen, dass es bei einer sehr hohen Komplexität des Modellierungsproblems nicht mehr möglich ist, alle im Prinzip vorhandenen Freiheitsgrade des Auswahlverfahrens und der nachgeschalteten datenbasierten Modellierung zu testen und so die absolut günstigste Lösung aufzufinden. Deshalb benötigt man ein effizientes Optimierungsverfahren, das den Suchraum bevorzugt in den besonders aussichtsreichen Gebieten erforscht. Für den Anwender ist dabei entscheidend, dass ein möglichst günstiger Kompromiss zwischen den Gesichtspunkten *Rechenaufwand* und *resultierende Modellgüte* hergestellt wird. Deshalb werden im Hinblick auf dieses Ziel bekannte Varianten evolutionärer Algorithmen analysiert und im Anschluss eine besonders leistungsfähige Variante vorgeschlagen, die auf einer adaptiven Diskretisierung des Suchraumes basiert.

Die Leistungsfähigkeit der neu entwickelten Strategieelemente wird in Kapitel 6 anhand von Benchmarkproblemen und industriellen Anwendungen demonstriert.

Eine Zusammenfassung schließt diese Arbeit mit Kapitel 7 ab.

Kapitel 2

Einordnung und Zielsetzung

2.1 Ausgangssituation

In diesem Kapitel wird, ausgehend von dem übergeordneten Themenkomplex der quantitativen Modellierung, eine Präzisierung der dieser Arbeit zugrunde liegenden Aufgabenstellung und Zielsetzung vorgenommen. Zunächst wird dafür die Leistungsfähigkeit der datenbasierten Fuzzy-Modellierung aufgezeigt. Von essenzieller Bedeutung für die Modellierungsverfahren ist dabei der Grad der Komplexität des zu lösenden Modellierungsproblems. Diese wird zum einen durch die Anzahl der beteiligten Systemgrößen bestimmt, die direkt als potenzielle Einflussgrößen bei der Fuzzy-Modellierung verwendet werden können. Zum anderen wird die Komplexität durch die Art der Abhängigkeiten zwischen den einzelnen Systemgrößen maßgeblich beeinflusst. Fehlendes Expertenwissen über das betrachtete System bedingt darüber hinaus Methoden zur automatisierten Fuzzy-Partitionierung der bei der Modellierung verwendeten Systemgrößen und führt dementsprechend ebenfalls beträchtlich zu einer Verminderung der Komplexität. Trotz vielfältiger methodischer Weiterentwicklungen vieler Verfahren zur datenbasierten Fuzzy-Modellierung sind damit vergleichsweise geringe Fortschritte in der Richtung erzielt worden, dass hochkomplexe Systeme besser beherrscht werden können. In den einschlägigen Publikationen sind keine wesentlichen Erfolge zur Ausdehnung ihrer Anwendungsgrenzen hinsichtlich dieser Komplexität von Modellierungsproblemen zu finden. Statt dessen zeigen Ansätze zu einer der Modellierung vorgeschalteten Komplexitätsreduktion wesentlich bessere Erfolge. Insbesondere Verfahren zur Selektion von Eingangsgrößen — der sogenannten *Merkmalsselektion* — entlasten die Modellierungsverfahren erheblich und ermöglichen meist die Generierung von genaueren Modellen mit zudem verminderter Komplexität. Vor diesem Hintergrund werden die generellen Anforderungen, die im Rahmen dieser Arbeit an die zu entwickelnden Verfahren zur Merkmalsselektion zu stellen sind, wie folgt for-

muliert: Durch aufeinander abgestimmte neue Verfahren zur Merkmalsselektion und Fuzzy-Partitionierung soll auch für sehr komplexe Anwendungen die Generierung interpretierbarer quantitativer Fuzzy-Modelle automatisch, d. h. auch bei wenig verfügbarem Systemwissen, ermöglicht werden.

Hinsichtlich dieses Anforderungsprofils wird zunächst in Abschnitt 2.1.1 die Aufgabenstellung der quantitativen Modellbildung sowie die prinzipielle Vorgehensweise der datenbasierten Modellierung vorgestellt. Darauf aufbauend wird in Abschnitt 2.1.2 der Einsatz der Fuzzy-Logik motiviert und die Vor- und Nachteile verschiedener Ansätze für die datenbasierte Fuzzy-Modellierung miteinander verglichen. Dabei wird auch eine Einordnung des in dieser Arbeit verwendeten Fuzzy-ROSA-Verfahrens vorgenommen. Nach einem Exkurs in die COD-Problematik¹ (Abschnitt 2.1.3) und einer kurzen Darstellung des Einflusses der Fuzzy-Partitionierung (Abschnitt auf das Modellierungsergebnis 2.1.4) werden schließlich in Abschnitt 2.2 Ansatzpunkte für eine Ausdehnung des Anwendungsfelds der datenbasierten Fuzzy-Modellierung in Richtung höherer Komplexität durch geeignete Methoden der Datenvorverarbeitung herausgearbeitet und konkrete Ziele für diese Arbeit formuliert.

2.1.1 Quantitative Modellierung

Aufgabenstellung

Ausgangspunkt dieser Arbeit ist die quantitative Modellierung, deren Aufgabenstellung in Abbildung 2.1 illustriert ist. Das grundlegende Ziel besteht dabei in einer möglichst guten Nachbildung des Eingangs-/Ausgangsverhaltens eines betrachteten MISO²-Systems mit n Eingangsgrößen X_i , $i \in \{1, \dots, n\}$ und einer Ausgangsgröße Y .

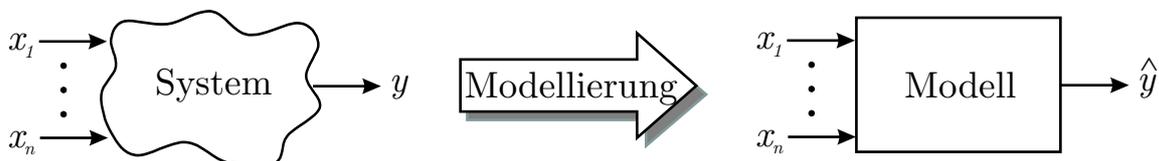


Abbildung 2.1: Aufgabenstellung der quantitativen Modellierung.

Der Begriff System ist dabei so weitläufig gefasst, dass damit sowohl konkrete reale technische Systemen als auch virtuelle Systemen, wie beispielsweise wirt-

¹engl.: *Curse of Dimensionality*

²engl.: *Multi Input Single Output*

schaftliche Prozessabläufe, erfasst werden können. Die Abbildung des Eingangs-/Ausgangsverhaltens erfolgt bei komplexen Systemen häufig vereinfachend, da in den meisten Fällen nie der ganze Umfang eines zu modellierenden Systems erfasst werden kann. Das entworfene Modell enthält in der Regel eine Berechnungsvorschrift, mit der der interessierende Ausgangsgrößenwert y_j des in der j -ten Situation (z. B. im Zeitpunkt t_j) betrachteten Systems auf Grundlage der anstehenden Eingangsgrößenwerte $x_{1,j}, \dots, x_{n,j}$ möglichst genau ermittelt werden kann. Die Modellierungsgüte ist dabei umso besser, je weniger die berechneten Werte \hat{y}_j von den tatsächlichen Ausgangsgrößenwerten y_j im Mittel aller j gegebenen oder beobachteten Eingangssituationen abweicht.

Typisierung der Systemgrößen

In Anlehnung an die deskriptive Statistik können die beobachtbaren Systemgrößen in drei verschiedene Typen eingeteilt werden:

- *Nominale Größen* nehmen ausschließlich symbolische Werte an, die nicht miteinander vergleichbar sind und keiner Rangfolge bezüglich einer kleiner–größer Relation angeordnet werden können. Beispielsweise können die Werte einer Systemgröße *Beruf* die Werte *Bäcker*, *Schlosser* und *Verkäufer* annehmen. Im Rahmen dieser Arbeit werden prinzipiell die verschiedenen symbolischen Werte einer nominalen Größe als von eins an aufsteigende natürliche Zahlen kodiert.
- *Ordinale Größen* nehmen ebenfalls wie nominale Größen ausschließlich symbolische Werte an, die sich jedoch in ihrer Intensität unterscheiden und sich somit bezüglich einer kleiner–größer Relation anordnen lassen. Die Abstände zwischen den symbolischen Werten lassen sich jedoch nicht interpretieren. Als Beispiel hierfür sei die Systemgröße *Temperatur* angeführt, die in Form qualitativer Temperaturangaben die symbolischen Werte *kalt*, *normal*, *warm* und *heiß* annehmen kann. Im Rahmen dieser Arbeit werden ordinale Größen ebenso wie nominale Größen behandelt und entsprechend kodiert.
- *Kontinuierliche Größen* können beliebige reelle Werte annehmen, die einer festen metrischen Skala unterliegen. Alle Werte lassen sich bezüglich einer kleiner–größer Relation anordnen. Die Abstände zwischen zwei Werten lassen sich interpretieren und mit anderen Abständen ebenfalls im Sinne einer kleiner–größer Relation vergleichen. Zum Beispiel stellen die realen in Grad Celsius gemessenen Werte einer Systemgröße *Temperatur* eine kontinuierliche Größe dar.

Eine andere gebräuchliche Art der Klassifizierung ist die Einteilung in *stetige* und *diskrete Größen*, anhand derer auch eine Typisierung der Modellierungsprobleme vorgenommen wird. *Diskrete Größen* können ausschließlich eine endliche Zahl von Werten annehmen. Hierzu zählen demnach die oben eingeführten nominalen und ordinalen Größen. Demgegenüber bezeichnet man die kontinuierlichen Größen auch als *stetige Größen*, die beliebig viele Werte in einem bestimmten Bereich annehmen können.

Typisierung der Modellierungsaufgaben

Generell werden zwei Einsatzmöglichkeiten der quantitativen Modellierung voneinander unterschieden. Unter *Approximation* wird im Rahmen dieser Arbeit die reine Nachbildung des Eingangs-/Ausgangsverhaltens eines Systems mit stetiger Ausgangsgröße y verstanden. Zur Beurteilung der Modellgenauigkeit des erhaltenen Modells wird in erster Linie der mittlere quadratische oder mittlere absolute Fehler herangezogen (Abschnitt 4.6). Anwendung finden solche Modelle in der Simulation, für die Kennfeldnachbildung, bei der Prognose oder beim Wissenserwerb im Sinne einer Prozessanalyse. Eine Simulation kann beispielsweise im Rahmen eines Reglerentwurfs oder dessen Optimierung sehr nützlich sein, indem man Kosten vermindert und Risiken einschränkt, die bei Experimenten am realen System auftreten würden. Ein Beispiel hierfür ist die Modellierung der Strömungskraft für ein Hydraulikventil (Abschnitt 6.2.4) oder die Adaption von Strategieparametern für ein Kraftwerksführungssystem in der Energieversorgung (Abschnitt 6.2.2). Eine spezielle Aufgabenstellung der Approximation ist die *Prognose*, bei der aus bekannten Werten ein zukünftiger Wert y vorhergesagt werden soll, wie bei der kurzfristigen Lastprognose in der Energieversorgung (Abschnitt 6.2.1).

Im Gegensatz zu der Problemklasse der Approximation wird bei der *Klassifikation* kein stetiger kontinuierlicher Ausgangsgrößenwert ermittelt, sondern immer ein diskreter Ausgangsgrößenwert (auch *Klasse* genannt) ausgegeben. Die zugrunde liegende Modellierungsaufgabe besteht somit in dem Entwurf eines Klassifikators, der anhand der anstehenden Eingangsgrößenwerte $x_{1,j}, \dots, x_{n,j}$ der j -Situation den geeignetsten Ausgangsklassenwert \hat{y}_j aus der Menge \mathcal{Y} mit $\mathcal{Y} = \{\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_\eta\}$ aller η möglichen Klassen ausgibt. In [BP92] wird diese Aufgabe auch Parameterbestimmung einer Klassifizierungsfunktion genannt, und die Lernaufgabe wird mit *Supervised Learning* bezeichnet, weil die Zuordnung zu den Klassen in den Lerndaten bereits vorliegen muss (*gelabelte Daten*). Zur Beurteilung der Qualität eines entworfenen Klassifikators wird üblicherweise ein Fehlermaß wie der relative Klassifikationsfehler (Abschnitt 4.6) ver-

wendet. Neben vielen etablierten Benchmarkproblemen (Abschnitt 6.1) wird in dieser Arbeit als Anwendungsbeispiel für ein Klassifikationsproblem die Getriebeklassifikation (Abschnitt 6.2.3) vorgestellt, bei der anhand akustischer Qualitätsmerkmale entschieden werden soll, ob ein Getriebe in Ordnung ist oder nicht. Abschließend sei hier noch angemerkt, dass in der Literatur (z. B. [Zim95, HKK97]) unter Klassifikation auch der Prozess der Zusammenfassung ähnlicher Objekte (z. B. mit Clusterverfahren) im Sinne einer *Klassenbildung* verstanden wird. Das Ziel einer derartigen Klassifikation ist somit das Aufspüren von Strukturen in gegebenen Daten und kann daher nicht der quantitativen Modellierung zugeordnet werden.

Ansätze zur quantitativen Modellierung

Infolge des breiten Anwendungsfeldes werden in der Literatur zahlreiche Ansätze für die quantitative Modellierung vorgeschlagen. Wie in Kapitel 1 bereits ausgeführt, unterscheiden sich die verschiedenen Modellierungsverfahren zum einen im Umfang des benötigten Vorwissens, um zum Beispiel die Wahl eines geeigneten Strukturansatzes für die Modellierung treffen zu können. Zum anderen differieren die generierten Modelle je nach verwendetem Modellierungsverfahren hinsichtlich der erzielbaren Modellierungsgüte und auch bezüglich der Interpretierbarkeit. Zusätzlich sind teilweise sehr unterschiedliche Anwendungsgrenzen der Modellierungsverfahren für komplexe Aufgabenstellungen zu verzeichnen. Im Folgenden wird ein kurzer Überblick über existierende Ansätze zur quantitativen Modellierung gegeben und darauf aufbauend genauer auf die datenbasierte Modellierung eingegangen.

Ein möglicher Ansatz zur Erstellung eines Modells ist die sogenannte *mathematisch-physikalische* Modellierung [LW98]. Dabei werden für die Wirkungszusammenhänge des Systems die entsprechenden beschreibenden physikalischen Gleichungen aufgestellt und gelöst. Für den Fall dass alle Systemparameter bekannt und die Gleichungen auch wirklich eindeutig lösbar sind, kann das ermittelte Modell direkt zur Berechnung des Eingangs-/Ausgangsverhaltens genutzt werden. Diese *wissensbasierte Modellierung*, auch *theoretische Analyse* genannt, setzt ein tiefgreifendes Verständnis der betrachteten Systeme voraus und ist für komplexere Anwendungen meist ungeeignet oder zumindest sehr zeitaufwändig. Für den Fall einiger unbekannter Systemparameter können diese im Rahmen einer so genannten *experimentellen Analyse* anhand am realen System gemessener Eingangs- und Ausgangssignale identifiziert werden. Diese Experimente bieten zudem eine Möglichkeit zur Verifikation des gewählten Strukturansatzes. Aufgrund der Nachvollziehbarkeit der ermittelten Modellglei-

chungen, die auf innerer Ursache–Wirkungszusammenhänge des Systems basieren, sind die erhaltenen quantitativen Modelle transparent. Zusätzlich ermöglichen sie analytische Betrachtungen, die unter Umständen auch Rückschlüsse auf das Systemverhalten für ein Kontinuum an möglichen Systemparametern oder Betriebssituationen zulassen. In Anlehnung an die Vorgehensweise dieser Modellbildung und die damit verbundene Transparenz der generierten Modelle wird dieser Ansatz auch *deduktive* oder *white–box* Modellierung genannt.

Dem gegenüber ist beim anderen Ansatz der *datenbasierten* Modellierung kein tieferes Verständnis der Systemzusammenhänge für eine Nachbildung des Eingangs-/Ausgangsverhaltens erforderlich. Bei diesem Ansatz werden zunächst Daten, die möglichst alle relevanten Systemzusammenhänge oder Betriebssituationen abdecken, am realen System erhoben. Auf Grundlage dieser Beobachtungsdaten wird dann mittels eines sogenannten *Lernverfahrens* induktiv ein Modell erstellt. Für eine solche rein quantitative Modellierung wird beispielsweise bei statistischen Modellierungsansätzen eine allgemeine Funktion angesetzt und deren freie Parameter werden mit Hilfe einer Regressionsanalyse anhand der Beobachtungsdaten ermittelt [SS95]. Im Bereich der Computational Intelligence werden für die gleiche Aufgabenstellung u. a. künstliche neuronale Netze eingesetzt [Kin94, Roj96]. Künstliche neuronale Netze haben den biologischen Aufbau eines Gehirns zum Vorbild. Sie liefern zwar quantitativ häufig genaue Modelle, sind aber von der Anschauung her in der Regel nicht für den Menschen verständlich und somit auch nicht interpretierbar. Dementsprechend werden solche Ansätze auch als *induktive* oder *black–box* Modellierung bezeichnet. Die generelle Vorgehensweise zur datenbasierten Modellierung wird im nachfolgenden Abschnitt 2.1.2 am Beispiel der datenbasierten Fuzzy–Modellierung vorgestellt.

Methoden zur Fuzzy–Modellierung generieren eine Wissensrepräsentation, bei der die in den Daten gefundenen Systemzusammenhänge in Form von intuitiv verständlichen WENN–DANN–Regeln ausgedrückt werden. Diese Methoden stehen gegenwärtig im Zentrum des Forschungsinteresses. Dies ist insbesondere dadurch motiviert, dass die Interpretierbarkeit solcher generierter Fuzzy–Modelle einen entscheidenden positiven Einfluss auf die Akzeptanz in industriellen Anwendungsfeldern ausübt. Von entsprechend hohem Interesse sind datenbasierte Ansätze, mit denen interpretierbare Modelle generiert werden können. In Abgrenzung zu der Exaktheit der allumfassenden mathematischen Modellierung wird diese mehr qualitative Beschreibung der Zusammenhänge durch Fuzzy–Regeln in der Literatur auch *grey–box* Modellierung genannt. Neben der Interpretierbarkeit bieten Fuzzy–Modelle den Vorzug, die generierte Wissensrepräsentation in Form der sogenannten *Regelbasis* nach eingehender Analyse

bearbeiten oder auch ergänzen zu können, indem bestehende Regeln geändert oder der Regelbasis neue Regeln hinzugefügt werden. Dies ist vor allem dann interessant, wenn für bestimmte Systemzustände keine Daten verfügbar sind oder etwa aus sicherheitstechnischen Gründen nicht generiert werden können.

2.1.2 Datenbasierte Fuzzy–Modellierung

Im Rahmen dieser Arbeit werden Ansätze zur Modellierung mit Hilfe von Daten nur dann als *datenbasierte Modellierung* bezeichnet, wenn für ihren Einsatz kein oder nur sehr wenig Systemverständnis notwendig ist. Damit ist eine klare Abgrenzung gegenüber den wissensbasierten Ansätzen möglich. Die generelle Vorgehensweise hierzu wird im Folgenden illustriert.

Ablauf einer datenbasierten Modellierung

Die datenbasierte Modellierung lässt sich grob in die drei Schritte *Datenaufbereitung*, *Merkmalsbildung* und *Modellbildung* einteilen. Diese werden im Folgenden, anhand von Abbildung 2.2 beschrieben. Speziell für die datenbasierte Fuzzy–Modellierung erforderliche zusätzliche Schritte werden im nachfolgenden Abschnitt bei der Beschreibung der verschiedenen Ansätze für eine datenbasierte Fuzzy–Modellierung vorgestellt. Weiterführende Betrachtungen hierzu sind z. B. in [BP92, Zim95, CM98] zu finden.

Datensatzbildung: Zunächst wird in Form einer *Datenaufnahme* am realen System eine Stichprobe mit N Rohdaten von *potenziellen Einflussgrößen* X_1, X_2, \dots, X_n und der zugehörigen *Ausgangsgröße* Y aufgenommen. Diese Datenpunkte sollten dabei für eine möglichst repräsentative Auswahl von Systemzuständen erhoben werden. Im Rahmen einer explorativen *Datenanalyse* werden fehlerhafte, inkonsistente oder mit Unsicherheiten behaftete Datenpunkte, welche beispielsweise durch Messfehler auftreten können, eliminiert. Ebenso können einfache Methoden zur *Datenaufbereitung*, wie beispielsweise eine Normierung, Filterung, Glättung oder Offset–Bereinigung, zum Einsatz kommen. Ergebnis der Datensatzbildung ist der *primäre Datensatz* \mathcal{D} , bestehend aus den N Datenvektoren der in unterschiedlichen Situationen j (z. B. Zeitpunkte t_j) gemessenen Datenpunkte $\mathbf{d}_j = (\mathbf{x}_{1,j}, \mathbf{x}_{2,j}, \dots, \mathbf{x}_{n,j}, \mathbf{y}_j)$, $j = 1, 2, \dots, N$.

Merkmalsbildung: Die Merkmalsbildung hat die Aufgabe aus der Menge \mathcal{X} aller primären potenziellen Einflussgrößen X_i einen möglichst kleinen Satz

von relevanten und nichtredundanten Größen, einen sogenannten *Merkmalsatz* \mathcal{M} , zu bilden. Die Werte der aus der Merkmalsbildung hervorgegangenen q Merkmale M_i und der zugehörige Ausgangsgröße Y bilden die *Lernstichprobe* $\mathcal{D}_L = (\mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_q, \mathbf{y})$ für das nachfolgende Modellierungsverfahren. Dieser Prozess der Merkmalsbildung hat eine entscheidende Bedeutung für den Erfolg der nachgeschalteten datenbasierten Modellierung: Eine Modellierung auf Grundlage des selektierten Merkmalsatzes \mathcal{M} kann zu einer höheren Modellgenauigkeit, als dies anhand aller primären Einflussgrößen möglich wäre, führen. Die nachgelagerten Verfahren zur Modellbildung werden durch die Beschränkung der Suchraums auf die relevantesten Größen erheblich entlastet und können zudem durch die verringerte Dimensionalität des Modellierungsproblems Modelle mit deutlich reduzierter Komplexität generieren (s. a. Abschnitt 2.1.3). Die einzelnen Merkmale M_i , $i = 1, 2, \dots, q$ des Merkmalsatzes \mathcal{M} sind somit die späteren Eingangsgrößen des Modells und können durch Verfahren der *Merkmalswahl* einerseits direkt aus der Menge der gemessenen potenziellen Einflussgrößen X_i selektiert werden. Andererseits können im Rahmen einer *Merkmalsgenerierung* bei der Merkmalsauswahl auch zusätzlich von den primären Einflussgrößen abgeleitete oder aggregierte *sekundäre Merkmale* berücksichtigt werden. Die Konzeption von Methoden zur datenbasierten Selektion relevanter und nichtredundanter Merkmale für die datenbasierte Fuzzy-Modellierung ist ein wesentlicher Gegenstand dieser Arbeit. Anforderungen und Ziele hierzu werden in Kapitel 2.2 ausführlich erläutert.

Modellbildung: Vor der eigentlichen *Modellidentifikation* sind in der Regel zunächst noch eine Reihe von *Strategieparametern* des ausgewählten Modellierungsverfahrens für ein möglichst gutes Modellierungsergebnis festzulegen. Daraufhin wird auf Basis der erstellten Lernstichprobe datenbasiert ein Modell generiert. Eine anschließende *Bewertung* wird häufig durch eine Validierung des Modells auf den verwendeten Lerndaten vorgenommen. Eine Analyse der erzielten Modellgüte sowie sonstiger interessierender Modelleigenschaften gibt Aufschluss darüber, ob eine erneute Modellidentifikation mit veränderten Strategieparametern erfolgversprechend erscheint oder ob mittels geeigneter Algorithmen zur Struktur- oder Parameteroptimierung noch leichte Verbesserungen von gewünschten Modelleigenschaften notwendig sind. Zur Vermeidung einer zu starken Anpassung des generierten Modells an die Lerndaten (*Overfitting*) wird die *Generalisierungsfähigkeit* meist durch eine Validierung auf einer *Teststichprobe* durchgeführt. Diese Teststichprobe besteht aus einem Teil der Lernstichprobe, der ausschließlich der Validierung vorbehalten ist und somit zur

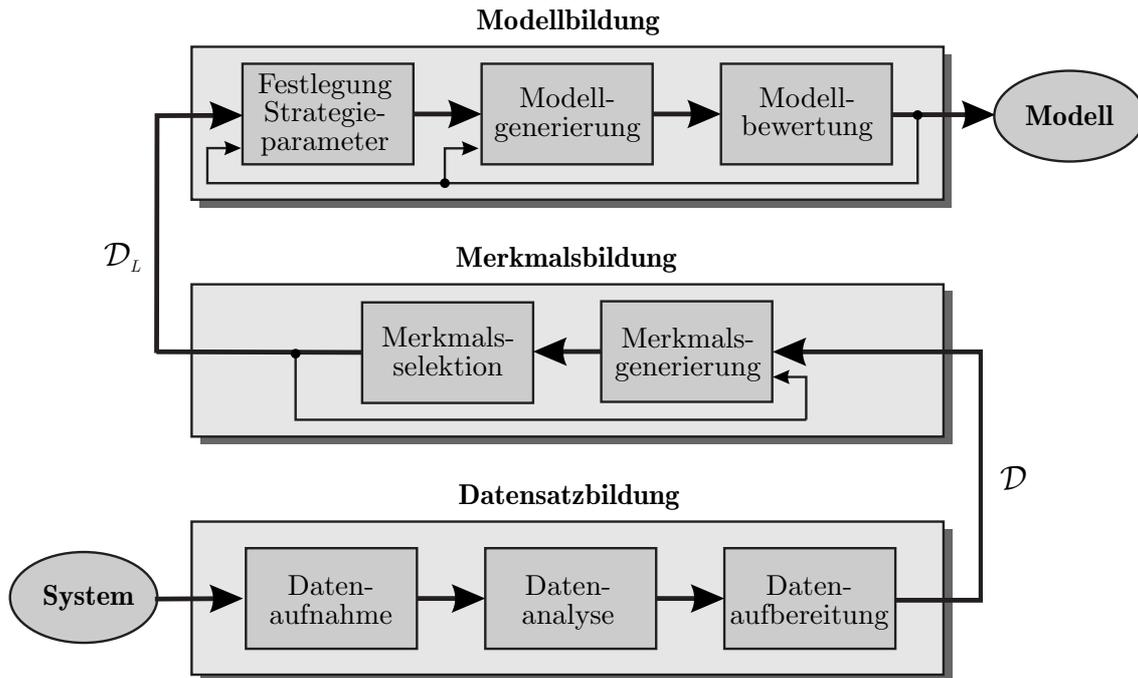


Abbildung 2.2: Ablauf der datenbasierten Modellierung.

eigentlichen Modellidentifikation nicht verwendet wird. Wird auf diesen Testdaten eine ähnliche Prognosegüte wie auf den Lerndaten, die zur Modellierung verwendet wurden, erzielt, so kann daraus eine gute Generalisierungsfähigkeit des generierten Modells abgeleitet werden. Eine detaillierte Beschreibung verschiedener Maße zur Bewertung der Prognosegüte findet sich in Kapitel 4.6.

In der Praxis zeigt sich, dass eine streng sukzessive Abarbeitung dieser Schritte selten möglich ist. Gerade bei komplexeren Problemen ist meistens eine iterative Vorgehensweise notwendig, bei dem für Teile des Modellierungsprozesses verschiedene Varianten und Algorithmen ausprobiert und somit mehrfach durchlaufen werden.

Methoden zur datenbasierten Fuzzy-Modellierung

In den Anfängen der Fuzzy-Modellierung wurde zunächst hauptsächlich wissensbasiert vorgegangen, bevor sich in den letzten Jahren stark gesteigerte Forschungsaktivitäten im Bereich der datenbasierten Fuzzy-Modellierung entwickelten. Von vielen Wissenschaftlern konnte die Möglichkeit der datenbasierten Generierung von Fuzzy-Modellen zur Approximation, Klassifikation, Prognose, Reglersynthese und dynamischen Prozessmodellierung belegt werden. In diesem Abschnitt wird zunächst ein Überblick über die grundlegend verschie-

denen Modellierungsansätze gegeben. Dabei wird insbesondere auf die Interpretierbarkeit der generierten Modelle, die verwendete Regelstruktur und die handhabbare Dimensionalität des Suchraums eingegangen. Weiterhin wird der Einsatz evolutionärer Algorithmen im Bereich der Fuzzy-Modellierung thematisiert. Dabei werden teilweise einige Grundprinzipien und Begriffe der Fuzzy-Logik als bekannt vorausgesetzt, auch wenn diese inhaltlich erst in Abschnitt 3 systematisch eingeführt werden.

Die beiden gebräuchlichsten Typen von Fuzzy-Modellen sind das Mamdani-Fuzzy-Modell [MG81] und das Takagi-Sugeno-Fuzzy-Modell [TS85]. Der Hauptunterschied dieser Modelle liegt in der Konklusionen der zugehörigen Fuzzy-Regeln. Mamdani-Fuzzy-Regeln haben die Form

WENN „*Prämisse*“ DANN „*Linguistischer Term der Ausgangsvariablen*“.

Takagi-Sugeno-Fuzzy-Regeln dagegen erscheinen in der Form

WENN „*Prämisse*“ DANN „*Auswertung einer Funktion*“.

Durch die Verwendung von linguistischen Termen in der Konklusion ist die Transparenz einer Mamdani-Fuzzy-Regel gewährleistet, während die Verwendung einer Funktion in der Konklusion einer Takagi-Sugeno-Fuzzy-Regel die Nachbildung eines Systems mit geringerem Approximationsfehler ermöglicht, wobei aber die Transparenz deutlich eingeschränkt wird.

Wie bereits im vorherigen Abschnitt angedeutet, sind demnach bei der datenbasierten Erstellung eines Fuzzy-Modells bei der Modellbildung zusätzliche Anforderungen zu erfüllen:

- Definition der linguistischen Werte $\mathcal{L}_{i,j}$ für die zur Modellierung selektierten Merkmale M_i ,
- Definition der linguistischen Werte $\mathcal{L}_{Y,j}$ für die Ausgangsgröße Y ,
- Bestimmung der Konklusionen C_r bei Takagi-Sugeno-Fuzzy-Regeln,
- Optimierung des generierten Regelsatzes \mathcal{R} (optional nach der Bewertung).

Die Erfüllung dieser einzelnen Anforderungen wird bei den verschiedenen Modellierungsverfahren in der Regel entkoppelt von der eigentlichen Modellidentifikation durch Regelgenerierung durchgeführt. Ausschließlich bei den Modellierungsverfahren, die auf dem sogenannten *Pittsburgh-Ansatz* [Smi80] basieren, können diese Anforderung direkt bei der Regelgenerierung mit berücksichtigt

werden. Bei diesem Modellierungsansatz werden für die Identifikation von Regelsätzen alle möglichen Kombinationen von Regeln aufgestellt und hinsichtlich der Modellierungsgüte bewertet. Diese vollständige Durchmusterung aller möglichen Regelsätze garantiert, dass der bezüglich des Gütekriteriums beste Regelsatz gefunden wird. Aufgrund der kombinatorischen Explosion ist eine solche Vorgehensweise jedoch nur für einfachere Anwendungsprobleme mit geringer Dimensionalität durchführbar und wird daher in der Praxis bei hochdimensionalen Problemen kaum angewendet. Eine simultane Identifikation bzw. Optimierung von Zugehörigkeitsfunktionen verschärft die Situation zusätzlich.

Durch eine individuelle Bewertung jeder einzelnen Regel anstelle der globalen Bewertung eines Regelsatzes kann diese Problematik umgangen werden. Dies ist der wesentliche Kern von Modellierungsverfahren, die auf dem *Michigan-Ansatz* [HHN86] basieren. Das komplexe Problem, einen guten Regelsatz zu finden, wird dadurch auf das einfachere Problem, einzelne gute Regeln zu finden, zurückgeführt. Dies ist auch die grundlegende Idee des im Rahmen dieser Arbeit eingesetzten Fuzzy-ROSA-Verfahrens [KK95, Kro99, SKKK01], das in Abschnitt 3.3 genauer beschrieben wird. Der Regelsatz wird dabei sukzessive aus den gefundenen Regeln aufgebaut. Die Strukturierung des Suchraums durch Definition der Zugehörigkeitsfunktionen wird dabei praktisch offline vor der eigentlichen Modellidentifikation durch Regelgenerierung vorgenommen. In Abhängigkeit von der festgelegten Struktur des Suchraums führt auch diese Vorgehensweise mit zunehmender Komplexität zu einem unvermeidbar hohen Aufwand. Somit sind auch bei diesem Ansatz der kompletten Durchmusterung, also dem Bewerten aller möglichen Regeln, Grenzen gesetzt [KSK00].

In Anbetracht dessen wird für die Identifikation von Regeln und Regelsätzen in der Literatur vielfach der Einsatz evolutionärer Algorithmen [Sch95, BHS97, SHF94] vorgeschlagen, die oftmals neben der Generierung der Regeln bzw. Regelsätze auch für die Optimierung weiterer freier Parameter, wie insbesondere Lage und Form der Zugehörigkeitsfunktionen [Kar91, HM92, KG93, LT93, HLV93, KKK94, Sur94, KKS00a], verwendet werden. Im Fall der Generierung von Regelsätzen mit Hilfe evolutionärer Algorithmen wird häufig der Pittsburgh-Ansatz verfolgt, bei dem ein einzelnes Individuum der Population einen gesamten Regelsatz repräsentiert [HM92, KG93, KKK94, Wol94, Xio01]. Die grundsätzliche Aufgabe besteht dann darin, aus einer vorgegebenen Menge möglicher Regeln eine bezüglich der Modellierungsgüte optimale Teilmenge als Regelsatz auszuwählen. Für diese kombinatorische Problemstellung werden vorzugsweise genetische Algorithmen [Hol75, Gol89] verwendet. Allerdings führt diese Vorgehensweise mit einer steigenden Zahl von Eingangsgrößen schnell zur kombinatorischen Explosion der Größe des resultierenden Suchraums. Der kom-

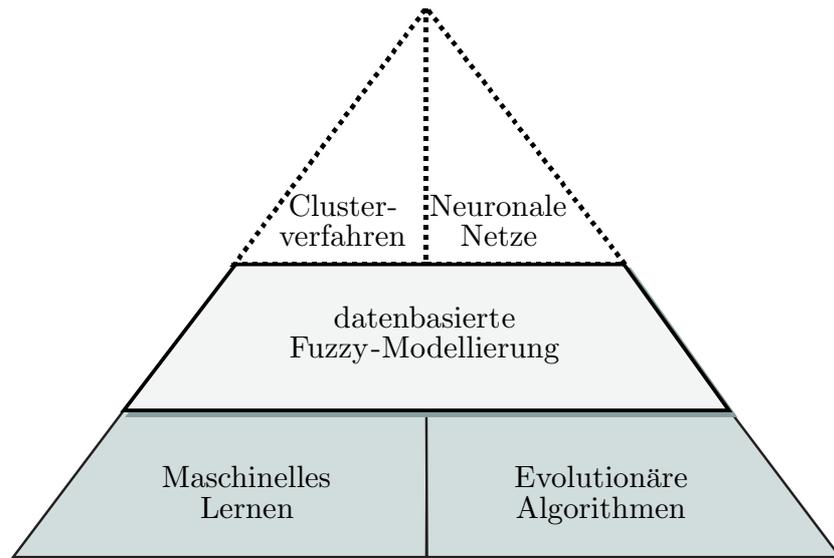


Abbildung 2.3: Zur datenbasierten Fuzzy-Modellierung genutzte etablierte Methoden.

binatorischen Explosion kann zumindest teilweise durch eine Einschränkung des Suchraums entgegengewirkt werden. Dies kann beispielsweise durch eine Begrenzung der Individuen (Regelsätze) auf eine maximale Größe erreicht werden. Noch viel effizienter dagegen ist eine Beschränkung der Suche auf einzelne relevante Regeln [KKS00b]. Dies entspricht einer Modellierung nach dem Michigan-Ansatz, wobei jede einzelne Regel von einem Individuum repräsentiert wird. Eine wesentliche Voraussetzung dafür ist, dass ein Bewertungsmaß für die Güte (Fitness) einzelner Regeln konzipiert wird. Die evolutionäre Regelsuche im Fuzzy-ROSA-Verfahren [KK97, SKH⁺99] beruht auf der grundsätzlichen Idee, dass während der Suche nach der im Sinne des Bewertungsmaßes besten Regel viele gute Regeln generiert werden, aus denen ein Regelsatz dann sukzessiv aufgebaut wird. Summa summarum bleibt deutlich herauszustellen, dass der Michigan-Ansatz auch für sehr komplexe Anwendungen geeignet ist [SPS⁺99] und somit der Mehraufwand beim Entwurf und die Tolerierung etwas geringerer Modellierungsgüten gerechtfertigt ist.

Ein alternativer Zugang zur datenbasierten Fuzzy-Modellierung bietet der Einsatz von *Clusterverfahren* [Bez81, JD88, HKK97, Zim95]. Diese *Instanzbasierten* Lernverfahren (*IBL*) [WM00] sind insbesondere sehr gut für eine Strukturierung des Suchraums und für eine erste grobe Modellierung geeignet. Die Strukturierung des Suchraums kann dabei in einer Merkmalsselektion [Bas96], in einer Festlegung der Zugehörigkeitsfunktionen [KS98] oder in einer Vorgabe der Anzahl der Regeln [PR96, RW96] bestehen. Sehr verbreitet ist auch eine direkte Modellierung durch Überführung des Ergebnisses der Clusteranalyse in Regeln [ACPF96, GSJ97, HC97, ON96, Chi96]. Da-

bei wird jedem Cluster eine Regel zugeordnet, deren Prämisse im Sinne von WENN „*Datenpunkt im Cluster*“ interpretiert werden kann. Bei dieser Sichtweise korrespondiert der Erfülltheitsgrad der Prämisse mit der Zugehörigkeit zum Cluster. Dies wird meist durch entsprechende eingangsseitige Zugehörigkeitsfunktionen individuell für jedes Cluster und damit individuell für jede Regel definiert. Nachteilig daran ist die mit Anzahl der Regeln stark ansteigende Zahl von Zugehörigkeitsfunktionen. Da diese sich zudem stark überlappen oder sogar vollständig überdecken können, sind die resultierenden Modelle kaum noch interpretierbar. Zur Ermittlung der Konklusionen der Fuzzy-Regeln werden verschiedene Verfahren zur Datenauswertung der einzelnen Cluster vorgeschlagen. Im einfachsten Fall erfolgt die Festlegung der Konklusion für Mamdani-Modelle analog zur Prämisse anhand der Cluster [GSJ97], und für Takagi-Sugeno-Kang-Modelle werden die freien Parameter der ausgangsseitigen Funktion mit einem klassischen Regressionsalgorithmus identifiziert [BDB97]. Insgesamt wird das Modellierungsergebnis sowie die handhabbare Komplexität bei der beschriebenen Vorgehensweise in starkem Maße vom verwendeten Clusterverfahren bestimmt. In hochdimensionalen Suchräumen komplexer Anwendungsprobleme können im Allgemeinen nur unbefriedigende Resultate erzielt werden. Neben dem Nachteil, dass bei der Mehrzahl der einsetzbaren Verfahren die Anzahl der Cluster und damit die Anzahl der Regeln vorgegeben werden muss, wirken sich zudem die starke Abhängigkeit des Ergebnisses von der Initialisierung des Clusterverfahrens sowie häufig auch von der Reihenfolge der Datenpunkte im Lerndatensatz negativ aus.

Mittlerweile finden zunehmend *künstliche neuronale Netze (KNN)* [Sar02] in der Fuzzy-Modellierung ihren Einsatz mit dem Ziel, die im Allgemeinen höhere Approximationsgüte ohne Verzicht auf die Interpretierbarkeit in sogenannten *Neuro-Fuzzy-Systemen* zu nutzen. Dazu werden in der Literatur vermehrt Verfahren vorgeschlagen, mit denen neuronale Netze in Fuzzy-Modelle überführt werden können [CW96, CERK96, CF97, IDC97, NK97, ISUT98, CL98, NK98, SKBB98]. Voraussetzung dafür sind spezielle Netzarchitekturen, bei denen im einfachsten Fall die Eingangsneuronen der Fuzzifizierung, die verborgenen Neuronen den Regeln und die Ausgangsneuronen der Inferenz entsprechen. Die Hauptschwierigkeit bei der Übertragung dieses Ansatzes auf komplexere Problemstellungen liegt darin, eine geeignete Netzstruktur zu finden, die sowohl eine hohe Approximationsgüte aufweist als auch zu interpretierbaren Ergebnissen führt. Eine weitere Einsatzmöglichkeit von neuronalen Netzen ist die Optimierung bereits bestehender Fuzzy-Modelle [JSS98, SM98, PR96].

Des Weiteren haben sich für die datenbasierte Fuzzy-Modellierung baumorientierte Verfahren [JMMB00] aus dem Bereich des *Maschinellen Lernens* etabliert

[Qui93, Wys97]. Die Verfahren zur Fuzzy-Regelgenerierung beruhen dabei meist auf dem von QUINLAN entwickelten ID3-Algorithmus zur Induktion von *Entscheidungsbäumen* [Qui93]. Entscheidungsbäume sind aus Knoten und Blättern bestehende Graphen. Ausgehend von einem sogenannten *Wurzelknoten* werden die Knoten des Graphen bis zu einem Blatt, d.h. einem Knoten ohne weitere Nachfolgeknoten, durchlaufen. In jedem Knoten wird dabei aufgrund einer Bedingung über die Auswahl des Nachfolgeknotens entschieden. Eine einfache Bedingung kann beispielsweise die Abfrage sein, ob eine bestimmte Eingangsgröße einen bestimmten Wert überschreitet. Ein Blatt schließlich enthält das Entscheidungsergebnis. Entscheidungsbäume sind jederzeit in einen Regelsatz aus Fuzzy-Regeln überführbar, indem das Entscheidungsergebnis eines Blattes als Konklusion im DANN-Teil der zugehörigen Regel gesetzt wird. Die Bedingungen, die auf dem graphischen Weg vom Wurzelknoten über die verschiedenen Nachfolgeknoten zu diesem betrachteten Blatt zu erfüllen sind, werden dann konjunktiv verknüpft als Prämisse im WENN-Teil der Regel übernommen. Der Effizienz dieser baumorientierten Regelinduktionsverfahren stehen verschiedene Nachteile gegenüber. Ähnlich wie bei vielen Clusterverfahren ist das Ergebnis in Form des resultierenden Entscheidungsbaums abhängig von der Reihenfolge der Datenpunkte im Lerndatensatz. Zusätzlich neigen die Verfahren zu einer starken Überanpassung an die Lerndaten (*Overfitting*) und induzieren häufig schwer interpretierbare Regeln von hohem Spezialisierungsgrad.

Die bisher vorgestellten Ansätze belegen eindrucksvoll, dass eine datenbasierte Generierung von Fuzzy-Modellen prinzipiell sehr gut möglich und Gegenstand vielfältiger Forschungsaktivitäten ist. In neueren Arbeiten sind im Bereich der Modellierungsverfahren zunehmende Bestrebungen festzustellen, die Modellierungsgüte weiter zu erhöhen, dabei möglichst ein *Overfitting* zu vermeiden und möglichst einfach interpretierbare Fuzzy-Modelle zu erhalten. Insbesondere die Interpretierbarkeit ist im Hinblick auf die Akzeptanz in industriellen Anwendungsfeldern ein außerordentlich wichtiger Aspekt, der nicht von allen Autoren angemessen berücksichtigt wird [KKN97, NK98]. Eine wesentliche Voraussetzung für die Nachvollziehbarkeit von Regeln ist, dass sie lokal vernünftig sind, d. h. einen tatsächlichen Einzelaspekt des Gesamtverhaltens beschreiben [INM96, CHL96, JS98, KT98, Kro99]. Im Fuzzy-ROSA-Verfahren wird dies dadurch gewährleistet, dass einzelne Regeln bezüglich ihrer Relevanz getestet werden (Abschnitt 3.3). Wie in [YWG98] gezeigt wird, kann bei Modellierungsansätzen, die ausschließlich den Modellierungsfehler minimieren, dies nicht garantiert werden. Auch bei neuronalen Netzen kann die Forderung nach lokal vernünftigen Regeln meist nur schwer berücksichtigt werden.

Hinsichtlich der Interpretierbarkeit der erzeugten Modelle ist zudem auf die Anzahl der generierten Regeln im Sinne der Übersichtlichkeit zu achten. Der Einsatz einer nachgelagerten Regelreduktion kann in diesem Zusammenhang eine Möglichkeit zur Erhöhung der Interpretierbarkeit darstellen [Kro96, KKS00b]. Zusätzlich ist bezüglich der Zugehörigkeitsfunktionen anzumerken, dass die Interpretierbarkeit der Fuzzy-Modelle steigt, wenn eine sogenannte *Standard-Fuzzy-Partitionierung* (Definition 3.5) oder zumindest eine geordnete Folge nicht zu stark überlappender Zugehörigkeitsfunktionen verwendet wird. Gerade bei Verfahren, die für jede Regel eigene Zugehörigkeitsfunktionen erzeugen, ist dies häufig nicht Fall. Teilweise kann sogar von einer Verlagerung der Komplexität von den Regeln hin zu den Zugehörigkeitsfunktionen gesprochen werden. Hieraus motiviert, rücken neuerdings zunehmend Methoden zur datenbasierten Merkmalsselektion (*feature selection*) zur Komplexitätsreduktion und Entlastung der Modellierungsverfahren in das Interesse der Forschungsaktivitäten. Im Hinblick auf die zur Fuzzy-Modellierung erforderlichen Zugehörigkeitsfunktionen sind dabei in der Literatur nur wenige Ansätze zur Berücksichtigung oder gar Optimierung einer Partitionierung des Suchraumes bekannt. Eine Einführung in diese bedeutsamen Aspekte sowie ein Überblick über die ersten bekannten Ansätze mit Literaturangaben wird in den folgenden beiden Unterkapiteln gegeben.

2.1.3 Merkmalsselektion zur Komplexitätsreduktion und Begegnung der COD-Problematik

Der Begriff der *Curse of Dimensionality*³ (COD) wurde 1961 von BELLMANN bei seiner Aufgabe, Wahrscheinlichkeitsdichten in hochdimensionalen Räumen anhand einer gezogenen Stichprobe zu schätzen, geprägt [Bel61]. Diese Aufgabe kann derart präzisiert werden, dass die Wahrscheinlichkeitsdichte in jeder Zelle eines mehrdimensionalen Gitternetzes zu bestimmen ist. Ist die Anzahl an Gitterlinien pro betrachteter Dimension fest vorgegeben, so steigt die Anzahl der Zellen exponentiell mit der Anzahl der Dimensionen an. Damit steigt ebenfalls die Größe der Stichprobe, die für eine Schätzung der gesuchten Wahrscheinlichkeitsdichte unter Berücksichtigung einer bestimmten statistischen Sicherheit benötigt wird.

Mittlerweile wird der Begriff der COD-Problematik für alle Probleme verwendet, die bei einem Algorithmus mit zunehmender Anzahl von zu berücksichtigenden Einflussgrößen auftreten können. Für sämtliche Methoden zur datenbasierten Modellierung ergeben sich daraus teilweise leicht unterschiedliche Aus-

³Fluch der Dimension oder Leere der hochdimensionalen Räume

wirkungen bezüglich Laufzeit, Speicherplatzbedarf und Qualität der Ergebnisse in Abhängigkeit von der Größe der verwendeten Lerndatenstichprobe.

So treten beispielsweise bei den Instanzbasierten Lernverfahren (*IBL*) Probleme mit der verwendeten Abstandsfunktion bei der Clusterbildung dahingehend auf, dass sich die ermittelten Abstandswerte für den nächsten und den fernsten Nachbarn einer betrachteten Eingangsraumposition mit zunehmender Anzahl der Dimension des Raumes immer mehr einander angleichen und damit letztlich bedeutungslos werden [HAK00].

Auch bei Modellierungsverfahren, die auf dem *Pittsburgh-Ansatz* oder *Michigan-Ansatz* basieren, wirkt sich in der Regel mit steigender Dimensionalität der Modellierungsprobleme die daraus resultierende Komplexität dahingehend störend aus, dass diese Verfahren mit akzeptablem Rechenaufwand keine handhabbaren und nachvollziehbaren Regelsätze generieren können. Das zeigt sich schon daran, dass viele veröffentlichte Ergebnisse über Weiterentwicklungen von Methoden zur datenbasierten Modellierung ihre Leistungsfähigkeit nur auf Anwendungsbeispiele mit höchstens drei Eingangsgrößen beziehen [ACPF96, BSKNL96, Bas96, CW96, Fag96, NFM96, ON96, PR96, RW96, ZGW96, Ben97, CF97, CH97, HC97, IDC97, MT97, GSJ97, TH97, CL98, JSS98, SKBB98, SM98, YWG98].

Nur wenige Autoren präsentieren Beispiele mit vier Eingangsgrößen [CFM96, Chi96, HL96, RY96, NGF96, BDB97, HC98, CHHL98, NK98] und kaum noch welche mit fünf bis zehn Eingangsgrößen [KFR95, CERK96, KS96, KK97, Hol98, ISUT98].

Abgesehen von Arbeiten, die auf dem Fuzzy-ROSA-Verfahren [KK97, SPS⁺99, SKH⁺99, SKK⁺99a, SKK99b, KKS00b, KKS00a, KSK00, SKKK01] basieren, wie sie im Rahmen dieser Arbeit zum größten Teil als Referenz verwendet werden (vergleiche Kapitel 6), finden sich für höherdimensionale Eingangsgrößenräume nur noch sehr wenige Arbeiten [INM96, KK97, Hol98]. Die großen Erfolge des Fuzzy-ROSA-Verfahrens zur effizienten Generierung relevanter Fuzzy-Regeln in hochdimensionalen Suchräumen ist ein Ergebnis konsequenter Forschungsarbeiten und Weiterentwicklungen von SLAWINSKI [Sla01], der unter anderem eine Reduzierung des erforderlichen Rechenaufwands durch sicheres Beschneiden des Suchraums erzielte.

Unabhängig vom gewählten Modellierungsverfahren sind erfahrungsgemäß nur selten alle gemessenen oder zur Verfügung stehenden Einflussgrößen des zu modellierenden Systems notwendig, um die zugrunde liegenden Prozesszusammenhänge mit ausreichender Genauigkeit zu modellieren. Eine dem Modellierungsprozess vorgelagerte Selektion von relevanten und nichtredundanten Ein-

flussgrößen wirkt sich aus mehreren Gründen oftmals vorteilhaft aus:

- Der Arbeitsaufwand zum Messen und zum Speichern der Daten wird reduziert.
- Der Rechenaufwand für die Modellidentifikation wird reduziert, da die Laufzeit vieler Modellierungsverfahren mit zunehmender Anzahl berücksichtigter Einflussgrößen stark ansteigt.
- Das generierte Modell ist häufig leichter zu interpretieren.
- Eine Elimination irrelevanter oder redundanter Einflussgrößen erhöht meistens die erzielbare Generalisierungsfähigkeit der generierten Modelle deutlich.

2.1.4 Einfluss der Fuzzy-Partitionierung des Merkmalsraumes

Zum einen beeinflusst die *Anzahl* der definierten linguistischen Werte für die Fuzzifizierung der eingangsseitigen Merkmale sowie der zu modellierenden Ausgangsgröße die Anzahl möglicher Regeln und damit die Komplexität des Suchraums, da hierdurch die Feinheit der sogenannten *Granularisierung* des Merkmalsraumes induziert wird. Hierdurch wird die Leistungsfähigkeit des gewählten Verfahrens zur datenbasierten Regelgenerierung stark tangiert. Zum anderen bestimmt die *Lage* der Zugehörigkeitsfunktionen die Partitionierung des Merkmalsraumes, also das jeweils adressierte Gebiet einer möglichen Regel. Wie in [Zhe92, BKC96, CHV00] gezeigt, ist der Einfluss dieser Fuzzy-Partitionierung auf die damit erzielbare Modellierungsgüte viel größer als der Einfluss des Zusammenspiels der generierten Regeln, welcher in unmittelbarem Zusammenhang mit dem gewählten Verfahren zur Regelgenerierung steht. Dies wird an einem einfachen zweidimensionalen fiktivem Klassifikationsbeispiel in Abbildung 2.4 verdeutlicht. Zur vereinfachten Illustration werden dabei scharfe Partitionierungen des zweidimensionalen Merkmalsraumes betrachtet. Ferner wird angenommen, dass die Ausgangsgröße genau zwei Klassenwerte annehmen kann, die durch die Symbole “o” und “+” dargestellt werden.

Die beiden einfachsten heuristisch motivierten und am häufigsten eingesetzten Verfahren sind die sogenannte *äquidistante* und die *äquifrequente* Partitionierung. Beide Verfahren arbeiten univariat und unüberwacht, d. h. die Einteilung des Raumes wird ohne Berücksichtigung der Werte einer anderen Einflussgröße oder der Werte der Ausgangsgröße vorgenommen. Für beide Verfahren wird die Vorgabe einer festen Anzahl S_c von Partitionierungsintervallen benötigt.

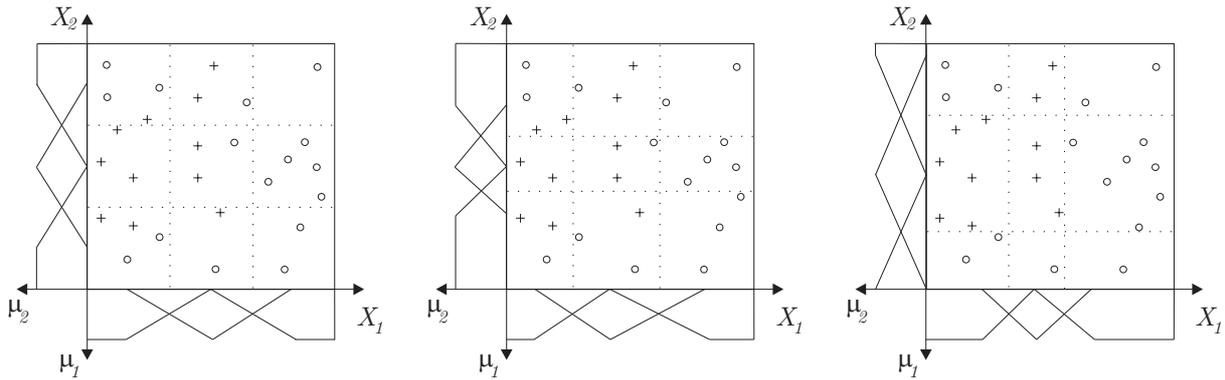


Abbildung 2.4: Äquidistante (links), äquifrequente (mitte) und optimal angepasste (rechts) Partitionierung des Merkmalsraumes.

- *Äquidistante Partitionierung*: Der Wertebereich einer Größe X wird unter Verwendung der anhand der Lernstichprobe ermittelten Spannweite X_{range} äquidistant in S_c Intervalle der Breite

$$w = \frac{X_{range}}{S_c} \quad (2.1)$$

eingeteilt. Zur Unterdrückung des Einflusses statistischer Ausreißer ist es ratsam, bei Ermittlung der Spannweite einen geringen Prozentsatz der kleinsten und größten Werte zu ignorieren, um ein stabileres Ergebnis zu erhalten.

- *Äquifrequente Partitionierung*: Als Nachteil der gleichmäßigen Aufteilung des Wertebereichs einer Größe erweist sich häufig die starke Streuung der Anzahl der Datenpunkte in den Diskretisierungsintervallen. Die äquifrequente Diskretisierung umgeht dieses Problem, indem die Intervallgrenzen so positioniert werden, dass sich die N Datenpunkte der Lernstichprobe zu gleichen Teilen

$$N_{S_c} = \text{floor} \left(\frac{N}{S_c} \right) \quad (2.2)$$

auf die S_c Intervalle verteilen. Der Operator $\text{floor}(x)$ liefert dabei den ganzzahligen Anteil einer Gleitkommazahl.

Beide Verfahren liefern eine scharfe Partitionierung, welche nachträglich, wie in Abbildung 2.4 an den Achsen angedeutet, leicht in eine *Standard-Fuzzy-Partitionierung* (Definition 3.5) umgewandelt werden kann (siehe Kapitel 4.9). In Abbildung 2.4 ist direkt erkennbar, dass eine Anwendung dieser Verfahren für

das fiktive Beispiel Partitionierungen liefert, bei denen *nicht* in allen resultierenden Zellen des Raumes ausschließlich Datenpunkte mit gleicher Ausgangsklasse zu finden sind. Unter Verwendung dieser Partitionierung kann daher mit keinem Verfahren zur automatischen Regelgenerierung ein vollständig exakt klassifizierendes Fuzzy-Modul generiert werden. Nur durch eine Berücksichtigung der Ausgangsgröße kann eine den Daten optimal angepasste Partitionierung gefunden werden.

Hieraus motiviert, wird in [KS98] ein Streifen-Cluster-Konzept vorgestellt, welches die Intervalleinteilung des Wertebereichs einer Einflussgröße X_i anhand einer achsparallelen streifenförmigen Clusterung der Datenpunkte im zweidimensionalen Raum $X_i \times Y$ ermittelt. Neben dem Einsatz von Clusterverfahren wird in der Literatur auch der Einsatz von Optimierungsverfahren vorgeschlagen, mit denen bei festgelegter Diskretisierung der Ausgangsgröße Y die Intervalleinteilung des Wertebereichs einer Einflussgröße X_i durch Maximierung der *Transinformation* $\mathcal{I}(X, Y)$ [JMMB00] ermittelt wird. Ferner sind im hier vorliegenden Kontext noch ähnliche Ansätze bekannt, die die intervallweise gemittelte *Entropie* $\mathcal{E}(X, Y)$ der Ausgangsgröße Y rekursiv minimieren [DKS95, HLTM99].

Der wesentliche Nachteil aller bekannten und hier vorgestellten Ansätze besteht darin, dass prinzipiell keine Wirkungszusammenhänge mehrerer Einflussgrößen X_i, X_j, \dots, X_k sowohl untereinander (*Redundanzanalyse*) als auch mit der Ausgangsgröße Y (*Relevanzanalyse*) festgestellt werden können.

2.2 Ziele dieser Arbeit

Generelles Ziel dieser Arbeit ist es, die Anwendungsgrenzen der Methoden zur datenbasierten Fuzzy-Modellierung in Richtung höherer Komplexität auszuweiten, wobei ein Verlust an erzielbarer Modellierungsgüte möglichst vermieden werden soll. Dieses Ziel soll insbesondere durch eine geeignete Beschneidung und Vorstrukturierung des Suchraums erzielt werden. Hierzu sind neue, speziell auf die datenbasierte Fuzzy-Modellierung ausgerichtete Verfahren zur Merkmalsselektion zu konzipieren, die im Rahmen einer Datenvoranalyse aus einer gegebenen Menge von potenziellen Einflussgrößen einen relevante Merkmalsatz selektieren. Besonderer Wert soll dabei im Hinblick auf eine größtmögliche Transparenz, Interpretierbarkeit und Generalisierungsfähigkeit der generierten Fuzzy-Modelle auf Modellierungsverfahren gelegt werden, die auf dem *Michigan-Ansatz* basieren. Durch die individuelle Bewertung jeder einzelnen Regel gewährleisten diese implizit, dass lokal vernünftige und damit interpretierbare Regeln erhalten werden. Für die Regelgenerierung müssen diesen

Verfahren die zu berücksichtigenden Einflussgrößen mit zugehörigen Fuzzy-Partitionierungen bereitgestellt werden. Eine Gewichtung der Berücksichtigung von Einflussgrößen oder einer Adaption der Partitionierungen während des Regelgenerierungsprozesses ist dabei prinzipiell nicht möglich. Als Repräsentant dieser Klasse von Modellierungsverfahren wird für die Regelgenerierung im Rahmen dieser Arbeit das von KIENDL, KRABS, KRONE und SLAWINSKI entwickelte Fuzzy-ROSA-Verfahren [KK89, Kra94, KK95, Kro99, Sla01] verwendet, bei dem einzelne Regeln daraufhin statistisch getestet werden, ob sie einen relevanten Teilaspekt des betrachteten Systems widerspiegeln.

Zum Einlösen des formulierten generellen Ziels werden in dieser Arbeit folgende Teilziele verfolgt:

Relevanzanalyse/Redundanzanalyse potenzieller Einflussgrößen

Das primäre Teilziel dieser Arbeit ist es, aus einer gegebenen Menge von potenziellen Einflussgrößen automatisiert einen relevanten Merkmalsatz zu selektieren. Dieser Merkmalsatz soll dadurch gekennzeichnet sein, dass ausschließlich solche Einflussgrößen darin enthalten sind, die entweder direkt oder im Zusammenspiel mit weiteren Einflussgrößen in engem Zusammenhang mit der Ausgangsgröße stehen. Letzteres erfordert die Möglichkeit zur Relevanzbewertung gesamter Sätze von Einflussgrößen (*potenzielle Merkmalsätze*), was die meisten etablierten Selektionsverfahren nicht ermöglichen. Zudem erlaubt der Ansatz zur gesamtheitlichen Relevanzbewertung mehrerer Einflussgrößen, Redundanzen innerhalb des Merkmalsatzes aufzuspüren und zu eliminieren. Dies ist ebenfalls von sehr großer Bedeutung, da redundante Einflussgrößen bei der Modellgenerierung zu einer Vervielfachung der Regeln führen. Diese Vervielfachung führt nicht zu einer Erhöhung der Modellgenauigkeit, wohl aber zu einem deutlichen Verlust an Transparenz und Interpretierbarkeit der generierten Modelle. Um die Handhabung möglichst komplexer Modellierungsprobleme zu ermöglichen, ist aus rechentechnischen Gesichtspunkten eine Entkoppelung der Verfahren zur Merkmalsselektion vom Modellierungsverfahren erforderlich. Die dazu zu entwickelten Verfahren sollen alle Typen von Modellierungsproblemen sowie auch alle möglichen Typen von Systemgrößen behandeln können.

Berücksichtigung/Optimierung der Partitionierung des Merkmalsraumes:

Wie vorab in 2.1.4 illustriert, ist für eine datenbasierte Generierung von Fuzzy-Modellen mit hoher Modellierungsgüte die Bereitstellung einer den Datenstrukturen gut angepassten Fuzzy-Partitionierung des Merkmalsraumes erforderlich. Daraus ergibt sich das zweite Teilziel dieser Arbeit:

die Berücksichtigung einer vorgegebene Partitionierung eines potenziellen Merkmalsatzes bei der Relevanzbewertung. Darüber hinaus sollen Verfahren zur Variation der Partitionierung im Sinne einer Abschätzung der maximal möglichen Relevanz eines potenziellen Merkmalsatzes entwickelt werden. Erst hierdurch wird die Möglichkeit zu einer einheitlichen vergleichenden Beurteilung verschiedener Merkmalsätze mit individuell partitionierten Merkmalsräumen geschaffen. Die Variation der Partitionierungen zur Abschätzung der maximal möglichen Relevanz soll als Optimierungsproblem definiert werden. Aufgrund des starken nichtlinearen Einflusses der Partitionierung soll hierzu eine Evolutionsstrategie eingesetzt werden.

Effizienzsteigerung von Evolutionsstrategien:

Die optimale Anpassung der Partitionierung eines Merkmalsraumes an die darin enthaltenen Datenpunkte ist ein äußerst komplexes Optimierungsproblem und stellt aufgrund der zu erwartenden Multimodalität der Bewertungsfunktion höchste Anforderungen an das eingesetzte Optimierungsverfahren. Hieraus ergibt sich das dritte Teilziel dieser Arbeit: die Entwicklung von Mechanismen zur Effizienzsteigerung von Evolutionsstrategien. Diese sollen ein sicheres und schnell konvergierendes Auffinden *robuster Lösungen* in hochgradig multimodalen Gütelandschaften ermöglichen. Die *Robustheit* einer gefundenen Lösung wird dabei durch die detektierbare Größe desjenigen umliegenden Gebietes gekennzeichnet, welches ausschließlich Punkte mit homogen verteilten geringfügig schlechteren Gütewerten beinhaltet.

Kapitel 3

Methodische Grundlagen

In diesem Kapitel wird auf die wesentlichen methodischen Grundlagen eingegangen, die inhaltlich für diese Arbeit einen zentralen Stellenwert einnehmen.

Für die Grundlagen zur datenbasierten Fuzzy-Modellierung wird zunächst dabei in Abschnitt 3.1 eine kurze Einführung in die von ZADEH im Jahre 1965 eingeführte Fuzzy-Logik gegeben. Die darauf basierenden gängigsten Typen von Fuzzy-Systeme werden in Abschnitt 3.2 beschrieben. Diese Ausführungen sind dabei an [Kie97a] angelehnt, da für die zentralen Aspekte dieser Arbeit eine anwendungsorientierte Betrachtungsweise von größerer Bedeutung ist. Die verwendeten Bezeichnungen richten sich zudem streng nach den in [MBC⁺99] aufgestellten VDI/VDE-Richtlinien 3550 „*Fuzzy-Logik und Fuzzy-Control — Begriffe und Definitionen*“. Schließend wird in Abschnitt 3.3 das hier verwendete Fuzzy-ROSA-Verfahren zur datenbasierten Generierung von Fuzzy-Systemen kurz vorgestellt sowie die detektierten Ansatzpunkte zur Effizienzsteigerung der datenbasierten Fuzzy-Modellierung durch die im Rahmen dieser Arbeit zu entwickelnden Methoden herausgearbeitet.

Anschließend werden die Grundlagen für diese neuen methodischen Ansätze für den Bereich der Merkmalsselektion in Abschnitt 3.4 und für das Gebiet der Evolutionsstrategien in Abschnitt 3.5 eingeführt.

3.1 Fuzzy-Logik

Die in [Zad65, Zad68] eingeführte Fuzzy-Logik basiert auf der Verwendung von Fuzzy-Mengen, die aus geordneten Wertepaaren

$$\mathcal{F} = \{x, \mu_{\mathcal{F}}(x)\} \quad \text{mit } x \in \mathcal{U} \quad \text{und } \mu_{\mathcal{F}}(x) \in [0, 1] \quad (3.1)$$

bestehen. Dabei repräsentiert x ein Element des betrachteten Universums \mathcal{U} , dessen Zugehörigkeit zur (unscharfen) Fuzzy-Menge \mathcal{F} durch den Zugehörigkeitsgrad $\mu_{\mathcal{F}}(x)$ spezifiziert wird. Die in dieser Arbeit betrachteten Ereignisse x

beinhalten immer die numerischen (scharfen) Werte x einer Eingangsvariablen X , so dass das Universum \mathcal{U} (auch *Grundbereich* genannt) mit der Menge \mathbb{R} der reellen Zahlen gleichgesetzt werden kann. Im Gegensatz zur booleschen Logik, bei der ein Element x ausschließlich mit den Wahrheitswerten $\mu_{bool}(x) = 1$ (*wahr*) oder $\mu_{bool}(x) = 0$ (*falsch*) zu einer Menge zugehörig sein kann, verwendet die Fuzzy-Logik somit auch Zwischenwerte aus dem Intervall $[0, 1]$ zur Spezifizierung einer (unscharfen) Zugehörigkeit. Die wesentlichen Kenngrößen und Eigenschaften von Fuzzy-Mengen und Zugehörigkeitsfunktionen sind im Anhang A aufgeführt.

Zur Ermöglichung logischer Verknüpfungen von Fuzzy-Mengen wurden die aus der Booleschen Logik bekannten Operatoren *UND*, *ODER* und *NICHT* durch verschiedene arithmetische Beziehungen zu Fuzzy-Operatoren so verallgemeinert, dass für den Fall von Booleschen Operanden das Ergebnis der Verknüpfung dem der Verwendung der klassischen Booleschen Operatoren entspricht. Es gibt unterschiedliche Möglichkeiten für die Definition von Fuzzy-Operatoren die dieser Forderung entsprechen. Die gebräuchlichsten Fuzzy-Operatoren sind in Tabelle 3.1 zu finden.

Tabelle 3.1: Arithmetische Realisierungen von Fuzzy-Operatoren

Symbol	Operator	Definition	Bezeichnung
\neg	Fuzzy-NICHT	$\neg\mu_{\mathcal{F}} := 1 - \mu_{\mathcal{F}}$	
\wedge	Fuzzy-UND	$\mu_{\mathcal{F}_1} \wedge \mu_{\mathcal{F}_2} := \mu_{\mathcal{F}_1} \cdot \mu_{\mathcal{F}_2}$	Algebraisches Produkt
		$\mu_{\mathcal{F}_1} \wedge \mu_{\mathcal{F}_2} := \min\{\mu_{\mathcal{F}_1}, \mu_{\mathcal{F}_2}\}$	Minimum
		$\mu_{\mathcal{F}_1} \wedge \mu_{\mathcal{F}_2} := \max\{0, \mu_{\mathcal{F}_1} + \mu_{\mathcal{F}_2} - 1\}$	Beschränkte Differenz
\vee	Fuzzy-ODER	$\mu_{\mathcal{F}_1} \vee \mu_{\mathcal{F}_2} := \mu_{\mathcal{F}_1} + \mu_{\mathcal{F}_2} - \mu_{\mathcal{F}_1} \cdot \mu_{\mathcal{F}_2}$	Algebraische Summe
		$\mu_{\mathcal{F}_1} \vee \mu_{\mathcal{F}_2} := \max\{\mu_{\mathcal{F}_1}, \mu_{\mathcal{F}_2}\}$	Maximum
		$\mu_{\mathcal{F}_1} \vee \mu_{\mathcal{F}_2} := \min\{1, \mu_{\mathcal{F}_1} + \mu_{\mathcal{F}_2}\}$	Beschränkte Summe

3.2 Fuzzy-Systeme

Für die Beschreibung der Struktur und der Funktionsweise von Fuzzy-Systemen bezeichnen im Folgenden x_1, \dots, x_n die numerischen (scharfen) Werte der n Eingangsvariablen X_1, \dots, X_n des Fuzzy-Systems zu einem festen Zeitpunkt, wodurch genau eine bestimmte Eingangssituation spezifiziert ist. Entsprechend ist der Ausgangsvariablen Y der numerische Wert y zugeordnet.

Fuzzy-Systeme basieren auf dem in Abschnitt 3.1 eingeführten Konzept der Fuzzy-Mengen und beschreiben im mathematischen Sinne einen funktionalen

Zusammenhang

$$f : \mathcal{X} \mapsto Y, \quad \text{mit } \mathcal{X} = X_1 \times \cdots \times X_n. \quad (3.2)$$

Definition 3.1 *Fuzzy-System:* Ein Fuzzy-System ist ein System, das linguistische Regeln verwendet und mit Hilfe der Teilblöcke Fuzzifizierung, Inferenz und Defuzzifizierung numerische Eingangsgrößen in numerische Ausgangsgrößen abbildet.

Im Teilblock der *Fuzzifizierung* werden dabei zunächst die Eingangsinformationen x_1, \dots, x_n auf Fuzzy-Mengen abgebildet. Diese werden einer Abbildungsvorschrift, die durch linguistische Regeln der Form WENN *Prämisse* DANN *Konklusion* repräsentiert wird, unterzogen (*Inferenz*). Als Inferenzergebnis resultiert wiederum eine Fuzzy-Menge, die letztendlich auf eine reellwertige Größe abgebildet wird (*Defuzzifizierung*).

Diese verschiedenen Teilbereiche eines Fuzzy-Systems sowie die wesentlichsten Definitionen, die zur Berechnung des Ausgangsgrößenwerts notwendig sind, werden in Abschnitt 3.2.1 exemplarisch an einem Fuzzy-System vom Typ Mamdani erläutert, der im Bereich Fuzzy Control und Fuzzy-Modellierung am weitesten verbreitet ist. Ein daraus weiter entwickelter funktionaler Typ — das Takagi-Sugeno-Fuzzy-System — mit dem in der Regel unter starker Beeinträchtigung der Transparenz bei Approximationsproblemen höhere Modellgüten erzielt werden können, wird im anschließenden Abschnitt 3.2.2 kurz vorgestellt.

3.2.1 Mamdani-Fuzzy-Systeme

Die wesentlichen Module eines Fuzzy-Systems vom Typ MAMDANI [Mam77, MA75] sind zur nachfolgenden Erläuterung in Abbildung 3.1 dargestellt.

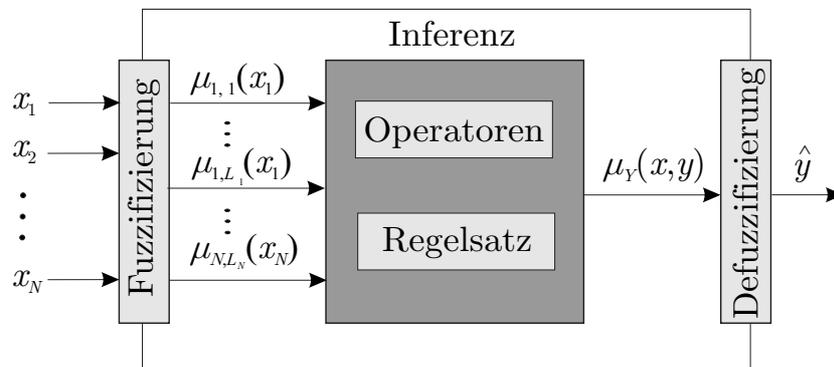


Abbildung 3.1: Vereinfachte modulare Struktur eines Mamdani-Fuzzy-Systems.

Fuzzifizierung

Die Eingangsgrößen eines Fuzzy-Systems werden *linguistische Variablen* genannt. Deren zu Grunde liegender Wertebereich wird im Modul der Fuzzifizierung einer unscharfen Granularisierung, der *Fuzzy-Partitionierung*, unterzogen.

Definition 3.2 *Fuzzy-Partitionierung*: Eine Fuzzy-Partitionierung ist eine Teilung eines (mehrdimensionalen) Raumes in unscharfe Teilräume, wodurch jeder Punkt allen Teilräumen zu einem bestimmten Zugehörigkeitsgrad zugeordnet wird.

Auf Grundlage von L_i Fuzzy-Partitionen für eine Eingangsvariable X_i werden die zugehörigen reellwertigen scharfen Werte x_i auf *linguistische Werte* $\mathcal{L}_{i,j}$ ($j = 1, 2, \dots, L_i$) abgebildet.

Definition 3.3 *Linguistischer Wert/Term*: Ein Linguistischer Wert/Term ist eine natursprachliche Bezeichnung zur Kennzeichnung der Eigenschaften einer Größe. Die gegenseitige Zuordnung zwischen numerischen und linguistischen Werten einer Größe erfolgt über Fuzzy-Mengen.

Die linguistischen Werte ermöglichen die Verarbeitung unscharfer qualitativer Aussagen der Form „Patient hat hohes Fieber“ mit Hilfe der Fuzzy-Logik (Abschnitt 3.1). Die Zuordnung eines numerischen Wertes x_i von einer Eingangsvariablen X_i (z. B. *Temperatur = 41.3 °C*) zu einem linguistischen Wert $\mathcal{L}_{i,5}$ gemäß Abbildung 3.3 *Temperatur = „sehr hoch“*, erfolgt über *Fuzzy-Zugehörigkeitsfunktionen* $\mu_{i,j}(x_i)$ mit $j = 1, 2, \dots, L_i$.

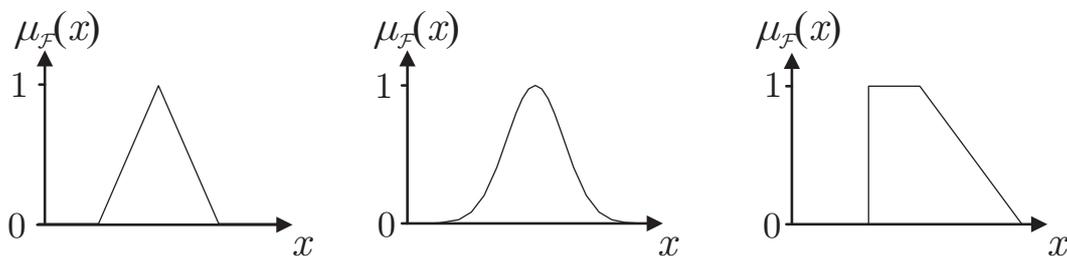


Abbildung 3.2: Gebräuchliche Formen für Zugehörigkeitsfunktionen $\mu_{\mathcal{F}}(x)$

Definition 3.4 *Fuzzy-Zugehörigkeitsfunktion*: Eine Fuzzy-Zugehörigkeitsfunktion ist Funktion, die jedem Element x aus dem im allgemeinen numerischen Grundbereich — dem Definitionsbereich — einen Zugehörigkeitsgrad $\mu_{\mathcal{F}}(x)$ zu einer Fuzzy-Menge \mathcal{F} nach Gleichung (3.1) zuordnet. Ein Zugehörigkeitsgrad

$\mu_{\mathcal{F}}(x) = 1$ bedeutet, dass der Wert x vollständig zur Fuzzy-Menge \mathcal{F} gehört. Bei $\mu_{\mathcal{F}}(x) = 0$ gehört x überhaupt nicht zur Fuzzy-Menge \mathcal{F} . Werte zwischen Null und Eins bedeuten, dass der Wert teilweise zur Fuzzy-Menge \mathcal{F} gehört.

Die gebräuchlichsten Zugehörigkeitsfunktionen für linguistische Werte sind dreieckförmig (triangular), trapezförmig, rechteckförmig oder gaußförmig (Abbildung 3.2). Häufig werden die Zugehörigkeitsfunktionen für jede Eingangsvariable X_i derart festgelegt, dass für jeden Wert x_i des zugehörigen Definitionsbereichs die Summe aller Zugehörigkeiten $\mu_{i,j}(x_i)$ Eins ergibt:

$$\sum_{j=1}^{L_i} \mu_{i,j}(x_i) = 1. \quad (3.3)$$

Definition 3.5 *Standard-Fuzzy-Partitionierung:* Gilt für alle Werte x des Definitionsbereichs einer linguistischen Variablen X Gleichung (3.3) und sind weiterhin alle Zugehörigkeitsfunktionen der linguistischen Werte normalisiert und konvex (Anhang A), dann wird diese Einteilung Standard-Fuzzy-Partitionierung oder auch Fuzzy-Informationssystem genannt.

Ein Beispiel für eine Standard-Fuzzy-Partitionierung ist in Abbildung 3.3 für die Modellierung der Körpertemperatur eines Menschen dargestellt.

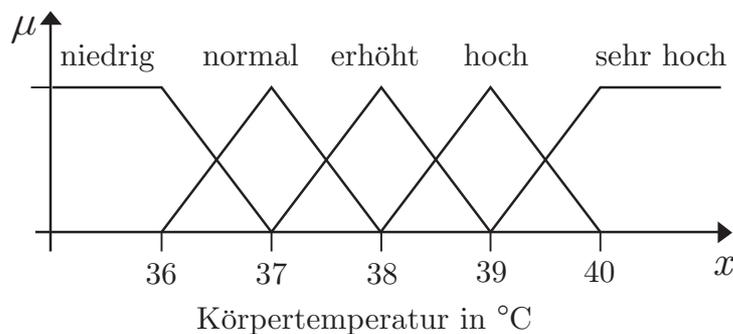


Abbildung 3.3: Beispiel für die Festlegung einer Standard-Fuzzy-Partitionierung.

Eine spezielle Form von Zugehörigkeitsfunktionen bilden *Singletons*, die fast ausschließlich als ausgangsseitige linguistische Werte oder zur eingangsseitigen Modellierung von Klassen verwendet werden.

Definition 3.6 *Singletons:* Singletons sind Zugehörigkeitsfunktionen, die an genau einer einzigen Stelle des Definitionsbereichs den Funktionswert Eins und für alle übrigen Werte des Definitionsbereichs den Wert Null annehmen.

Insgesamt werden aber im Modul der Fuzzifizierung alle Eingangsgrößenwerte x_i einer gegebenen Eingangssituation $\mathbf{x}^T = (x_1, \dots, x_n)$ in Zugehörigkeitsgrade $\mu_{i,j}(x_i)$ mit $i = 1, 2, \dots, n$ und $j = 1, 2, \dots, L_i$ umgewandelt. Für eine konkrete statische Eingangssituation werden im Folgenden die Werte für $\mu_{i,j}(x_i)$ zur Vereinfachung mit $\mu_{i,j}$ abgekürzt.

Inferenz

Die Generierung einer ausgangsseitigen Fuzzy-Menge $\mu_Y(\mathbf{x}, y)$ auf Grundlage der fuzzifizierten Eingangsgrößenwerte \mathbf{x} wird als Inferenz bezeichnet. Sie basiert auf der Auswertung der *Regelbasis* \mathcal{R} , die aus R Regeln der Form

$$\text{WENN } P_r \text{ DANN } C_r \quad \text{mit } r = 1, 2, \dots, R \quad (3.4)$$

besteht. Der so genannte WENN-Teil enthält die *Prämisse* P_r der r -ten Regel, die aus einer oder mehrerer konjunktiv verknüpfter¹ Partialprämissen A_i besteht.

Definition 3.7 *Partialprämisse: Eine Partialprämisse A_i ist eine Elementaraussage „ X_i ist $\mathcal{L}_{i,j}$ “ ($X_i = \mathcal{L}_{i,j}$) in einer Regel, in der nur eine linguistische Variable X_i und ein linguistischer Wert $\mathcal{L}_{i,j}$ vorkommen.*

Eine *vollständige Regel* r verknüpft für jede der n Eingangsvariablen des Fuzzy-Systems genau eine Partialprämisse $A_{r,i}$ gemäß

$$\text{WENN } A_{r,1} \wedge A_{r,2} \wedge \dots \wedge A_{r,n} \text{ DANN } C_r, \quad (3.5)$$

so dass die Prämisse P_r durch

$$P_r = \bigwedge_{i=1}^n A_{r,i} = \bigwedge_{i=1}^n (X_i = \mathcal{L}_{i,l_i}) \quad \text{mit } l_i \in \{1, 2, \dots, L_i\} \quad (3.6)$$

definiert wird. Für Fuzzy-Systeme mit vielen Eingangsgrößen besteht eine Regelbasis häufig aus unvollständigen (*generalisierenden*) Regeln. Die *Kombinationstiefe* c_r gibt dann die Anzahl der verknüpften Partialprämissen innerhalb der Prämisse P_r an.

Der DANN-Teil einer Regel r wird *Konklusion* C_r genannt und beinhaltet eine Elementaraussage, bezogen auf die Ausgangsgröße Y

$$C_r = (Y = \mathcal{L}_{Y,l_Y}) \quad \text{mit } l_Y \in \{1, 2, \dots, L_Y\}. \quad (3.7)$$

¹In dieser Arbeit wird die Betrachtung auf konjunktive Verknüpfungen von Partialprämissen beschränkt.

Die Inferenz wertet die vollständige Regelbasis \mathcal{R} aus und erzeugt für eine gegebene Eingangssituation \mathbf{x} eine ausgangsseitige Fuzzy-Menge, die auch *Empfehlungsgebirge* genannt wird. Dieses Empfehlungsgebirge gibt für jeden Ausgangsgrößenwert y des Definitionsbereichs genau den Grad $\mu_Y(\mathbf{x}, y)$ der Empfehlung an. Die dafür notwendigen Teilschritte sind die *Aggregation* (Prämissenauswertung), die *Aktivierung* (Komposition) und die *Akkumulation*.

Definition 3.8 *Aggregation:* Unter Aggregation wird die Bestimmung des Erfülltheitsgrads μ_P der Prämisse P_r einer linguistischen Regel in Abhängigkeit der Eingangsgrößenwerte \mathbf{x} durch Verknüpfung der Erfülltheitsgrade μ_{i,l_i} aller Partialprämissen $A_{r,i}$ mittels Fuzzy-Operatoren verstanden:

$$\mu_{P_r}(\mathbf{x}) := \bigwedge_{k=1}^c \mu_{i_k j_k}(x_{i_k}) .$$

Dabei entspricht k dem k -ten linguistischen Ausdruck in der Prämisse P_r . Eine Übersicht über die Definitionen der gebräuchlichsten Fuzzy-Operatoren ist in Tabelle 3.1 zusammengestellt. Aus der ermittelten Erfülltheit der Prämisse P_r wird die Aktivierung der Zugehörigkeitsfunktion der Konklusion C_r bestimmt.

Definition 3.9 *Aktivierung:* Mit Aktivierung wird die Bestimmung des Erfülltheitsgrads μ_C der Konklusion $C_r = (Y = \mathcal{L}_{Y,l_Y})$ einer linguistischen Regel aus dem Erfülltheitsgrad der Prämisse μ_{P_r} und einem eventuell vorhandenen Wichtigkeitsfaktor (Glaubensmaß) $J_r \in [0, 1]$ bezeichnet:

$$\mu_{C_r}(\mathbf{x}, y) := \mu_{P_r}(\mathbf{x}) \wedge \mu_{Y,l_Y}(y) \wedge J_r .$$

Das Glaubensmaß J_r wird häufig bei Regeln verwendet, die datenbasiert generiert wurden. Es basiert auf dem bei der Regelsuche verwendeten statistischen Test und ist ein Maß dafür, wie gut die Regel r durch die Daten gestützt wird. Abschließend werden im Prozess der Inferenz die aktivierten Zugehörigkeitsfunktion der Konklusion C_r akkumuliert.

Definition 3.10 (*Akkumulation*) Unter Akkumulation wird die Zusammenfassung der Erfülltheitsgrade $\mu_{C_r}(\mathbf{x}, y)$ der Konklusionen aller R linguistischen Regeln einer Regelbasis \mathcal{R} zu einer ausgangsseitigen Zugehörigkeitsfunktion $\tilde{\mu}_Y$, auch *Empfehlungsgebirge* genannt, verstanden:

$$\mu_Y(\mathbf{x}, y) := \bigvee_{r=1}^R \mu_{C_r}(\mathbf{x}, y) .$$

Die bisher betrachteten Regeln beziehen sich auf positives Erfahrungswissen, die Empfehlungen für ausgangsseitige linguistische Werte aussprechen. Eine Erweiterung der herkömmlichen Inferenz bildet die von KIENDL eingeführte Verwendung *negativer Regeln* [Kie93b, Kie93a, Kie94b, KKN96, Kie97a, Kra99], mit denen auch negatives Erfahrungswissen im Sinne von Warnungen und Verboten implementiert werden kann. Dabei handelt es sich um Regeln der Form

$$\text{WENN } P_r \text{ DANN } C_r \text{ VERBOTEN .} \quad (3.8)$$

Für diese Form der Regeln wird die Aggregation und Aktivierung wie zuvor beschrieben durchgeführt. Die Akkumulation wird jedoch für die positiven und negativen Regeln getrennt durchgeführt, so dass zwei ausgangsseitige Zugehörigkeitsfunktionen $\mu_Y^+(y)$ für *Empfehlungen* und $\mu_Y^-(y)$ für *Warnungen* berechnet werden. Die Verrechnung der beiden Zugehörigkeitsfunktionen erfolgt mit Hilfe der *Hyperinferenz* gemäß

$$\mu_Y(\mathbf{x}, y) = \mu_Y^+(y) \wedge \neg \mu_Y^-(y) . \quad (3.9)$$

Defuzzifizierung

Die Abbildung der akkumulierten ausgangsseitigen Zugehörigkeitsfunktion $\mu_Y(\mathbf{x}, y)$ auf einen numerischen (scharfen) Ausgangsgrößenwert y_D wird *Defuzzifizierung* genannt. Die beiden gräuchlichsten Defuzzifizierungsmethoden sind in Tabelle 3.2 angegeben.

Tabelle 3.2: Definitionen der gräuchlichsten Defuzzifizierungsmethoden.

Abk.	Bezeichnung	Definition
COG	Schwerpunktmethode	$y_D^{COG} := \int y \cdot \mu_Y(\mathbf{x}, y) \cdot dy / \int \mu_Y(\mathbf{x}, y) \cdot dy$
MAX	Maximum-Defuzzifizierung	$y_D^{MAX} := \{y_D \mu_Y(\mathbf{x}, y_D) = \max\{\mu_Y(\mathbf{x}, y)\}, \forall y \in Y\}$

Die MAX-Defuzzifizierung liefert dabei den am meisten empfohlenen Wert als Ausgangsgrößenwert $\hat{y} = y_D^{MAX}$. Für den Sonderfall, dass mehrere Maxima vorliegen, wird als häufigste Variante der mittlere Wert y_D^{MOM} (**M**ean **O**f **M**aximum) als Ausgangsgrößenwert verwendet. Sehr selten dagegen wird von der Möglichkeit der Selektion des minimalen Wertes y_D^{min} oder des maximalen Wertes y_D^{max} Gebrauch gemacht. Demgegenüber bildet die COG-Defuzzifizierung (**C**enter **O**f **G**ravity) aus allen Empfehlungen anhand ihrer Gewichte einen Kompromiss zum Ausgangsgrößenwert $\hat{y} = y_D^{COG}$.

In der Praxis werden sehr häufig ausgangsseitig Singletons (Definition 3.6) als Zugehörigkeitsfunktionen verwendet, weil diese im Allgemeinen einfacher und rechentechnisch schneller ausgewertet werden können. Die in Tabelle 3.2 angegebenen Methoden sind einfach auf den Fall ausgangsseitiger Singletons übertragbar und werden dann COS (**C**enter **O**f **S**ingletons) bzw. MOS (**M**aximum **O**f **S**ingletons) genannt.

Weitere Methoden zur Defuzzifizierung sind u. a. die von KIENDL eingeführte Drehmomentmethode (TOR/TOS) [Kie94b, Kie97b, Kie99] zur verstärkten Überlagerung gleichsinniger Empfehlungen sowie das Inferenzfilter [Kie94a, Kie96, KKN96, Kie97a, Kie99], welches herkömmliche Defuzzifizierungsmethoden in einem parametrisierten Ansatz verallgemeinert und durch *stufenlose Übergänge* zwischen den Defuzzifizierungsmethoden weiteres Optimierungspotenzial erschließt.

3.2.2 Takagi–Sugeno–Fuzzy–Systeme

Durch TAKAGI, SUGENO und KANG ist aus dem Mamdani–Fuzzy–System das ebenso verbreitete Takagi–Sugeno–Fuzzy–System entwickelt worden [TS85].

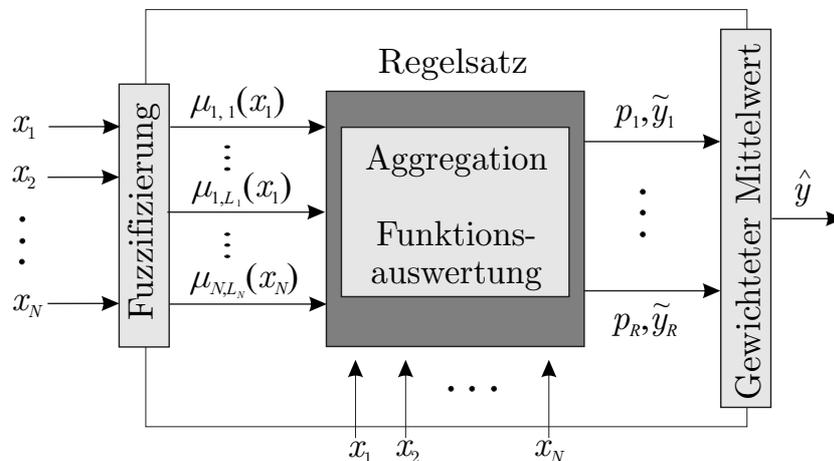


Abbildung 3.4: Takagi–Sugeno–Fuzzy–System.

Mit dem Ziel der Ermöglichung höherer Approximationsgüten wurden dabei die linguistischen Werte der Konklusion eines Mamdani–Systems durch funktionale Abhängigkeiten von den Eingangsgrößen \mathbf{x} ersetzt, so dass sich für Takagi–Sugeno–Regeln die Form

$$\text{WENN } P_r \text{ DANN } \tilde{y}_r = f_r(\mathbf{x}) \quad (3.10)$$

ergibt. Die zugehörige Struktur ist in Abbildung 3.4 dargestellt. Das Modul der Fuzzifizierung sowie die Aggregation werden mit analoger Funktionsweise

zum Mamdani-Fuzzy-System verwendet. Für die Auswertung des Regelsatzes kommt eine erweiterte Auswertung der Funktionen $f_r(\mathbf{x})$ der Regeln hinzu. Aus den Erfüllungstufen p_r der Prämissen P_r und den Funktionswerten \tilde{y}_r wird durch eine gewichtete Mittelwertbildung der Ausgangsgrößenwert \hat{y} mit

$$\hat{y} = \frac{\sum_{r=1}^R \mu_{P_r} \cdot \tilde{y}_r}{\sum_{r=1}^R \mu_{P_r}} \quad (3.11)$$

bestimmt. In [Kra00, KS01] wird eine Vorgehensweise zur datenbasierten Generierung von interpretierbaren Takagi-Sugeno-Kang-Systemen (TSK-Systeme) vorgeschlagen. Sie basiert auf einer Transformation von Mamdani-Regeln, die mit dem Fuzzy-ROSA-Verfahren generiert werden, in TSK-Regeln, wie sie auch im Rahmen dieser Arbeit verwendet wird.

3.3 Datenbasierte Fuzzy-Modellierung

Die im Rahmen dieser Arbeit zu entwickelnden Methoden zur Selektion relevanter nichtredundanter Merkmalsätze und Generierung von den Daten angepassten Fuzzy-Partitionen definieren die linguistischen Werte für alle selektierten Größen. Daraus ergibt sich direkt die Menge aller möglichen Fuzzy-Regeln ergibt. Die zwei grundlegenden Ansätze zur Regelgenerierung, die zur Bestimmung einer möglichst geeigneten Teilmenge als resultierender Regelsatz verfolgt werden können, sind in Abschnitt 2.1.2 beschrieben und hinsichtlich Interpretierbarkeit und Genauigkeit diskutiert worden. Ausgehend von diesen Kriterien ist für die im Rahmen dieser Arbeit zu generierenden Fuzzy-Systeme das von KIENDL, KRABS, KRONE und SLAWINSKI entwickelte Fuzzy-ROSA-Verfahren [KK89, Kra94, KK95, Kro99, Sla01, SKKK01] ausgewählt worden. Dieses datenbasierte Verfahren zeichnet sich durch eine effiziente Generierung kompakter und transparenter Regelsätze mit im Allgemeinen auch hoher Modellierungsgüte aus. Neben diesen Eigenschaften ist der Aspekt der Verfügbarkeit sowie die akzeptable durchschnittliche Generierungszeit von Regelsätzen bei der Auswahl des Fuzzy-ROSA-Verfahrens berücksichtigt worden.

Die Grundidee des Fuzzy-ROSA-Verfahrens wird in Abschnitt 3.3.1 beschrieben. Eine zusammenfassende Übersicht über die einzelnen Teilschritte des zugehörigen Regelgenerierungsprozesses erfolgt in Abschnitt 3.3.2. Daran angelehnt werden in Abschnitt 3.3.3 die wesentlichen Ansatzpunkte zur Effizienzsteigerung von datenbasierten Modellierungsverfahren konkretisiert, die im Rah-

men dieser Arbeit konzipiert und am Beispiel des Fuzzy–ROSA–Verfahrens erprobt werden.

3.3.1 Grundidee des Fuzzy–ROSA–Verfahrens

Ausgangspunkt des Fuzzy–ROSA–Verfahrens ist die 1989 von KIENDL und KRABS vorgeschlagene **R**egel–**O**rientierte **S**tatistische **A**nalyse (ROSA) [KK89] zur automatischen Generierung regelbasierter Systeme. Die grundlegende Idee des Verfahrens besteht darin, einzelne potenzielle Regeln anhand von Beobachtungsdaten mit Hilfe eines statistischen Tests daraufhin zu testen, ob sie einen relevanten Teilaspekt des betrachteten Systems beschreiben. Der dafür entwickelte Relevanzindex als Regeltest- und Bewertungsstrategie vereinigt ein aus der Literatur bekanntes Schema zur Bewertung der Unsicherheit von Regeln in Expertensystemen mit dem statistischen Konzept der Konfidenzintervalle [Kra94].

Im Gegensatz zu anderen Verfahren, die gemäß des Pittsburgh–Ansatzes den vollständigen Regelsatz evaluieren und optimieren, wird mit dem ROSA–Verfahren nach dem Michigan–Ansatz der Regelsatz sukzessive aus individuell getesteten und bewerteten Regeln aufgebaut. Damit wird das komplexe Problem, einen guten Regelsatz zu generieren, auf das einfachere Problem, einzelne relevante Regeln zu finden, zurückgeführt. Hieraus resultiert im Vergleich eine Verkleinerung des Suchraumes, da ausschließlich Regeln statt Mengen von Regeln betrachtet werden müssen. Des Weiteren wird durch die statistische Absicherung implizit gewährleistet, dass die Regeln in ihrem lokalen Wirkungsbereich vernünftige Aussagen beinhalten und somit auch wirklich interpretierbar sind. Infolgedessen ermöglicht dieser Ansatz einerseits, befriedigende Lösungen auch für komplexere Aufgabenstellungen in akzeptabler Zeit zu erzielen, andererseits wird das globale Optimum im Allgemeinen nicht erreicht.

Mit dem Ziel, diese Vorteile auch für die datenbasierte Fuzzy–Modellierung zu nutzen, ist 1994 die Vorgehensweise des ROSA–Verfahrens von KIENDL und KRONE auf den Fall von Fuzzy–Regeln übertragen worden [KK94, Kro99]. Die dabei verwendete Regelstruktur basiert auf Mamdani–Fuzzy–Regeln gemäß Gleichung (3.4). Das Fuzzy–ROSA–Verfahren berücksichtigt grundsätzlich auch generalisierende Regeln, deren Kombinationstiefe c kleiner als die gesamte Anzahl linguistischer Eingangsvariablen n ist. Insbesondere bei Modellierungsproblemen mit vielen Eingangsgrößen können die hochdimensionalen Suchräume mit weniger generalisierenden Regeln abgedeckt werden, so dass der resultierende Regelsatz im Vergleich zur Verwendung vollständiger Regeln meist kleiner und entsprechend besser interpretierbar ist. Weitere Strategieelemente zur Ana-

lyse und effizienten Generierung relevanter Fuzzy-Regeln in hochdimensionalen Suchräumen sind von SLAWINSKI zur Erweiterung der Anwendungsfelder des Fuzzy-ROSA-Verfahrens implementiert worden [Sla01]. Als jüngste Erweiterung des Fuzzy-ROSA-Verfahrens ist zur Erhöhung der Modellierungsgüte bei Approximationsproblemen von KRAUSE ein Verfahren entwickelt und implementiert worden, das ein datenbasiert generiertes Mamdani-Fuzzy-Modell in ein Takagi-Sugeno-Fuzzy-Modell umwandelt, ohne die Interpretierbarkeit und Transparenz des Mamdani-Fuzzy-Modells zu verlieren [Kra01].

3.3.2 Der Regelgenerierungsprozess

Der Regelgenerierungsprozess des Fuzzy-ROSA-Verfahrens lässt sich grob in die folgenden vier Schritte einteilen [SKKK01]:

Projektdefinition: Vor der eigentlichen Regelgenerierung müssen die linguistischen Werte für die Eingangs-/Ausgangsvariablen des betrachteten Systems festgelegt werden. Dies kann wissensbasiert, datenbasiert oder heuristisch erfolgen [KS98]. Außerdem kann die maximale Kombinationstiefe c_{max} für die Prämisse und gegebenenfalls eine maximale Zeittiefe t_{max} festgelegt werden, um den Rechenaufwand zu beschränken.

Regelgenerierung: Abhängig von der Suchraumgröße kann eine komplette Suche, eine evolutionäre Suche oder eine Kombination von beiden gewählt werden [SKKK01]. Der Regelsatz wird sukzessive aus allen relevanten, nicht redundanten Regeln aufgebaut. Die Bewertungsgrundlagen, nach denen Regeln getestet und bewertet werden, werden in Anhang B beschrieben.

Regelreduktion: Die Anzahl der Regeln kann anschließend durch Offline-Regelreduktionsverfahren verringert werden [Kro96]. Dabei können verschiedene Anforderungen, wie z. B. komplette Überdeckung aller Eingangssituationen, gleichmäßige Ausnutzung der Daten, Verringerung des Modellierungsfehlers oder der Regelanzahl, berücksichtigt werden [KKS00b].

Regelsatzanalyse und -optimierung: Durch die abschließende Analyse des Regelsatzes kann der Modellierungsprozess und das Modellierungsergebnis bewertet und gegebenenfalls Feedback für die Problemformulierung erhalten werden. Zusätzlich kann das Eingangs-/Ausgangsverhalten des erhaltenen Fuzzy-Systems durch Anpassung der verbleibenden freien Parameter optimiert werden [KKS00b, KKS00a, KWS00].

3.3.3 Ansatzpunkte für die vorliegende Arbeit

Wie in Kapitel 2.2 illustriert, werden mit der *Projektdefinition* direkt zu Beginn des Regelgenerierungsprozesses die erfolgsbestimmenden Einstellungen bezüglich der erzielbaren Modellgüte gelegt. Typischerweise werden für die Modellierung mit dem Fuzzy–ROSA–Verfahren gemäß der von SLAWINSKI entwickelten *Systematischen Vorgehensweise* alle potenziellen Einflussgrößen X_i unter Verwendung von äquidistant verteilten Standard–Fuzzy–Partitionen bei der Regelgenerierung berücksichtigt. Im Rahmen dieser Arbeit sollen Verfahren zur Effizienzsteigerung des Regelgenerierungsprozesses erarbeitet werden, die automatisiert und datenadaptiv besonders günstige Vorgaben dieser entscheidenden Einstellungen zur direkten Verwendung in der *Projektdefinition* erstellen:

Selektion relevanter Einflussgrößen Für den Regelgenerierungsprozess ist aus einer gegebenen Menge \mathcal{X} von potenziellen Einflussgrößen eine Teilmenge \mathcal{M} von denjenigen Einflussgrößen X_i zu selektieren, die entweder direkt oder im Zusammenhang mit weiteren Einflussgrößen X_j Informationen über die zu modellierende Ausgangsgröße Y beinhalten.

Eliminierung redundanter Einflussgrößen Für den Regelgenerierungsprozess sind aus der ermittelten Teilmenge \mathcal{M} relevanter Einflussgrößen diejenigen Einflussgrößen X_i zu eliminieren, deren Informationsgehalt bezüglich der zu modellierenden Ausgangsgröße Y bereits von anderen relevanteren Einflussgrößen X_j abgedeckt wird.

Optimierung der Fuzzy–Partitionen Für die selektierten relevanten und nichtredundanten Einflussgrößen X_i der Menge \mathcal{M} sind zur Bestimmung der linguistischen Werte $\mathcal{L}_{i,j}$ zugehörige datenadaptierte Fuzzy–Partitionen derart zu generieren, dass vom Regelgenerierungsprozess ein möglichst transparentes und generalisierungsfähiges Modell mit hoher Modellgenauigkeit gebildet werden kann.

3.4 Selektion relevanter Merkmale

Ein Hauptproblem bei der datenbasierten Fuzzy–Modellierung besteht in der Auswahl der zu berücksichtigenden Einflussgrößen. Erfahrungsgemäß sind nur selten alle gemessenen oder zur Verfügung stehenden Größen des zu modellierenden Systems notwendig, um die zugrunde liegenden Prozesszusammenhänge

mit ausreichender Genauigkeit zu modellieren. Der Erfolg einer datenbasierten Fuzzy-Modellierung hängt somit häufig entscheidend von einer geeigneten Auswahl der tatsächlich berücksichtigten Einflussgrößen aus einer Vielzahl von potenziell wichtigen Einflussgrößen ab. Eine dem Modellierungsprozess vorgelagerte Selektion von relevanten und nichtredundanten Einflussgrößen wirkt sich aus mehreren Gründen oftmals vorteilhaft aus:

- höhere erzielbare Modellgenauigkeit
- deutliche Reduzierung des Modellierungsaufwandes
- erhöhte Transparenz des generierten Fuzzy-Modells
- verringerter Aufwand der Modellauswertung.

Eine Einteilung der aus der Literatur bekannten Selektionsverfahren kann zum einen anhand der verwendeten *Bewertungsfunktion* und zum anderen auf der Basis der verfolgten *Suchstrategie* vorgenommen werden. Dabei ist es die Aufgabe der Bewertungsfunktion, einzelnen Einflussgrößen oder einem ganzen Satz betrachteter Größen (einem potenziellen *Merkmalsatz*) einen Relevanzwert zuzuordnen. Die verwendete *Suchstrategie* bestimmt die Vorgehensweise, wie im Raum aller möglichen Merkmalsätze ein im Sinne der verwendeten Bewertungsfunktion möglichst günstigster Merkmalsatz gesucht wird.

3.4.1 Bewertungsfunktion

In der Literatur werden bezüglich der Wahl der Bewertungsfunktion zwei grundsätzlich verschiedene Ansätze [JKP94, KJ97, KW00] voneinander unterschieden.

Beim sogenannten *Wrapperansatz* (engl. *wrapper* = einwickeln) wird die Relevanz eines Satzes von Einflussgrößen dadurch ermittelt, dass mit diesem Satz von Einflussgrößen anhand der verfügbaren Messdaten eine Modellidentifikation mit Hilfe des vorgesehenen Modellierungsverfahrens vorgenommen und die resultierende Modellgüte als Maß für die Relevanz angesehen wird. Ein wesentlicher Vorteil dieses Ansatzes besteht darin, dass der *Bias* der Bewertungsfunktion mit dem des hinterher verwendeten Modellierungsverfahrens übereinstimmt. Demgegenüber wird bei dem sogenannten *Filteransatz* ein im Allgemeinen einfach und schnell zu berechnendes Maß zur Bewertung eines Merkmalsatzes oder einer einzelnen potenziellen Einflussgröße verwendet. Bei einem Filteransatz kann somit nicht gewährleistet werden, dass eine im Sinne der Bewertungsfunktion als günstig bewertete Lösung sich auch als günstig für

das nachgeschaltete Modellierungsverfahren erweist. Der wesentliche Vorteil des Filteransatzes hingegen liegt in dem meist vergleichsweise wesentlich geringeren Rechenaufwand. In Anbetracht der Komplexität und Rechenintensivität des Fuzzy–ROSA–Verfahrens wird der Filteransatz im Folgenden verfolgt.

Eine etablierte Bewertungsfunktion für die Relevanzbewertung einzelner reellwertiger Einflussgrößen X im Sinne eines linearen Zusammenhangs mit der Ausgangsgröße Y ist die aus der Statistik bekannte *Korrelation* [HEK95]. Sie bewertet die Relation zwischen genau einer Einflussgröße und einer Ausgangsgröße. Die ebenfalls aus der Statistik bekannte *Hauptkomponentenanalyse* [Jac91] verallgemeinert eine solche Bewertung auf eine beliebige Anzahl n von Einflussgrößen X_1, X_2, \dots, X_n . Hierzu werden diese durch eine lineare Transformation in q möglichst unkorrelierte, also orthogonal zueinander stehende Komponenten \mathbf{M} , die sogenannten Hauptkomponenten, transformiert. Das Verfahren der Hauptkomponentenanalyse hat neben der Beschränkung auf einen linearen Strukturansatz den weiteren wesentlichen Nachteil, dass die neu ermittelten Komponenten M_1, M_2, \dots, M_q aufgrund ihrer Berechnung als Linearkombination der ursprünglichen Einflussgrößen X_1, X_2, \dots, X_n im Allgemeinen keine interpretierbaren Systemgrößen darstellen. Die erzielbare Transparenz eines darauf aufsetzenden Modells wäre dementsprechend äußerst gering.

Zur Analyse nichtlinearer Abhängigkeiten zweier Einflussgrößen mit diskret verteilten Werten ist aus der Informationstheorie die *Transinformation* bekannt [Mat96]. Von [Pra99] wird für diese Aufgabenstellung ein Eindeutigkeitsmaß definiert, welches insbesondere auch für fuzzifizierte Größenwerte anwendbar ist. Es untersucht potenzielle Einflussgrößen auf lineare und nichtlineare Zusammenhänge mit der Ausgangsgröße (Relevanzanalyse) und mit den übrigen potenziellen Einflussgrößen (Redundanzanalyse). Das Verfahren berücksichtigt eine durch Fuzzy–Zugehörigkeitsfunktionen vorzugebende Granularisierung (*Fuzzy–Partitionierung*) der Einflussgrößen und der Ausgangsgröße und ist daher auf eine nachgeschaltete Fuzzy–Modellierung zugeschnitten. Allerdings bewertet das Verfahren die potenziellen Einflussgrößen einzeln und nicht — was angemessener wäre — ganze daraus zusammengestellte Sätze von Einflussgrößen. Zudem verbleibt das Problem einer günstigen Granularisierung der Einflussgrößen. Insbesondere bei hochdimensionalen Modellierungsproblemen komplexer Systeme wird, wie in 2.1.4 dargelegt, eine Partitionierung häufig heuristisch durch äquidistante Raumeinteilung vorgenommen. Aufgrund einer derartigen, nicht optimal gewählten, Granularisierung kann die Relevanz einer potenziellen Einflussgröße falsch bewertet sowie die mit dem Fuzzy–ROSA–Verfahren erzielbare Modellierungsgüte erheblich beeinträchtigt werden.

Der beschriebene Mangel des Verfahrens aus [Pra99], dass die potenziellen Einflussgrößen einzeln bewertet werden, wird in [Xio01] abgestellt. Dort wird ein Verfahren zur Auswahl eines relevanten Satzes von potenziellen Einflussgrößen vorgestellt. Allerdings erfolgt die Relevanzbewertung ohne Berücksichtigung einer Granularisierung des Eingangsraumes. Dies ist nachteilig für eine nachgeschaltete Fuzzy-Modellierung.

3.4.2 Suche von relevanten Sätzen von Einflussgrößen

Für die Suche eines Merkmalsatzes, der im Sinne der verwendeten Bewertungsfunktion als möglichst relevant bewertet wird, werden in der Literatur verschiedene Ansätze vorgeschlagen.

Die *vollständige Suche* durchmustert den Raum aller möglichen Kombinationen der vorhandenen Einflussgrößen und findet daher garantiert das globale Optimum. Die exponentiell mit der Anzahl vorhandener Einflussgrößen anwachsende Größe des Suchraums führt unweigerlich zu einem immens hohen Rechenaufwand, so dass diese Strategie in der Praxis nur selten anwendbar ist. Meistens werden daher sogenannte *Greedy-Algorithmen*, etwas seltener auch *Genetische Algorithmen*, verwendet [IV94, DM98].

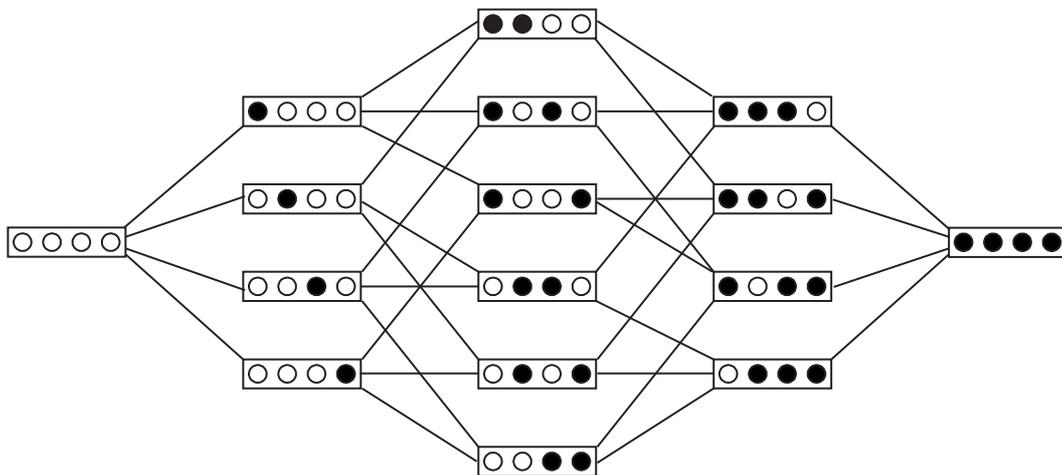


Abbildung 3.5: Veranschaulichung des Ablaufs von Greedy-Algorithmus am Beispiel von vier potenziellen Einflussgrößen. Jeder Knoten des Graphen repräsentiert einen Merkmalsatz, der durch die Teilmenge der berücksichtigten Einflussgrößen (markiert durch gefüllte Kreise) definiert wird. Zwei miteinander verbundene Merkmalsätze unterscheiden sich in genau einer Einflussgröße.

Bekannte *Greedy-Algorithmen* sind die sogenannte *Rückwärts-* und *Vorwärtsselektion*. Bei der Vorwärtsselektion wird zunächst die Relevanz einer jeden potenziellen Einflussgröße separat ermittelt und die Einflussgröße mit der höchsten

Relevanz als erste Größe des Merkmalsatzes übernommen. Darauf aufbauend wird der Merkmalsatz iterativ um diejenige noch nicht ausgewählte Einflussgröße erweitert, deren Hinzufügung zum bereits selektierten Merkmalsatz die größte Relevanzsteigerung erbringt. Dieser sequentielle Erweiterungsprozess wird beendet, wenn keine oder keine „signifikante“ Relevanzsteigerung durch Hinzufügung einer weiteren potenziellen Einflussgröße erzielt werden kann. Bei der Festlegung, was als „signifikant“ anzusehen ist, ist ein meist anwendungsspezifischer Kompromiss zwischen den gegenläufigen Zielen „hohe Relevanz“ und „möglichst wenig Einflussgrößen“ zu schließen. Demgegenüber startet die *Rückwärtsselektion* mit der Relevanzanalyse des vollständigen Satzes aller potenziellen Einflussgrößen und verkleinert diesen iterativ immer um diejenige Einflussgröße, die bei Entfernung aus dem bestehenden Merkmalsatz die größte Relevanzsteigerung (oder auch geringste Relevanzverminderung) auslöst. Falls sich bei der Relevanzanalyse eines verkleinerten Merkmalsatzes in jedem Fall eine Verschlechterung der Relevanz ergibt, wird derjenige Merkmalsatz ausgewählt, der zur geringsten Relevanzverschlechterung führt, sofern diese als Preis für die Verkleinerung des Merkmalsatzes noch akzeptabel ist.

3.5 Evolutionsstrategien

3.5.1 Einführung

Evolutionsstrategien ordnen sich in die Klasse der *Evolutionären Algorithmen* ein, die durch die biologische Evolution inspiriert wurden. Dabei geht man von der Sichtweise aus, dass die natürliche Evolution selbst ein Optimierungsprozess ist, dessen Ziel es ist, die Fähigkeit des Überlebens eines Individuums zu verbessern. Im Jahre 1859 veröffentlichte Charles Darwin diese Erkenntnis in seinem Buch „*On the Origin of Species by Means of Natural Selection*“. Was seitdem als Evolutionstheorie bekannt und bisher nicht widerlegt ist, beruht auf drei grundlegenden Beobachtungen und Annahmen:

- Nahezu alle Lebewesen erzeugen mehr Nachkommen, als tatsächlich überleben können.
- Lebewesen einer Art können sich zwar sehr stark ähneln, sind aber nie vollständig identisch.
- Erbinformationen werden an eine Nachfolgeneration nur weitergegeben, wenn sich diese im Kampf um Anpassung und Überleben bewährt haben.

Die meisten Nachkommen einer Art sterben also, bevor diese sich wieder vermehren können. Lebewesen einer Art unterliegen demnach einem *Selektionsdruck*, welcher die *Populationsgröße*, also die Anzahl aller Individuen einer Art, auf natürliche Art und Weise konstant hält. Die genetische Information eines Lebewesens ist in seinem Genom oder auch Gensatz dargestellt, welches sich aus einzelnen Genen zusammensetzt (*Genotyp*). Lebewesen unterscheiden sich in der Art ihres genetischen Materials, so dass einige aufgrund ihres Erscheinungsbildes, dem sogenannten *Phänotyp*, besser an ihre Umgebung angepasst sind als andere. Ein solcher Kampf um die beste Anpassung an Umweltbedingungen wird als *natürliche Selektion* oder *survival of the fittest* bezeichnet [Nis97]. So werden vor allem nur die besten Lebewesen ihrer Art Nachkommen zeugen, deren Fitness also am höchsten ist.

Das Erbgut eines Nachkommen setzt sich dann aus den Eigenschaften beider Elternteile zusammen, in der Hoffnung, dass so nur die besten Eigenschaften weitergegeben werden. Gelegentlich unterliegt das Erbmaterial auch der Mutation, die zu zufälligen Änderungen einzelner Erbinformationen führt. Dabei können sowohl eine Verschlechterung als auch eine Verbesserung der Fitness eintreten, die das weitere Überleben des Individuums entsprechend beeinflusst. Auf diese Weise entwickeln sich Nachfolgenerationen, die immer besser an ihre Lebensräume (die sich im Laufe der Zeit auch verändern können) angepasst sind.

Evolutionäre Algorithmen simulieren genau diesen Evolutionsprozess anhand eines als *Gütefunktion* (häufig auch als *Zielfunktion* bezeichnet) definierten Lebensraumes, des mathematisch formulierten Optimierungsproblems. Durch die Wahl einer Menge zufällig erzeugter Individuen entspricht die Anwendung der Selektion einem stochastischen Suchprozess im Raum aller möglichen Chromosome, der zu einer optimalen Lösung eines Optimierungsproblems führen kann. Im Bereich der Evolutionären Algorithmen haben sich im Laufe der Zeit folgende Ausprägungen etabliert:

Genetische Algorithmen (*GA*) wurden als erste jemals entworfene und angewendete evolutionäre Paradigma von HOLLAND und GOLDBERG in den 70er Jahren publiziert [Hol75, Gol89]. Die Codierung der Informationen eines Individuums erfolgt in der Regel durch einzelne Bits, ein ganzes Chromosom wird durch einen Bit-String repräsentiert. Alternativ gibt es aber auch vektorbasierte Darstellungen, Codierungen in Form von Matrizen oder Permutationen. Eine geeignete Wahl der Repräsentation ist nicht allgemeingültig formulierbar, sondern problemspezifisch zu definieren. Evolutionäre Operatoren hängen somit auch von der Codierung ab, im Falle

von Bit-String-Darstellung werden binäre Operationen auf die einzelnen Individuen angewendet. Der maßgebliche Operator zur Reproduktion für genetische Algorithmen ist die Rekombination. Mutation spielt nur eine geringe Rolle und tritt nur mit einer sehr kleinen Wahrscheinlichkeit auf.

Genetische Programmierung (GP) kann als Spezialisierung genetischer Algorithmen betrachtet werden. Auch hier unterliegen lediglich die Genotypen der Evolution, wobei sich nur in der Repräsentation von Individuen Unterschiede ergeben. Hierfür werden keine Zahlen, Vektoren oder Bits verwendet, sondern ganze ausführbare Programme herangezogen. Dabei löst jedes Programm für sich ein Teilproblem. Gesucht wird die Kombination von Teilprogrammen, welche ein größeres komplexeres Problem am effizientesten löst. Dabei werden die Individuen in Baumstrukturen verwaltet. Die Bewertung der Fitness eines Individuums erfolgt aufgrund der Leistung des so generierten Computerprogramms bei Anwendung auf das Optimierungsproblem.

Evolutionäre Programmierung (EP) unterscheidet sich grundlegend von den beiden zuvor genannten evolutionären Paradigmen. Anders als GA und GP liegt hierbei der Schwerpunkt auf der Entwicklung von Verhaltensmodellen, es wird die Anpassung von Verhalten in der Evolution simuliert. Demzufolge werden hier die Phänotypen verändert, so dass eine Menge von optimalen Verhalten aus einem Raum von beobachtbaren Verhalten erzeugt wird. Von den evolutionären Operatoren wird Rekombination nicht verwendet. Die Nachfolgenerationen werden ausschließlich durch Mutation erzeugt.

Evolutionsstrategien (ES) wurden in den sechziger Jahren von RECHENBERG und SCHWEFEL [Sch77] entwickelt und erlebte ihren Durchbruch mit der zunehmenden Rechenleistung der heutigen Computer. Als wesentlicher Grundbestandteil dieser Arbeit werden diese daher im Folgenden ausführlich beschrieben.

3.5.2 Ablauf einer Evolutionsstrategie

Der Aufgabenbereich von Evolutionsstrategien besteht in der Optimierung einer gegebenen reellwertigen Zielfunktion $f(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}$ mit n reellen zu optimierenden Parametern x_1 bis x_n . In der Regel beschränken sich technische Optimierungsprobleme aufgrund der gegebenen Randbedingungen auf Suchräume S in Form eines Teilraumes $S \subset \mathbb{R}^n$.

Ein Individuum **Ind** beinhaltet eine potenzielle Lösung \mathbf{x} (*Zielvariablenvektor*) des Optimierungsproblems sowie einen *Strategieparametervektor* $\boldsymbol{\sigma}$, dessen Komponenten die Schrittweiten für den genetischen Operator *Mutation* sind:

$$\mathbf{Ind} = \left(\underbrace{x_1, x_2, \dots, x_n}_{\text{Zielvariablen}}, \underbrace{\sigma_1, \sigma_2, \dots, \sigma_n}_{\text{Strategieparameter}} \right). \quad (3.12)$$

Der Wert $f(\mathbf{x})$ entspricht dem *Fitness-* oder *Gütwert* eines Individuums. Gesucht im Sinne des Optimierungsproblems ist das Individuum mit der besten Fitness, das heißt mit dem kleinsten bzw. größten Wert $f(\mathbf{x})$. Dazu arbeitet die Evolutionsstrategie gemäß Abbildung 3.6 über mehrere *Generations* j mit einer Population \mathcal{P}_j von μ Individuen, von denen über die genetischen Operatoren der *Rekombination* und *Mutation* in jeder Generation λ Nachkommen erzeugt werden. Eine nachfolgende *Selektion* entscheidet, welche der Individuen die Elternpopulation \mathcal{P}_{j+1} der nachfolgenden Generation $j + 1$ werden.

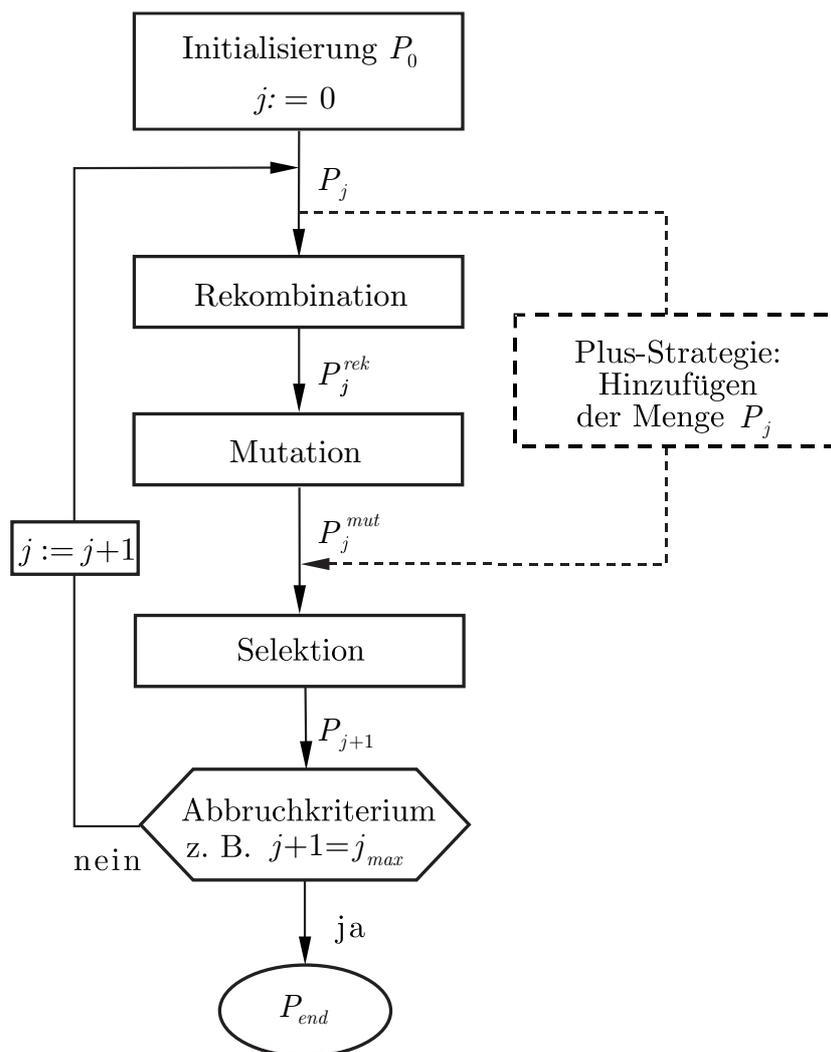


Abbildung 3.6: Üblicher Ablauf einer Evolutionsstrategie.

Startpopulation

Als erster Schritt wird eine Startpopulation \mathcal{P}_0 erzeugt. Im Interesse einer guten Breitensuche wird diese in der Regel zufällig im Suchraum S generiert. Wenn der Suchraum nicht begrenzt ist und bereits Vorwissen über eine relativ gute Lösung des Optimierungsproblems vorliegt, wird häufig eine normalverteilte Startpopulation um diesen Lösungspunkt im Suchraum erzeugt. Andererseits besteht hierbei auch die Gefahr, dass der Optimierungsprozess gegen diese bereits bekannte Lösung konvergiert oder diese nur noch wenig verbessern kann, obwohl an anderen Stellen im Suchraum bessere Lösungen zu finden sind. In diesem Fall wäre der Suchraum nicht breit genug durchsucht worden. Die Startpopulation kann auch deterministisch, zum Beispiel über ein Gitternetz, festgelegt werden. Sind bereits mehrere relativ gute Lösungen des Optimierungsproblems bekannt, so kann die Menge dieser Vektoren als Startpopulation verwendet werden. Allerdings besteht hierbei ebenfalls die oben genannte Gefahr der frühzeitigen Konvergenz gegen eine nur suboptimale Lösung.

Die Strategieparametervektoren $\boldsymbol{\sigma}$ mit den komponentenweisen Schrittweiten der Mutation werden zufällig gemäß

$$\sigma_i = |\tilde{\sigma}_i \cdot N_i(0, 1)| \quad \text{mit } i = 1, \dots, n \quad (3.13)$$

erzeugt. Hierbei ist $N(0, 1)$ eine normalverteilte skalare Zufallsvariable mit Erwartungswert Null und Standardabweichung Eins,

$$P(s) = \frac{1}{\sqrt{2 \cdot \pi}} \cdot e^{-\frac{1}{2} \cdot s^2}, \quad (3.14)$$

deren Wert unabhängig für jede Dimension i des Optimierungsproblems erzeugt wird. $\tilde{\sigma}_i$ ist eine auf die Breite des Suchraums in der Dimension i normierte Standardabweichung:

$$\tilde{\sigma}_i = 0.05 \cdot (x_{i,max} - x_{i,min}). \quad (3.15)$$

Die Schrittweitenvektoren werden in dieser Arbeit komponentenweise durch $\boldsymbol{\sigma}_{min} = 0.001 \cdot (\mathbf{x}_{max} - \mathbf{x}_{min})$ und $\boldsymbol{\sigma}_{max} = 0.1 \cdot (\mathbf{x}_{max} - \mathbf{x}_{min})$ begrenzt. Kleinere oder größere Suchschritte sind dennoch möglich, da die Schrittweitenvektoren der einzelnen Individuen ausschließlich die Standardabweichungen der im Programmschritt Mutation auftretenden normalverteilten Suchschritte festlegen. Die Individuen der Startpopulation werden in dieser Arbeit für reale Optimierungsprobleme immer gleichverteilt im Suchraum erzeugt. Lediglich zum Vergleich neuer Strategieelemente für den evolutionären Optimierungsprozess werden alle Individuen der Startpopulation auf einem einzigen Punkt im Suchraum platziert, um so künstlich erschwerte Anfangsbedingungen zu schaffen.

Rekombination

Die Rekombination — auch Fortpflanzung genannt — erzeugt in der Generation j aus mehreren bereits bestehenden Vektoren \mathbf{x} der Menge \mathcal{P}_j (der sogenannten Elternpopulation) neue potenzielle Lösungen des Optimierungsproblems. Dabei werden auch die Strategieparameter $\boldsymbol{\sigma}$ jedes neu erzeugten Individuums in Abhängigkeit von den Strategieparametern der Eltern-Individuen, die zur Erzeugung dieses Individuums verwendet wurden, ebenfalls zufällig neu generiert.

Die Idee der Rekombination ist, dass gute Individuen ihre Gene austauschen, um eventuell noch bessere Lösungen zu erzielen. Die Anzahl der “Sexualpartner”, die zur Erzeugung eines Nachkommens herangezogen wird, wird mit der *Gruppenpaarungszahl* ρ bezeichnet. Die Rekombination ist ein ungerichteter Operator, das heißt, die ρ Individuen werden unabhängig von ihrer Qualität zufällig mit gleicher Wahrscheinlichkeit aus der Menge der μ Eltern gewählt. Für jeden Nachkommen wird eine neue Menge von ρ Eltern ermittelt. Es ergeben sich drei übliche Formen der Rekombination [Eng03]:

Lokale diskrete Rekombination Ein Nachkomme \mathbf{Ind}_j^{rek} wird aus zwei Eltern erzeugt, wobei die Parameter zufällig gleichverteilt aus beiden Eltern gewählt werden.

Globale diskrete Rekombination Jede Komponente des Nachkommens \mathbf{Ind}_j^{rek} wird durch die entsprechende Position eines zufällig aus der Menge der ρ Eltern gewählten Individuums gebildet. Für jede Komponente wird zufällig ein neues Individuum bestimmt.

Intermediäre Rekombination In diesem Falle wird das arithmetische Mittel aller involvierten Eltern als Parameterwert des Nachkommens verwendet. Der Nachkomme \mathbf{Ind}_j^{rek} entspricht also dem Schwerpunkt der ρ Individuen \mathbf{Ind}_j^k der Elternpopulation \mathcal{P}_j :

$$\mathbf{Ind}_j^{rek} = \frac{1}{\rho} \sum_{k=1}^{\rho} \mathbf{Ind}_j^k. \quad (3.16)$$

In den meisten Anwendungsfällen erweist sich die Strategie, Optimierungsparameter \mathbf{x} mit einer lokal diskreten Rekombination und Strategieparameter $\boldsymbol{\sigma}$ über eine intermediäre Rekombination zu verändern, als besonders effizient [BHS97].

Mutation

Im Anschluss an die Rekombination wird üblicherweise die Menge \mathcal{P}_j^{rek} dem Programmschritt Mutation unterworfen und dadurch die Menge \mathcal{M}_j^{mut} erzeugt. Hierbei können alle oder nur einige der Optimierungsparameter \mathbf{x} aus \mathcal{P}_j^{rek} durch die Mutation verändert werden. Dabei sind kleine Änderungen wahrscheinlicher als große. Bei Evolutionsstrategien wird die Mutation durch Addition eines n -dimensionalen Zufallsvektors $\boldsymbol{\rho}$ zu den bereits bestehenden Optimierungsparametern \mathbf{x} realisiert. Üblicherweise wird hierbei ein normalverteilter Zufallsvektor $\boldsymbol{\rho}$ mit Erwartungswert Null und vektorieller Standardabweichung σ verwendet. Mit

$$\boldsymbol{\rho} = (N_1(0, \sigma_1), \dots, N_n(0, \sigma_n)) \quad (3.17)$$

wird der Mutationsschritt für ein Individuum \mathbf{Ind}_i folgendermaßen vorgenommen:

$$\mathbf{x}_i^{mut} = \mathbf{x}_i + \boldsymbol{\rho}. \quad (3.18)$$

Die Mutation stellt eine wesentliche Komponente der Evolutionsstrategie dar, denn nur durch die Mutation ist es möglich, in bestimmte Bereiche des Suchraums S zu gelangen. Dies können beispielsweise Bereiche sein, die am Rand des Suchraums liegen. Bei den gängigen Rekombinationsmethoden treten nur Nachkommen-Individuen auf, deren Parameterwerte zwischen denen der Eltern-Individuen liegen (intermediäre Rekombination) oder mit diesen identisch sind (diskrete Rekombination). Die Mutation dient dazu, auch Individuen außerhalb dieses Gebiets zu erzeugen.

Von entscheidender Bedeutung für die erfolgreiche Wirkungsweise des Programmschritts Mutation ist eine geeignete Schrittweite für die zufallsgesteuerten Mutationsschritte. Zu große Schrittweiten können dazu führen, dass die Suche einer reinen Zufallssuche gleich wird, zu kleine Schrittweiten bringen kaum Änderung und ermöglichen daher kaum das Verlassen eines lokalen Minimums. Die Festlegung der Startschrittweiten erfolgt bereits bei Generierung der Startpopulation. Für die Veränderung der Schrittweiten sowie der Individuen \mathbf{Ind}_i der Menge \mathcal{P}_j^{rek} im Programmschritt Mutation existieren zwei grundlegende Strategien.

Nach RECHENBERG ist die 1/5-Erfolgsregel anzuwenden, nach der während des Optimierungsprozesses 1/5 der Mutationen erfolgreich sein sollten, also zu einer Verbesserung des Gütewertes führen. Sind weniger als 1/5 der Mutationen erfolgreich, so ist die Schrittweite zu verkleinern, sind mehr als 1/5 der Mutationen erfolgreich, so ist die Schrittweite zu vergrößern.

Von SCHWEFEL wurde die 1/5–Erfolgsregel konkretisiert, d. h. es wurde eine algorithmische Umsetzung erarbeitet. Eine Erweiterung der Vorgehensweise von RECHENBERG stellen Mutationsschritte unter Verwendung eines n -dimensionalen Zufallsvektors mit *vektorieller* Schrittweite σ dar. In diesem Zusammenhang geht auf SCHWEFEL das Verfahren der *Selbstadaption der Schrittweiten* zurück [Sch77]. Hierbei wird bei jedem Mutationsschritt zunächst die Schrittweite des betrachteten Nachkommen–Individuums aus \mathcal{P}_j^{rek} mutiert, d. h. einer zufälligen Änderung unterworfen, und anschließend unter Verwendung dieser mutierten Schrittweite der eigentliche Mutationsschritt des betrachteten Individuums durchgeführt. Zunächst wird die neue Schrittweite des zu mutierenden Nachkommen–Individuums erzeugt:

$$\sigma_i^{mut} = \sigma_i \cdot e^{N(0,\tau_1)+N_i(0,\tau_2)} \quad \text{für } i = 1, \dots, n \quad (3.19)$$

mit

$$\tau_1 = \frac{c}{\sqrt{2 \cdot n}}, \quad \tau_2 = \frac{c}{\sqrt{2 \cdot \sqrt{n}}} \quad \text{und } c = 1. \quad (3.20)$$

Die Standardabweichungen der Normalverteilungen in Gleichung (3.19) sind über τ_1 und τ_2 nach Gleichung (3.20) parametrisiert, wodurch sich der Optimierungsprozess speziell an die Gütefunktion anpassen lässt. Dabei wirkt τ_1 global für alle Komponenten eines Individuums, τ_2 wirkt individuell für jede einzelne Komponente. Das Nachkommen–Individuum \mathbf{Ind}_i^{rek} wird analog Gleichung 3.17 über:

$$\boldsymbol{\rho} = (N_1(0, \sigma_1^{mut}), \dots, N_1(0, \sigma_n^{mut})) \quad (3.21)$$

zu

$$\mathbf{x}_i^{mut} = \mathbf{x}_i + \boldsymbol{\rho} \quad (3.22)$$

durchgeführt. Nähere Ausführungen hierzu sind in [Bäc94, BS96] zu finden.

Selektion

Der letzte Programmschritt innerhalb einer Iteration der Optimierung ist die Selektion. Aus der gesamten Population werden die μ Individuen \mathbf{Ind}_i bestimmt, die als Teil der Elternpopulation \mathcal{P}_{j+1} der nächsten Generation überleben. Es werden dabei zwei verschiedene Strategien unterschieden:

Plus-Strategie Bei der $(\mu + \lambda)$ –ES werden die gemäß ihrer Fitness μ besten Individuen aus der Vereinigungsmenge der Eltern– und Nachkommen–Individuen gewählt.

Komma-Strategie Bei der (μ, λ) -ES werden die gemäß ihrer Fitness μ besten Individuen nur aus der Menge der λ Nachkommen gewählt. Dabei ist natürlich die Bedingung $\lambda \geq \mu$ einzuhalten. Dabei wird sehr häufig nach dem *Eliteprinzip* selektiert, das heißt, die besten μ Individuen der aktuellen Population \mathcal{P}_j werden die Eltern der nächsten Generation \mathcal{P}_{j+1} . Der Fitnesswert dient somit dabei als alleiniges Bewertungskriterium. Damit ist die Selektion das einzige gerichtete Element der Evolutionsstrategie, da sie vorgibt, wohin sich die Population der μ Eltern im Suchraum S bewegt.

Bei der Plus-Strategie bleibt somit die beste je gefundene Lösung auf jeden Fall in der Population erhalten und wird in die nachfolgende Generation übernommen. Somit wird die beste Lösung von Generation zu Generation nie schlechter. Allerdings birgt die Plus-Strategie die Gefahr, dass das Verfahren schnell gegen ein lokales Optimum konvergiert. Eine gemäß der Fitness sehr gute Lösung kann die gesamte Population dominieren. Liegt diese sehr gute Lösung in der Nähe eines lokalen Optimums, dann konvergiert mit der Zeit die gesamte Population gegen dieses lokale Optimum. Die Varianz der Gene innerhalb der Population geht verloren und der Suchraum wird nur noch an einer lokalen Stelle untersucht. Das globale Optimum kann dann nicht mehr gefunden werden. Diese Gefahr ist bei der Komma-Strategie geringer. Sie hat aber den Nachteil, dass die beste je gefundene Lösung verloren gehen kann. Eine ausführliche Darstellung dieser Problematik ist in [SHF94] zu finden. Strategien zur Festlegung von μ und λ sind in [Sch87] zu finden.

Eine typischerweise bei Genetischen Algorithmen übliche Selektionsmethode ist die sogenannte *Roulette-Wheel-Selektion* (auch *Proportionale Selektion* genannt) [Spr95, Spr99]. Hierbei geschieht die Auswahl der Individuen der Menge \mathcal{P}_{j+1} zufallsabhängig, wobei die Selektionswahrscheinlichkeit für ein Individuum umso größer ist, je besser der ihm zugeordnete Fitnesswert ist. Diese Strategie kann in geeignet modifizierter Weise auch bei Evolutionsstrategien eingesetzt werden.

Bei *multikriteriellen* Optimierungsproblemen, wo mehrere, meist gegenläufige Zielfunktion synchron optimiert werden, besteht die wesentliche Schwierigkeit der Selektion darin, eine in der Regel *kontinuierliche* Menge von Vektoren \mathbf{x} , die Pareto-Menge \mathcal{X} , durch eine *endliche* Anzahl von Individuen möglichst gleichmäßig abzudecken. Bei allen bisher bekannten Varianten der Evolutionsstrategien liegen die Defizite darin, dass die Individuen auf einen oder mehrere ausgezeichnete Punkte der Pareto-Menge zustreben, aber keine ausreichend gleichmäßige Abdeckung der gesamten Pareto-Menge erreicht wird [SV01]. Verschiedene aus der Literatur bekannte Vorgehensweisen bei multikriteriellen Optimierungsproblemen sind in [Hor97] übersichtlich zusammengefasst.

3.5.3 Abbruchkriterium

Der Abbruch des iterativen evolutionären Optimierungsprozesses liefert die Lösungspopulation \mathcal{P}_{end} , wovon das Individuum $\mathbf{Ind}_{end}^{best}$ mit dem besten Fitnesswert die beste gefundene Lösung des Optimierungsproblems darstellt. Der Optimierungsprozess wird beendet, wenn erstmalig mindestens eine der folgenden Abbruchbedingungen erfüllt ist:

- Eine vom Anwender vorgegebene maximale Generationenzahl j_{max} wurde erreicht.
- Eine vom Anwender vorgegebene maximale Anzahl an Fitnessbewertungen wurde durchgeführt. Insbesondere bei der Optimierung an realen Prozessen oder rechenintensiven Simulationen kosten diese Gütebewertungen in vielen praktischen Anwendungsfällen viel Zeit. In vielen Fällen kann von der Generationenzahl ausgehend recht einfach auf die Anzahl an Gütebewertungen umgerechnet werden. Dies ist jedoch nicht immer möglich. Manchmal fehlen hierfür notwendige Angaben oder es wurde nicht mit einer konstanten Anzahl an Gütebewertungen pro Generation gearbeitet. Deshalb kann nur anhand der Angabe dieser Anzahl ein fairer Vergleich zwischen unterschiedlichen Optimierungsverfahren oder verschiedenen Strategievarianten vorgenommen werden.
- Bei der Lösung skalarer Optimierungsprobleme kann ein Schwellwert für die Fitnesswerte angegeben werden. Wird dieser innerhalb des Optimierungsprozesses erreicht, so kann das Optimierungsproblem als gelöst angesehen und der Optimierungsprozess abgebrochen werden. Liegt hingegen ein multikriterielles Optimierungsproblem vor, so ist das Auffinden der gesamten Pareto-Menge – bzw. deren möglichst genaue Approximation – Ziel des Optimierungsprozesses. Da diese Pareto-Menge wohl kaum vorab bekannt ist, macht ein Abbruchkriterium, das auf einem Schwellwert für das Gütemaß basiert, bei multikriteriellen Optimierungsproblemen keinen Sinn.
- In praktischen Anwendungsfällen ist neben der Lösung des Optimierungsproblems auch der zugehörige Fitnesswert völlig unbekannt. In diesen Fällen bietet es sich an, für die letzten k Populationen \mathcal{P}_j jeweils den mittleren Fitnesswert \bar{f} über alle Individuen \mathbf{Ind}_i^j zu ermitteln. Hat dieser Mittelwert sich nicht um einen vorgegebenen Schwellwert ρ verbessert, kann analog zum vorherigen Fall abgebrochen werden.

Kapitel 4

Selektion von Sätzen relevanter nichtredundanter Merkmale und effiziente Fuzzy-Partitionierung

4.1 Problemstellung

Zur Präzisierung der Problemstellung diene die Aufgabenstellung der Modellierung eines Prozesses mit einer Menge \mathcal{X} von n bekannten potenziell relevanten Einflussgrößen X_i

$$\mathcal{X} = \{X_1, X_2, \dots, X_n\} \quad (4.1)$$

und einer Ausgangsgröße Y . Durch Prozessbeobachtung wird für einen datenbasierten Modellierungsansatz eine Datenmenge \mathcal{D} , bestehend aus N Datenpunkten \mathbf{d}_s , akquiriert

$$\mathcal{D} = \{\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_N\}. \quad (4.2)$$

Jeder Datenpunkt \mathbf{d}_j umfasst dabei die zugehörigen Werte $x_{i,j}$ aller Einflussgrößen X_i sowie den dazugehörigen Wert y_j der Ausgangsgröße Y

$$\mathbf{d}_s^T = (\mathbf{x}_j^T \quad y_j), \quad \mathbf{x}_j^T = (x_{1,j} \dots x_{n,j}). \quad (4.3)$$

Ferner werden alle N erhaltenen Ausgangsgrößenwerte y_s in der Menge \mathcal{Y} zusammengefasst

$$\mathcal{Y} = \{y_1, y_2, \dots, y_N\}. \quad (4.4)$$

Die dieser Arbeit zugrunde liegende Problemstellung besteht nun in der Selektion eines Satzes

$$\mathcal{M} \subseteq \mathcal{X}, \quad (4.5)$$

von q möglichst relevanten und nichtredundanten Einflussgrößen X_i . Um zu kennzeichnen, dass ein solcher Satz aus gerade q Einflussgrößen besteht, wird er

im Folgenden mit \mathcal{M}_q bezeichnet. Dabei soll für die Relevanzbewertung insbesondere auch das Zusammenspiel der zu bewertenden Einflussgrößen $X_i \in \mathcal{M}_q$ zu berücksichtig werden. Die Anzahl q der selektierten Einflussgrößen sollte dabei so groß wie nötig und so klein wie möglich sein, um einen handhabbaren Suchraum für eine nachgeschaltete Fuzzy-Modellierung bereitzustellen. Zusätzlich soll bei der Relevanzbewertung die Partitionierung des zugehörigen Merkmalsraumes berücksichtigt oder zusätzlich optimiert werden können. Hierdurch wird die Grundlage zur Generierung eines Fuzzy-Systems mit kompakter Regelbasis und hoher Modellierungsgüte gelegt.

4.2 Grundidee der rekonstruktionsbasierten Einflussgrößenselektion

Den im Folgenden vorgestellten Verfahren zur Relevanzbewertung liegt die Idee der Rekonstruktion eines Datenpunktes auf der Basis *aller* anderen Datenpunkte zu Grunde [Xio01]. Das in Abbildung 4.1 schematisch dargestellte Konzept kann durch eine entsprechende Wahl der Bewertungsfunktion für eine Relevanzanalyse potenzieller Einflussgrößen sowohl von Approximationsproblemen als auch von Klassifikationsproblemen angewendet werden. Den wesentlichen Kern des hier vorgeschlagenen Konzepts zur gesamtheitlichen Relevanzbewertung von Merkmalsätzen bildet ein Rekonstruktionsalgorithmus, der für jeden Datenpunkt \mathbf{d}_j , $j = 1, \dots, N$ mit zugehörigem Eingangsvektor \mathbf{x}_j einen Ausgangsgrößenwert \hat{y}_j auf der Basis der übrigen Datenpunkte $\tilde{\mathbf{d}}_j$ schätzt. Aus den tatsächlichen und rekonstruierten Ausgangsgrößenwerten y_j bzw. \hat{y}_j wird der Rekonstruktionsfehler berechnet. Je kleiner der Rekonstruktionsfehler ist, für desto relevanter wird der zugrunde liegende Satz von Einflussgrößen angesehen. Einflussgrößen, die bei Entfernung aus dem Satz den Rekonstruktionsfehler nicht oder nicht nennenswert vergrößern, werden als redundant angesehen.

Ein mit dem Rekonstruktionsalgorithmus gekoppelter Minimierungsalgorithmus ermittelt denjenigen Satz von Einflussgrößen, der zu dem geringsten Rekonstruktionsfehler führt oder einen guten Kompromiss zwischen den Zielen „möglichst geringer Rekonstruktionsfehler“ und „möglichst geringe Anzahl von Einflussgrößen“ herstellt. Um alle vorhandenen Einflussgrößen X_i gleichwertig zu berücksichtigen, werden die zugehörigen Einflussgrößenwerte $x_{i,j}$ geeignet normiert. Eine entsprechende Normierung der Ausgangsgrößenwerte y_j ist nicht erforderlich.

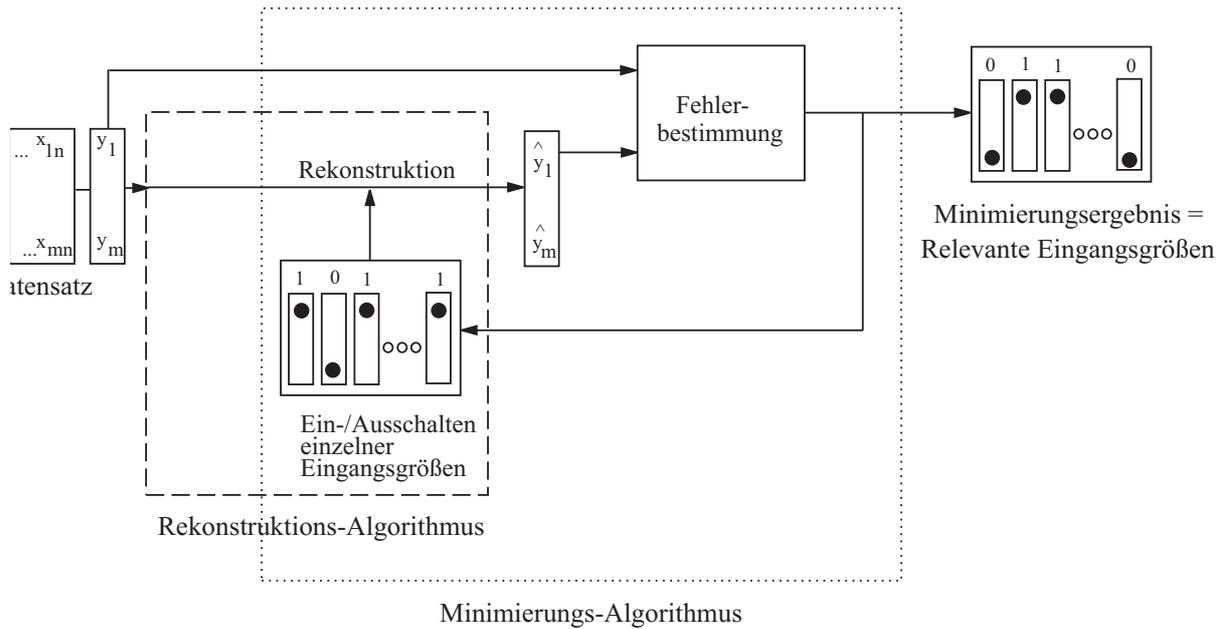


Abbildung 4.1: Konzept zur rekonstruktionsbasierten Selektion relevanter nichtredunder Einflussgrößen.

4.3 Normierung der Einflussgrößen

4.3.1 Kontinuierliche Einflussgrößen

Die im Folgenden vorgestellten Rekonstruktionsverfahren für kontinuierliche Einflussgrößen verwenden zur Rekonstruktion Datenpunkte, die innerhalb lokaler durch Nachbarschaftsfunktionen definierte Gebiete des Merkmalsraumes M liegen. Typischerweise unterscheiden sich die Wertebereiche kontinuierlicher Einflussgrößen realer Anwendungsprobleme meist sehr stark. Beispielsweise kann der Wertebereich einer Einflussgröße im Intervall $[0, 1]$ und der einer anderen im Intervall $[0, 100]$ liegen. Um zu vermeiden, dass derartige Unterschiede zu einer unerwünschten ungleichen Behandlung der Einflussgrößen bei der Merkmalsselektion führt, werden die Werte $x_{i,j}$ jeder Einflussgröße X_i gemäß

$$\tilde{x}_{i,j} = \frac{x_{i,j} - x_i^{\min}}{\delta_i} \quad (4.6)$$

so normiert, dass die auftretenden Werte der komponentenweisen Abstände zweier Eingangsvektoren \mathbf{x}_a und \mathbf{x}_b ganz oder zumindest überwiegend im Intervall $[0, 1]$ liegen. In Gleichung 4.6 bezeichnet x_i^{\min} den minimalen Wert der Einflussgröße X_i und δ_i einen zu wählenden Normierungsfaktor. Für die Wahl von δ_i werden statistische Kennzahlen gemäß Tabelle 4.1 berücksichtigt, die die Lerndaten charakterisieren.

Tabelle 4.1: Zur Normierung verwendete statistische Kennzahlen einer Einflussgröße X_i .

Bezeichnung	Definition
Minimalwert	$x_i^{min} := \min_{j=1}^N \{x_{i,j}\}$
Maximalwert	$x_i^{max} := \max_{j=1}^N \{x_{i,j}\}$
Spannweite	$x_i^{range} := x_i^{max} - x_i^{min}$
Arithmetischer Mittelwert	$\bar{x}_i := \frac{1}{N} \sum_{j=1}^N x_{i,j}$
Standardabweichung	$\sigma_i := \frac{1}{N-1} \sum_{j=1}^N (x_{i,j} - \bar{x}_i)^2$

Vielfach anwendbar ist bereits der einfache Ansatz

$$\delta_i = x_i^{range} . \quad (4.7)$$

Bei einer derart parametrisierten Normierung der Daten nach Gleichung 4.6 liegen die möglichen komponentenweise Abstände zweier Lerndatenpunkte $x_{i,a}$ und $x_{i,b}$

$$\Delta(x_{i,a}, x_{i,b}) = |x_{i,a} - x_{i,b}| \quad (4.8)$$

stets im Intervall $[0, 1]$. Dagegen können sich mit dieser an den Lerndaten orientierten Normierung für die Testdaten unter Umständen auch komponentenweise Abstände $d_i > 1$ ergeben, was aber keine störende Auswirkungen hat. Die Normierungsvorschrift (4.7) hat jedoch den Nachteil, dass eventuelle Ausreißer, wie sie beispielsweise durch Messfehler entstehen können, das Normierungsergebnis unerwünscht stark tangieren. Bei verrauschten Daten ist es daher im Interesse einer großen Robustheit sinnvoll, zur Ermittlung der Spannweite x_i^{range} einen gewissen Prozentsatz von beispielsweise $\alpha = 5\%$ der kleinsten und größten Werte zu ignorieren, um so robustere Kennzahlen zu erhalten¹. Alternativ ist es bei stark verrauschten Daten sinnvoll in Gleichung 4.6

$$\delta_i = 4\sigma_i \quad (4.9)$$

zu wählen, wobei σ_i die Standardabweichung ist. Dies liegt darin begründet, dass bei normalverteilten Werten $x_{i,j}$ einer Größe X_i insgesamt 95% aller Werte innerhalb des Intervalls $[-2\sigma_i, +2\sigma_i]$ liegen, so dass bei dieser Normierung der Abstand $\Delta(x_{i,a}, x_{i,b})$ zweier zufällig ermittelter Werte $x_{i,a}$ und $x_{i,b}$ meist kleiner als 1 sein wird.

¹Dies entspricht der Verwendung des Abstandes des $(1 - \alpha)$ -Quantils zum α -Quantil der betrachteten Einflussgröße.

4.3.2 Gemischt kontinuierliche und nominale Einflussgrößen

Bei Vorliegen sowohl kontinuierlicher als auch nominaler Einflussgrößen liegen nach der obigen Normierung die komponentenweisen Abstände zwar ganz oder überwiegend innerhalb des Intervalls $[0, 1]$, jedoch ist der arithmetische Mittelwert dieser Abstände meist wesentlich größer als derjenige, der sich bei nur kontinuierlicher Eingangsgrößen ergibt. Dies verletzt bei der in Abschnitt 4.4 beschriebenen Rekonstruktion die angestrebte gleichgewichtete Berücksichtigung kontinuierlicher und nominaler Einflussgrößen.

Eine einfache Möglichkeit zur Vermeidung dieser Nachteile besteht darin, den Normierungsfaktor δ_i für kontinuierliche Eingangsgrößen entweder gemäß Gleichung 4.7 an die Spannweite, oder gemäß Gleichung (4.9) an die Standardabweichung zu koppeln, während davon abweichend für nominale Eingangsgrößen als Normierungsfaktor δ_{Symb} beispielsweise die Anzahl der vorhandenen Symbole gewählt wird.

Eine aufwändigere aber leistungsfähigere Möglichkeit zur Überwindung des oben beschriebenen Nachteils besteht darin, die Normierungsfaktoren gleichermaßen für nominale als auch für kontinuierliche Eingangsgrößen auf Basis eines ermittelten durchschnittlichen komponentenweisen Abstands Δ_{avr} zu ermitteln [HC99]. Die Idee dieses Ansatzes besteht darin, die komponentenweisen Abstände jeder Eingangsgröße derart zu normieren, dass sich im arithmetischen Mittel der Wert 1 ergibt. Dazu wird anhand der Datenpunkte der jeweils betrachteten Lernstichprobe für jede Einflussgröße der durchschnittliche komponentenweise Abstand zweier Datenpunkte zueinander bestimmt. In Gleichung (4.10) wird der durchschnittliche komponentenweise Abstand der Einflussgröße X_i als arithmetischer Mittelwert über den komponentenweisen Abstand aller möglichen Kombinationen zweier Eingangsvektoren gemäß

$$\delta_i = \Delta_{avr} = \frac{\sum_{j=1}^N \sum_{z=j+1}^N \Delta_i(x_{i,j}, x_{i,z})}{\frac{1}{2}N(N-1)} \quad (4.10)$$

ermittelt. Der Rechenaufwand steigt dabei quadratisch mit der Anzahl der Lerndatenpunkte an. Zur Verminderung des Rechenaufwands kann daher die Betrachtung nur einer Teilmenge von Datenpunkte sinnvoll sein.

4.4 Abstandsbasierte Ausgangsgrößenwertschätzung

Nach der in 4.2 erläuterten Grundidee stützt sich die Relevanzbewertung eines Satzes von Einflussgrößen auf einen Rekonstruktionsalgorithmus, der für jeden Datenpunkt $\mathbf{d}_a^T = (\mathbf{x}_a^T, y_a)$, im Folgenden auch Aufpunkt genannt, den Wert y_a auf der Basis der übrigen Datenpunkte \mathbf{d}_i schätzt. Dabei sollten diese Punkte \mathbf{d}_i mit umso größerem Gewicht in den gesuchten Schätzwert \hat{y}_a einfließen, je geringer ihr Abstand, gemessen im betrachteten Merkmalsraum \mathcal{M} , zum Aufpunkt ist.

4.4.1 Abstandsbasierte Zugehörigkeitsfunktionen

Eine naheliegende Möglichkeit zur Umsetzung dieses Konzepts basiert auf der Einführung von abstandsbasierten Zugehörigkeitsfunktionen:

Definition 4.1 *Abstandsbasierte Zugehörigkeitsfunktion $\mathcal{N}_{\mathcal{M}}(i, a)$: Eine Funktion, die jedem Datenpunkt \mathbf{d}_i einen Funktionswert $\mathcal{N}_{\mathcal{M}}(i, a) \in [0, 1]$ zuordnet, der umso größer ist, je geringer der Abstand zwischen dem Datenpunkt \mathbf{d}_i und dem Aufpunkt \mathbf{d}_a , gemessen im betrachteten Merkmalsraum \mathcal{M} , ist.*

Speziell verwendet werden hierzu in Anlehnung an [SKMM03] Zugehörigkeitsfunktionen der Form

$$\mathcal{N}_{\mathcal{M}}(i, a) = e^{-\lambda \cdot \Delta_{\mathcal{M}}(i, a)} . \quad (4.11)$$

Darin ist λ ein wählbarer Anpassungsparameter und $\Delta_{\mathcal{M}}(i, a)$ der Abstand zwischen den Punkten \mathbf{d}_i und \mathbf{d}_a gemessen im p -dimensionalen Merkmalsraum \mathcal{M} (deshalb die Kennzeichnung mit dem Index \mathcal{M}). Zur Bestimmung dieses Abstands wird die *Minkowski-Metrik* mit

$$\Delta_{\mathcal{M}}(i, a) = \sqrt[p]{\sum_{j=1}^p d_{i,a,j}^p} \quad (4.12)$$

verwendet. Die darin vorkommenden insgesamt p komponentenweisen Abstände werden gemäß

$$d_{i,a,j} = |x_{i,j} - x_{a,j}| \quad (4.13)$$

bestimmt. Dabei bezeichnen $x_{i,j}$ und $x_{a,j}$ die jeweilige j -te Komponente des Eingangsvektors \mathbf{x}_i bzw. \mathbf{x}_a .

Die Minkowski-Metrik (4.12) enthält in der Wahl des Parameters ρ noch einen Freiheitsgrad. Die gebräuchlichsten Werte für ρ führen auf die in Tabelle 4.2 erläuterten unterschiedlichen Abstandsmaße.

Tabelle 4.2: Übliche Wahl des Parameters ρ bei der Minkowski-Metrik.

Komponentenweiser Abstand	$\rho = 1$	$\Delta_{\mathcal{M}}(i, a) = \sum_{j=1}^p d_{i,a,j}$
Euklidischer Abstand	$\rho = 2$	$\Delta_{\mathcal{M}}(i, a) = \sqrt{\sum_{j=1}^p d_{i,a,j}^2}$
Chebyshev-Abstand	$\rho = \infty$	$\Delta_{\mathcal{M}}(i, a) = \max_{j=1}^p d_{i,a,j} $

4.4.2 Rekonstruktionsvorschriften

Die oben eingeführten Zugehörigkeitsfunktionen $\mathcal{N}_{\mathcal{M}}(i, a)$ stellt die Schlüssel-funktion zur Rekonstruktion (Schätzung) des Ausgangsgrößenwertes y_a (zugehörig zum Aufpunkt $\mathbf{d}_a^T = (\mathbf{x}_a^T, y_a)$) auf Basis aller anderen Datenpunkte \mathbf{d}_i dar. Zur Ermittlung eines günstigen Schätzwertes \hat{y}_j , der im Folgenden auch mit $\mathcal{E}(a)$ bezeichnet wird, sind unterschiedliche Verfahren anzuwenden, je nachdem ob es sich um ein Problem der *Klassifikation* oder *Approximation* handelt. Bei einer *Klassifikation* wird dem Eingangsvektor \mathbf{x}_a ein Schätzwert \hat{y}_a zugeordnet, der einem der in den Daten auftretenden η diskreten Ausgangsgrößenwerte $\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_\eta$ entspricht. Hierzu werden alle erhobenen Datenpunkte $\tilde{\mathbf{d}}_i$ in Klassen $\mathcal{K}_1, \mathcal{K}_2, \dots, \mathcal{K}_\eta$ derart eingeteilt, dass alle Daten einer Klasse \mathcal{K}_t den gleichen Ausgangsgrößenwert \tilde{y}_t aufweisen. Zur Bestimmung des Rekonstruktionswertes \hat{y}_j für einen Aufpunkt \mathbf{d}_a wird zunächst die gewichtete mittlere Häufigkeit

$$H_a(\tilde{y}_t) = \frac{\sum \mathcal{N}_{\mathcal{M}}(i, j)}{\text{card}(\mathcal{K}_t)} \quad \forall j \text{ mit } \tilde{\mathbf{d}}_i \in \mathcal{K}_t \quad (4.14)$$

berechnet. Der Schätzwert \hat{y}_a ergibt sich schließlich durch Zuweisung desjenigen Klassenwertes \tilde{y}_t als Rekonstruktionswert $\mathcal{E}(a)$, für den die gewichtete mittlere Häufigkeit $H_a(y_t)$ maximal wird:

$$\hat{y}_a = \mathcal{E}(a) = \{\tilde{y}_t | H_a(\tilde{y}_t) = \max\{H_a(\tilde{y}_1), H_a(\tilde{y}_2), \dots, H_a(\tilde{y}_\eta)\}\}. \quad (4.15)$$

Demgegenüber wird im Falle einer *Approximation* die Rekonstruktion eines reellwertigen Ausgangsgrößenwertes anhand einer gewichteten Mittelwertbildung vollzogen:

$$\mathcal{E}(a) = \frac{\sum_{j \neq a} \mathcal{N}_{\mathcal{M}}(j, a) \cdot y_j}{\sum_{j \neq a} \mathcal{N}_{\mathcal{M}}(j, a)}. \quad (4.16)$$

4.4.3 Anwendungsgrenzen der abstandsbasierter Rekonstruktion

In [SKMM03] wird die prinzipielle Tauglichkeit der abstandsbasierter Rekonstruktion zur Relevanzbewertung von Merkmalsätzen anhand eines synthetischen Beispiels sowie durch Ergebnisse bei Anwendung auf einschlägigen Benchmarkproblemen nachgewiesen. Allerdings hat diese Vorgehensweise auch Anwendungsgrenzen. Sie liegen vor allem in dem relativ hohen Rechenaufwand: Für jede der N erforderlichen Rekonstruktionen eines Ausgangsgrößenwertes \hat{y}_a muss jeweils die Zugehörigkeit $\mathcal{N}_{\mathcal{M}}(i, a)$ aller anderen $(N - 1)$ Datenpunkte \mathbf{d}_i ermittelt werden. Die daraus resultierende Rechenzeit macht eine Analyse von Datensätzen mit mehr als 10000 Datenpunkten, unabhängig von der Anzahl n der vorhandenen Einflussgrößen X_i , unpraktikabel. Ferner ist zu berücksichtigen, dass die Relevanzbewertung hier im Hinblick auf eine nachgeschaltete Fuzzy-Modellierung vorgenommen werden soll. Ein wesentlicher Nachteil der abstandsbasierter Rekonstruktion besteht darin, dass der zelluläre Charakter von Fuzzy-Modellen damit nicht von vornherein berücksichtigt werden kann. Beide Nachteile werden im Folgenden abgestellt.

4.5 Zellenbasierte Ausgangsgrößenwertschätzung

Wie oben erwähnt, ist der Wrapper-Ansatz, d. h. die Verwendung des später zur eigentlichen Modellierung verwendeten Verfahrens auch zur Relevanzbewertung, an sich die Methode der ersten Wahl. Da eine nachgeschaltete Fuzzy-Modellierung vorgesehen ist, muss dieser Ansatz aus Aufwandsgründen ausscheiden und stattdessen ein Filteransatz verfolgt werden. Der oben beschriebene Filteransatz, der auf abstandsbasierter Rekonstruktion basiert, hat die erläuterten Nachteile. Diese werden im Folgenden dadurch minimiert, dass ein Filteransatz gewählt wird, der „näher“ am Wrapper-Ansatz liegt: Es wird dabei nämlich für die Relevanzbewertung dieselbe Partitionierung verwendet, wie sie auch im späteren Fuzzy-Modell vorgesehen wird.

4.5.1 Partitionierung des Merkmalsraumes

Die Granularisierung einer Einflussgröße X_j wird durch einen Partitionierungsvektor \mathbf{s}_j beschrieben, dessen k Komponenten die Stützstellen zur Festlegung der Partitionierung definieren:

$$\mathbf{s}_j^T = (s_{j,1}, s_{j,2}, \dots, s_{j,k}) . \quad (4.17)$$

Diese Stützstellen teilen den Wertebereich jeder Größe X_j in $k - 1$ Intervalle ein (Abbildung 4.2). Die Anzahl der zu verwendenden Intervalle hängt dabei von der Art des Modellierungsproblems ab. Zum Beispiel sollte die Anzahl der Intervalle je Einflussgröße bei Klassifikationsproblemen mindestens der Anzahl η der in den Daten vorhandenen verschiedenen Klassen \mathcal{K}_i entsprechen. Die Verwendung einer größeren Anzahl von Intervallen führt zu einer feineren Partitionierung und damit erhöhten Auflösung des Eingangsraums, auf Grundlage dessen möglicherweise auch exaktere Klassifikatoren entworfen werden können. Demgegenüber kann jedoch eine zu groß gewählte Anzahl von Intervallen neben einer geminderten Transparenz auch zum Overfitting führen. In diesem Fall liefern die anschließend generierten Klassifikatoren zwar für die verwendeten Lerndaten sehr gute, aber für die nicht verwendeten Validierungsdaten unbefriedigende Ergebnisse.

Die Menge der Partitionierungsvektoren $\mathcal{S}_{\mathcal{M}}$ für alle m Einflussgrößen eines potenziellen Merkmalsatzes \mathcal{M}

$$\mathcal{S}_{\mathcal{M}} = \{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_m\} \quad (4.18)$$

induzieren eine Zerlegung des Merkmalsraumes in eine Menge von $q = (k - 1)^m$ Hyperquadern Q_h :

$$\mathcal{Q}_{\mathcal{M}} = \{Q_1, Q_2, \dots, Q_q\} . \quad (4.19)$$

4.5.2 Partialfunktionen

Ausgehend von einer solchen Partitionierung werden Partialfunktionen $\mathcal{E}_h(\mathbf{x})$ definiert, deren Funktionswerte außerhalb des Quaders Q_h Null sind und innerhalb des Quaders nur von den darin liegenden Datenpunkten $\mathbf{d}_i^T = (\mathbf{x}_i^T, y_i)$ abhängen. Die Anzahl der in Q_h gelegenen Datenpunkte wird im Folgenden mit $m(h)$ bezeichnet.

Klassifikation

Für das Anwendungsproblem einer *Klassifikation* werden die Partialfunktionen $\mathcal{E}_h(\mathbf{x})$ so festgelegt, dass sie den am häufigsten vertretenen Ausgangsgrößenwert im Hyperquader Q_h ausgeben. Hierzu werden alle diskreten Ausgangsgrößenwerte der in Q_h liegenden Datenpunkte derart in Klassen $\mathcal{K}_1, \mathcal{K}_2, \dots, \mathcal{K}_\eta$ eingeteilt, dass jede Klasse nur Punkte aus Q_h mit dem gleichen Ausgangsgrößenwert enthält.

Aus den daraus abgeleiteten Häufigkeiten

$$k_t = \text{card}(\mathcal{K}_t) \quad \forall t \in \{1, 2, \dots, \eta\} \quad (4.20)$$

wird der am häufigsten auftretende Ausgangsgrößenwert als Funktionswert der Partialfunktion im Quader Q_h festgelegt (Abbildung 4.2a):

$$\mathcal{E}_h(\mathbf{x}) = \{\tilde{y}_t | k_t = \max\{k_1, k_2, \dots, k_\eta\}\}. \quad (4.21)$$

Approximation

Bei dem Anwendungsproblem einer *Approximation* kann die nachgeschaltete Fuzzy-Modellierung mit Fuzzy-Systemen vom Typ *Mamdani* (Abschnitt 3.2.1) und vom Typ *Takagi-Sugeno* (Abschnitt 3.2.2) erfolgen. Beide Fälle werden bei der Definition der Partialfunktionen in adäquater Weise berücksichtigt.

Da ein Mamdani-Fuzzy-System durch die Konklusion einer Regel konstant einen kleinen Wertebereich der Ausgangsgröße empfiehlt, wird der Mittelwert aller Ausgangsgrößenwerte, die zu den Datenpunkten aus Q_i gehören, als konstanter Funktionswert

$$\mathcal{E}_h(\mathbf{x}) = \frac{1}{m(h)} \cdot \sum y_i \quad ; \quad \mathbf{x}_i \in Q_i \quad (4.22)$$

der Partialfunktion festgelegt (Abbildung 4.2b).

Soll ein Approximationsproblem mit einem Fuzzy-System vom Typ *Takagi-Sugeno-Kang* (TSK) gelöst werden, so wird im Hyperquader Q_i die Partialfunktion $\mathcal{E}_h(\mathbf{x})$ in Form einer linearen Funktion

$$\mathcal{E}_h(\mathbf{x}) = c_{h,0} + \sum c_{h,j} \cdot x_j \quad ; \quad \mathbf{x}_i \in Q_i \quad (4.23)$$

definiert (Abbildung 4.2c).

Dabei ergeben sich die Koeffizienten $c_{h,0}$ und $c_{h,j}$, aus einer *Least-Square-Approximation* [PTVF99], die innerhalb des Hyperquaders Q_h zu einem minimalen quadratischen Fehler

$$\sum_{j=1}^{m(h)} (y_j - \mathcal{E}_h(\mathbf{x}_j))^2 \quad (4.24)$$

führt. Unabhängig vom gewählten Typ der Partialfunktion wird die gesuchte Schätzfunktion im vollständigen Merkmalsraum \mathcal{M} durch

$$\mathcal{E}(\mathbf{x}) = \mathcal{E}_a(\mathbf{x}) \quad \text{für } \mathbf{x} \in Q_a \quad (4.25)$$

erklärt.

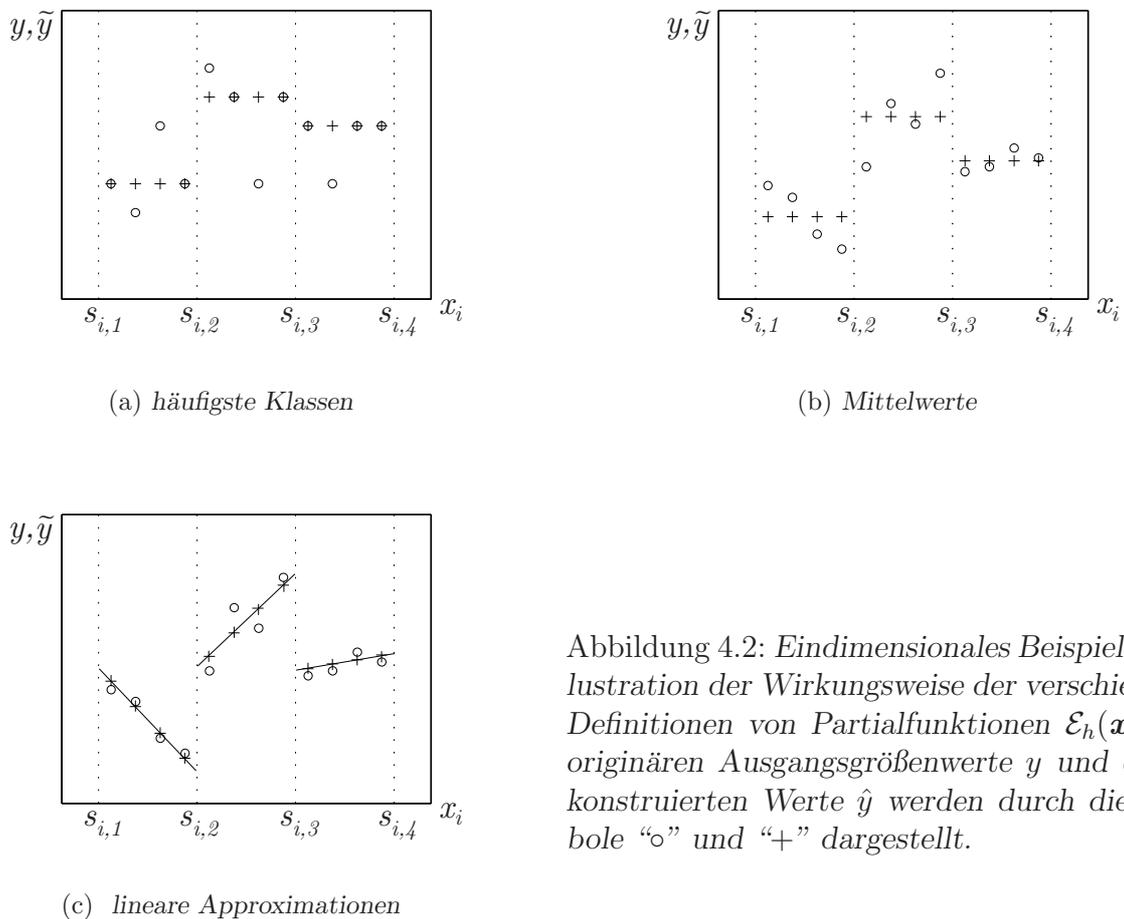


Abbildung 4.2: Eindimensionales Beispiel zur Illustration der Wirkungsweise der verschiedenen Definitionen von Partialfunktionen $\mathcal{E}_h(\mathbf{x})$. Die originären Ausgangsgrößenwerte y und die rekonstruierten Werte \hat{y} werden durch die Symbole “o” und “+” dargestellt.

4.6 Relevanzbewertung eines Merkmalsatzes

Wie einleitend in 4.2 als Grundidee skizziert, spiegelt der Grad der Übereinstimmung zwischen den rekonstruierten Ausgangsgrößenwerten \hat{y}_j und den tatsächli-

chen Ausgangsgrößenwerten y_j die Relevanz des zu bewertenden Merkmalsatzes \mathcal{M} wider. Im Folgenden werden anwendungsspezifische Präzisierungen dieses Schätzfehlers beschrieben. Die dazu definierten Funktionen dienen zudem auch als Zielfunktion für eine Optimierung der Partitionierung des Merkmalsraumes durch evolutionäre Variation der Partitionierungsvektoren $\mathcal{S}_{\mathcal{M}}$.

Klassifikation

In Falle eines Klassifikationsproblems wird der relative Klassifikationsfehler

$$\text{RCE}(\mathcal{S}_{\mathcal{M}}) = \frac{1}{N} \sum_{i=1}^N \text{sgn} |y_i - \mathcal{E}(\mathbf{x}_i)| \quad (4.26)$$

über alle N Datenpunkte $\mathbf{d}_i^T = (\mathbf{x}_i^T, y_i)$ des Datensatzes \mathcal{D} als Relevanzmaß $\eta_{\mathcal{M}}$ für den betrachteten Merkmalsatzes \mathcal{M} verwendet. Ersichtlich bedeutet der Wert $\text{RCE}(\mathcal{S}_{\mathcal{M}}) = 0$, dass stets richtig klassifiziert wird, was einer größtmöglichen Relevanz entspricht und auf einen optimal partitionierten Merkmalsatzes \mathcal{M} hinweist.

Approximation

Für den Fall eines Approximationsproblems wird der empirische Korrelationsfehler

$$\text{ECE}(\mathcal{S}_{\mathcal{M}}) = 1 - \left| \frac{\sum_{i=1}^N (y_i - \bar{y}) \sum_{i=1}^N (\mathcal{E}(\mathbf{x}_i) - \hat{y})}{\sqrt{\sum_{i=1}^N (y_i - \bar{y})^2 \sum_{i=1}^N (\mathcal{E}(\mathbf{x}_i) - \hat{y})^2}} \right| \quad (4.27)$$

mit

$$\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i \quad \text{und} \quad \hat{y} = \frac{1}{N} \sum_{i=1}^N \mathcal{E}(\mathbf{x}_i) \quad (4.28)$$

als Relevanzmaß $\eta_{\mathcal{M}}$ des betrachteten Merkmalsatzes \mathcal{M} verwendet. Auch hier entspricht der Wert $\text{ECE}(\mathcal{S}_{\mathcal{M}}) = 0$ der größtmöglichen Relevanz. Der empirische Korrelationsfehler beinhaltet neben der reinen Bewertung des Rekonstruktionsfehlers, wie z. B. bei der Betrachtung des mittleren quadratischen Fehlers

$$\text{MSE}(\mathcal{S}_{\mathcal{M}}) = \frac{1}{N} \sum_{i=1}^N (y_i - \mathcal{E}(\mathbf{x}_i))^2 \quad (4.29)$$

oder des mittleren absoluten Fehlers

$$\text{MAE}(\mathcal{S}_{\mathcal{M}}) = \left| \frac{1}{N} \sum_{i=1}^N (y_i - \mathcal{E}(\mathbf{x}_i)) \right| \quad (4.30)$$

auch eine Bewertung der Ähnlichkeit im Sinne der Korrelation der Zeitreihe der rekonstruierten Ausgangsgrößenwerte $\hat{\mathcal{Y}}$ mit der Zeitreihe der tatsächlichen Ausgangsgrößenwerte \mathcal{Y} .

4.7 Optimierung der Partitionierung des Merkmalsraumes

Zur Ermittlung der maximal möglichen Relevanz eines Merkmalsssatzes \mathcal{M} werden die Elemente der zugehörigen Partitionierungsvektoren \mathbf{s}_i aus $\mathcal{S}_{\mathcal{M}}$ (Gleichung 4.18) im Sinne einer möglichst guten Partitionierung des Merkmalsraumes optimiert. Hierzu wird als zu optimierende Zielfunktion das verwendete Relevanzmaß $\eta_{\mathcal{M}}$ angesetzt, welches es zu minimieren gilt. Aufgrund der stückweisen Definition dieser Zielfunktion weist diese im Allgemeinen viele Unstetigkeiten, Plateaus und nichtlineare Charakteristiken auf. Daher wird als Optimierungsverfahren eine Evolutionsstrategie vorgesehen. Diese erlaubt zudem eine multikriterielle Optimierung, indem weitere Zielfunktionen, die unabhängig vom Relevanzmaß $\eta_{\mathcal{M}}$ die Partitionierung des Merkmalsraumes unter anderen Gesichtspunkten bewerten, synchron mitberücksichtigt werden.

4.7.1 Kodierung der Partitionierung

Sämtliche Komponenten aller Stützstellenvektoren \mathbf{s}_i der Menge $\mathcal{S}_{\mathcal{M}}$ des betrachteten Merkmalsssatzes \mathcal{M} werden als Individuum einer Evolutionsstrategie repräsentiert. Dabei werden folgende Aspekte besonders beachtet:

- Die Stützstellen $s_{i,j}$ können jeweils ausschließlich innerhalb der Spannweite x_j^{range} des Wertebereichs der Eingangsgröße X_j liegen.
- Die Stützstellen $s_{i,j}$ unterliegen innerhalb des Wertebereichs der Eingangsgröße X_j einer geordneten Rangfolge, die es aufrechtzuerhalten gilt.

Prinzipiell gibt es zwei mögliche Kodierungsarten zur Repräsentation einer Stützstellenkonfiguration \mathcal{S} .

Abstandskodierung

Zur garantierten Einhaltung der geforderten Stützstellenrelationen

$$s_{j,min} \leq s_{j,1} \leq s_{j,2} \leq \dots \leq s_{j,k_j^{max}} \leq s_{j,max} \quad (4.31)$$

kann eine Abstandskodierung derart definiert werden, dass für jede Dimension i die erste Stützstelle $s_{j,1}$ absolut und alle weiteren Stützstellen $s_{j,k}$ relativ in Form des Abstandes $\lambda_{j,k}$ zum linken Nachbarn $s_{j,k-1}$ erfolgt.

Hieraus resultiert das Optimierungsproblem zur Minimierung der anwendungsabhängig gewählten Zielfunktion, beispielsweise $ECE(\mathcal{S}_{\mathcal{M}})$ nach Gleichung 4.27 für ein Approximationsproblem unter Einhaltung der Nebenbedingungen

$$s_{j,1} + \sum_{k=2}^{k_j^{max}} (s_{j,k} - s_{j,k-1}) = x_j^{max} . \quad (4.32)$$

Die Optimierungsparameter — die sogenannten *Zielvariablen* — werden daher für den Einsatz der Evolutionsstrategie als Individuum

$$\mathbf{Ind} = \left(\underbrace{s_{1,1}, \lambda_1, s_{2,1}, \lambda_2, \dots, s_{m,1}, \lambda_m}_{\text{Zielvariablen}}, \underbrace{\sigma_1, \sigma_2, \dots, \sigma_m}_{\text{Strategieparameter}} \right) \quad (4.33)$$

mit

$$\lambda_{j,k} = s_{j,k} - s_{j,k-1} \quad \forall k \in \{2, \dots, k_j^{max}\} \wedge \forall j \in \{1, 2, \dots, m\} \quad (4.34)$$

und

$$\sigma_j = (\sigma_{j,1}, \sigma_{j,2}, \dots, \sigma_{j,k_j^{max}}) \quad \forall j \in \{1, 2, \dots, m\} \quad (4.35)$$

kodiert. Des Weiteren wird dem Individuum je Zielvariable ein Strategieparameter $\sigma_{j,k}$ hinzugefügt, der nach SCHWEFEL [Sch95] zur Schrittweitenadaption der Mutationsweiten im evolutionären Optimierungsprozesses dient (vergleiche Abschnitt 3.5).

Auch wenn mit dieser Repräsentation die geforderten Stützstellenrelationen garantiert eingehalten werden, so wirkt sich die hierzu eingeführte Verkopplung der Zielvariablen extrem ungünstig auf den Optimierungsprozess aus. Dies liegt darin begründet, dass eine Variation der ersten Zielvariablen $s_{j,1}$ sich auf alle weiteren Stützstellen der j -ten Dimension auswirkt. Auch eine Veränderung einer Komponente von λ_j wirkt sich auf alle weiteren „hinteren“ Stützstellen aus.

Stützstellenkodierung

Bei der Stützstellenkodierung werden in jedem Individuum als zu optimierende Zielvariablen die Stützstellen $s_{j,k}$ voneinander entkoppelt angesetzt:

$$\mathbf{Ind} = \left(\underbrace{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_m}_{\text{Zielvariablen}}, \underbrace{\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2, \dots, \boldsymbol{\sigma}_{m_\sigma}}_{\text{Strategieparameter}} \right) \quad (4.36)$$

mit

$$\mathbf{s}_j = (s_{j,1}, \dots, s_{j,k_j^{max}}) \wedge \boldsymbol{\sigma}_j = (\sigma_{j,1}, \dots, \sigma_{j,k_j^{max}}) \quad \forall j \in \{1, \dots, m\}. \quad (4.37)$$

Ein Nachteil der Stützstellenkodierung besteht allerdings darin, dass während der evolutionären Optimierung, insbesondere bei der Ausführung der Mutationen, sich die Stützstellen $s_{j,k}$ einer Dimension j möglicherweise „überholen“ und damit die vorgesehenen Relationen nach Gleichung 4.31 verletzt werden. In [Bri98] wird zur Überwindung dieses Problems ein relativ aufwändiger Ansatz zur Adaption der Strategieparameter $\boldsymbol{\sigma}_j$ an die jeweils aktuelle Konfiguration der Zielvariablen \mathbf{s}_j vorgestellt und die deutliche Überlegenheit dieser Stützstellenkodierung gegenüber der Abstandskodierung bezüglich der Effizienz des Optimierungsprozesses nachgewiesen.

Da die Einhaltung der Stützstellenrelationen in diesem Anwendungsfall rein rechen-technisch begründet und nicht beispielsweise zur Aufrechterhaltung semantischer Beziehungen damit verknüpfter Regeln erforderlich ist, wird im Rahmen dieser Arbeit ein weniger aufwändiger und dennoch effizienterer pragmatischer Ansatz eingeführt. Dieser besteht in der dimensionsweisen aufsteigenden Sortierung der Komponenten einer jeden Zielvariablen \mathbf{s}_j zu $\bar{\mathbf{s}}_j$

$$\overline{\mathbf{Ind}} = \text{sort}(\mathbf{Ind}) = (\bar{\mathbf{s}}_1, \bar{\mathbf{s}}_2, \dots, \bar{\mathbf{s}}_m, \bar{\boldsymbol{\sigma}}_1, \bar{\boldsymbol{\sigma}}_2, \dots, \bar{\boldsymbol{\sigma}}_{m_\sigma}), \quad (4.38)$$

wobei die Zuordnung eines Strategieparameters $\bar{\sigma}_{j,k}$ zu $\bar{s}_{j,k}$ analog der Zuordnung von $\sigma_{j,k}$ zu $s_{j,k}$ erhalten bleibt.

Um die Komplexität des evolutionären Optimierungsproblems nicht künstlich zu erhöhen, wird diese Sortierung als Bestandteil der Evolutionsstrategie derart implementiert, dass diese auf jedes neu erzeugte Individuum unverzüglich, also im Genotypraum, ausgeführt wird. Eine ausschließliche Anwendung der Sortierung im Phänotypraum zur Fitnessauswertung würde die Repräsentation einer bestimmten Stützstellenkonfiguration $\mathcal{S}_{\mathcal{M}}$ des betrachteten Merkmalsatzes \mathcal{M} durch die immens hohe Anzahl von

$$\prod_{j=1}^m (k_j^{max}!) \quad (4.39)$$

verschiedenen Repräsentationsformen ermöglichen und somit die Effizienz des Optimierungsprozesses beeinträchtigen.

4.7.2 Zusätzliche Zielfunktionen

Zusätzlich wird für die Optimierung einer Stützstellenkonfiguration $\mathcal{S}_{\mathcal{M}}$ ein weiterer Vorteil Evolutionärer Algorithmen ausgenutzt, indem eine multikriterielle Optimierung angestrebt wird. Damit ist es möglich, mehrere — auch tendenziell gegenläufige — Zielfunktionen gleichzeitig zu berücksichtigen und die zugehörige Paretomenge zu ermitteln. Dies wird hier dahingehend eingesetzt, dass neben der hauptsächlich interessierenden Relevanz weitere Eigenschaften wie z. B. die Robustheit oder Generalisierungsfähigkeit einer Stützstellenkonfiguration $\mathcal{S}_{\mathcal{M}}$ berücksichtigt wird. Dies ist insbesondere bei der Verwendung des global operierenden Selektionsprozesses (Abschnitt 4.8.2) von wichtiger Bedeutung. Damit wird dem Anwender die Möglichkeit in die Hand gegeben, unter einer Vielzahl möglicher Lösungen die für ihn geeignetste Kompromisslösung aus der gefundenen Paretomenge auszuwählen.

Merkmalsanzahl

Ein grundlegender eingangs motivierter Gütegesichtspunkt besteht in der Selektion ausschließlich relevanter und nichtredundanter Einflussgrößen. Daraus resultiert die Anforderung der Minimierung der Dimensionalität des Merkmalsraumes \mathcal{M} :

$$G_2 = \text{card}(\mathcal{M}) . \quad (4.40)$$

Zellenanzahl

Im Sinne einer guten Generalisierungsfähigkeit ist es erstrebenswert, eine Stützstellenkonfiguration $\mathcal{S}_{\mathcal{M}}$ zu finden, die eine möglichst geringe Anzahl mit Datenpunkten besetzter Hyperquader $\mathcal{Q}_{\mathcal{M}}^*$ induziert. Die Anzahl von unbesetzten Hyperquadern $\mathcal{Q}_{\mathcal{M}}^o$ sollte ebenfalls möglichst gering sein. Dies führt insgesamt zur Anforderung der generellen Minimierung der Hyperquaderanzahl:

$$G_3 = \text{card}(\mathcal{Q}_{\mathcal{M}}) . \quad (4.41)$$

Datendichte

Ebenfalls im Sinne einer guten Generalisierungsfähigkeit sind solche Stützstellenkonfigurationen $\mathcal{S}_{\mathcal{M}}$ zu bevorzugen, die zu einer möglichst gleichmäßigen Ver-

teilung aller Datenpunkte über die q^* besetzten Hyperquader $Q_{\mathcal{M}}^*$ führen. Dies führt zur Anforderung der Minimierung der Varianz der Datenpunktverteilung:

$$G_4 = \frac{1}{q^*} \sum_{i=1}^{q^*} (N_i^2 - \bar{N}^2), \quad (4.42)$$

wobei

$$\bar{N} = \frac{N}{\text{card}(Q_{\mathcal{M}}^*)} \quad (4.43)$$

die mittlere Datendichte über alle Hyperquader $Q_{\mathcal{M}}^*$ darstellt und N_i die Anzahl von Datenpunkten innerhalb eines Hyperquaders $Q_{\mathcal{M}}^{*i}$ quantifiziert.

4.8 Selektionsstrategien

Für die dieser Arbeit zugrunde liegende Problemstellung der Selektion relevanter und nichtredundanter Einflussgrößen kommen im Wesentlichen zwei grundlegend verschiedene Ansätze in Betracht. Diese unterscheiden sich einerseits in der jeweils damit verbundenen Handhabbarkeit im Sinne der Dimensionalität des Modellierungsproblems und andererseits bezüglich der Möglichkeit zur Quantifizierung von Relevanzrelationen zwischen einzelnen Einflussgrößen. Zusätzlich kann noch die Möglichkeit bzw. Notwendigkeit einer erweiterten multikriteriellen Optimierung als Unterscheidungsmerkmal herangezogen werden.

4.8.1 Iterativer Selektionsprozess

Zur Generierung eines Merkmalsssatzes $\mathcal{M}_q \subseteq \mathcal{X}$ von q möglichst relevanten und nichtredundanten Einflussgrößen $X_i \in \mathcal{X}$ werden beim *Iterativen Selektionsprozess* gemäß einer *Vorwärtsselektion* iterativ q Selektionsstufen durchlaufen. In der ersten Selektionsstufe $h = 1$ wird dabei zunächst die maximal erzielbare Relevanz η_i einer jeden potenziellen Einflussgröße $X_i \in \mathcal{X}$ durch eine evolutionäre Optimierung EFS(X_i) der zugehörigen Stützstellenverteilung \mathcal{S}_i ermittelt. Die Relevanzbewertung erfolgt dabei anwendungsspezifisch anhand eines der eingeführten Maße (4.26), (4.27), (4.29) oder (4.30). Auf Grundlage der ermittelten Relevanzwerte wird die Einflussgröße mit der höchsten Relevanz η_1 als erste Größe des Merkmalsssatzes \mathcal{M}_1 übernommen.

Darauf aufbauend wird der Merkmalsatz iterativ um diejenige noch nicht ausgewählte Einflussgröße erweitert, deren Hinzufügung zum bereits selektierten Merkmalsatz die größte Relevanzsteigerung erbringt. Hierzu wird nach einem

abgeschlossenen Selektionsschritt h der dabei gebildete Merkmalsatz \mathcal{M}_h in Kombination mit jeweils einer verbliebenen, noch nicht selektierten Einflussgröße $X_i \notin \mathcal{M}_h$ untersucht, indem die Stützstellenverteilungen \mathcal{S}_{h+1} der damit jeweils betrachteten $h+1$ Einflussgrößen mit $\text{EFS}(\mathcal{M}_h \cup \{X_i\})$ simultan gemäß des gewählten Relevanzmaßes optimiert werden. Diejenige Einflussgröße X_j , die die größte Relevanzsteigerung induzieren kann, wird für die Erweiterung des Merkmalsatzes \mathcal{M}_h ausgewählt. Auf diese Weise wird der neue Merkmalsatz \mathcal{M}_{h+1} mit optimierten Stützstellen für X_j sowie neu optimierten Stützstellen aller Größen von \mathcal{M}_h gebildet. Dieser sequentielle Erweiterungsprozess wird beendet, wenn keine oder keine „signifikante“ Relevanzsteigerung $\Delta\eta_{h+1}$ durch Hinzufügung einer weiteren potenziellen Einflussgröße erzielt werden kann. Folglich beinhaltet der letztendlich selektierte Satz \mathcal{M}_q keine überflüssigen — und somit redundanten oder gar irrelevanten — Einflussgrößen. Bei der Festlegung, was als „signifikant“ anzusehen ist, ist durch die Wahl einer Signifikanzschwelle ε_η ein meist anwendungsspezifischer Kompromiss zwischen den gegenläufigen Zielen „hohe Relevanz“ und „möglichst wenig Einflussgrößen“ zu schließen. Der nachfolgende Algorithmus 4.1 beschreibt detailliert die genaue Funktionsweise dieser iterativen Selektionsstrategie.

Algorithmus 4.1 Iterative Selektion des Merkmalsatzes \mathcal{M}_q

Gegeben: Menge \mathcal{X} aller n gegebenen Einflussgrößen X_i , Schwellwert ε_η

Ausgabe: Merkmalsatz \mathcal{M}_q

```

1:  $\mathcal{M}_0 \leftarrow \emptyset, \quad \eta_0 = 0, \quad \Delta\eta_0 = \varepsilon_\eta, \quad h = 0$ 
2: while  $\Delta\eta_h < \varepsilon_\eta \wedge h \leq n$  do
3:    $h \leftarrow h + 1$ 
4:   for all  $X_i \notin \mathcal{M}_{h-1}$  do
5:      $\tilde{\mathcal{M}}_i \leftarrow \mathcal{M}_{h-1} \cup \{X_i\}$ 
6:      $\eta_i = \text{EFS}(\tilde{\mathcal{M}})$ 
7:   end for
8:    $t = \{i \mid \eta_i = \max\{\eta_i\}, \forall i\}$ 
9:    $\eta_h = \eta_t$ 
10:   $\Delta\eta_h = \eta_h - \eta_{h-1}$ 
11:  if  $\Delta\eta_h \geq \varepsilon_\eta$  then
12:     $\mathcal{M}_h \leftarrow \mathcal{M}_{h-1} \cup \{X_t\}$ 
13:  end if
14: end while
15:  $q = h$ 
16:  $\mathcal{M}_q \leftarrow \mathcal{M}_h$ 

```

4.8.2 Global operierender Selektionsprozess

Der oben beschriebene iterative Selektionsprozess eröffnet dem Anwender aufgrund der erzielten Reihenfolge der Erweiterung des Merkmalsatzes \mathcal{M}_q in Verbindung mit den jeweils erzielten Relevanzsteigerungen $\Delta\eta_j$ eine entsprechende

Relevanzrelation zwischen den selektierten Merkmalen $X_j \in \mathcal{M}_q$. Aufgrund der COD-Problematic ist die Anwendung des iterativen Greedy-Algorithmus bei großen Datenmengen \mathcal{D} und einer hohen Anzahl n potenziell relevanter Einflussgrößen $X_i \in \mathcal{X}$ nicht mehr praktikabel. Unter Verzicht auf das Ergebnis einer Relevanzrelation zwischen selektierten Einflussgrößen kann für solche Anwendungsfälle eine Verbindung von Merkmalsselektion und Suchraumpartitionierung innerhalb einer einzigen — dafür entsprechend aufwändigeren — evolutionären Optimierung als global operierender Selektionsprozess wie folgt implementiert werden:

Die Granularisierung einer Einflussgröße X_i wird nicht mehr absolut über einen Partitionierungsvektor \mathbf{s}_i definiert. Stattdessen wird über einen Parameter $k_i \in \{0, 1, 2, \dots\}$ die *Anzahl* und über einen zweiten Parameter $p \in [-2, 2]$ die *Lage* der Stützstellen zur Partitionierung der Einflussgröße X_i kodiert.

Eine Einflussgröße X_i ist während der evolutionären Optimierung zur Relevanzbewertung genau dann selektiert, wenn

$$k_i \neq 0 \quad \wedge \quad p_i \neq 0 \quad (4.44)$$

gilt. Zur Ermittlung des Partitionierungsvektors $\mathbf{s}_i^T = (s_{i,1}, s_{i,2}, \dots, s_{i,k})$ wird zunächst ein Vektor \mathbf{a} mit einer äquidistanten Verteilung der Stützstellen in einem Bereich von $[0, 1]$ für $p < 0$ bzw von $[-1, 1]$ für $p > 0$ ermittelt:

$$\mathbf{a} = \begin{cases} -1 + \frac{2}{k-1}, -1 + \frac{4}{k-1}, \dots, 1 - \frac{4}{k-1}, 1 - \frac{2}{k-1} & \text{für } p > 0 \\ \frac{1}{k-1}, \frac{2}{k-1}, \dots, 1 - \frac{2}{k-1}, 1 - \frac{1}{k-1} & \text{für } p < 0 \end{cases} \quad (4.45)$$

Anschließend wird in Anlehnung an [CHMV01] die Lage der Stützstellen über eine nichtlineare Skalierungsfunktion verändert und gegebenenfalls in den Wertebereich $[0, 1]$ transformiert:

$$f(\mathbf{a}) = \begin{cases} \text{sign}(\mathbf{a}) \cdot |\mathbf{a}|^{|p|} & \text{für } p > 0 \\ 1 - (1 - |\mathbf{a}|)^{\frac{1}{|p|}} & \text{für } p < 0 \end{cases} \quad (4.46)$$

Abbildung 4.3 illustriert am Beispiel von sechs Stützstellen, inwieweit sich in Abhängigkeit des Parameters p die Lage der Stützstellen zur Mitte, zu beiden Rändern oder zum linken bzw. rechten Rand des normierten Merkmalsraums hin verdichten.

Das derart formulierte Optimierungsproblem ist im Verhältnis zum iterativen Selektionsprozess deutlich komplexer. Von daher werden neben dem eigentlichen anwendungsabhängig gewählten Relevanzmaß aus Abschnitt 4.6 auch alle zusätzlichen Bewertungskriterien aus Abschnitt 4.7.2 multikriteriell optimiert.

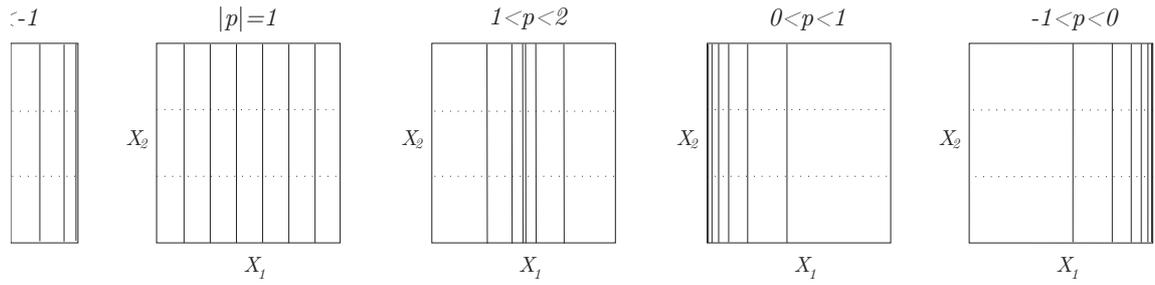


Abbildung 4.3: Exemplarische Darstellung möglicher Stützstellenkonfigurationen in Abhängigkeit von p .

4.9 Generierung effizienter Fuzzy-Zugehörigkeitsfunktionen

Für die selektierten Einflussgrößen X_i aus \mathcal{M}_q werden automatisch adäquate trapezförmige Fuzzy-Zugehörigkeitsfunktionen auf Grundlage der evolutionär optimierten Stützstellenverteilungen \mathcal{S}_i in Form eines Fuzzy-Informationssystems generiert (Abschnitt 4.9.1). Die Generierung der Zugehörigkeitsfunktionen für die Ausgangsvariable Y erfolgt für Approximationsprobleme, basierend auf einer Auswertung der Verteilungen der Ausgangsgrößenwerte y_s , innerhalb der mit Datenpunkten besetzten Hyperquader Q_δ (Abschnitt 4.9.2). Für Klassifikationsprobleme ist diese Auswertung nicht erforderlich. In diesem Fall wird für jede in den Daten vorhandene Klasse ein Singleton definiert.

4.9.1 Eingangsseitige Zugehörigkeitsfunktionen

Die Definition der Fuzzy-Zugehörigkeitsfunktionen $\mu_{i,j}(X_i)$ für jede der selektierten Einflussgrößen X_i erfolgt durch Spezifizierung von vier Parametern $a_{i,j}$, $b_{i,j}$, $c_{i,j}$ und $d_{i,j}$, basierend auf der ermittelten Stützstellenverteilung \mathcal{S}_i . Zwei benachbarte Stützstellen $s_{i,j}$ und $s_{i,j+1}$ bilden dabei jeweils die linke und rechte Grenze eines Intervalls j , $1 \leq j \leq k-1$, worin mittig der Kern der trapezförmigen Zugehörigkeitssfunktion $\mu_{i,j}(x_i)$ definiert wird (Abbildung 4.4). Zur Bestimmung der steigenden und fallenden Flanken wird der Grad der Zugehörigkeit für zwei benachbarte überlappende Zugehörigkeitsfunktionen in der Stützstelle $s_{i,j}$ auf jeweils $\mu_{i,j}(s_{i,j}) := \mu_{i,j-1}(s_{i,j}) := 0.5$ gesetzt.

Die Berechnung der vier Parameter erfolgt gemäß

$$a_{i,j} = s_{i,j} - \gamma \quad (4.47)$$

$$b_{i,j} = s_{i,j} + \gamma \quad (4.48)$$

$$c_{i,j} = s_{i,j+1} - \gamma \quad (4.49)$$

$$d_{i,j} = s_{i,j+1} + \gamma \quad (4.50)$$

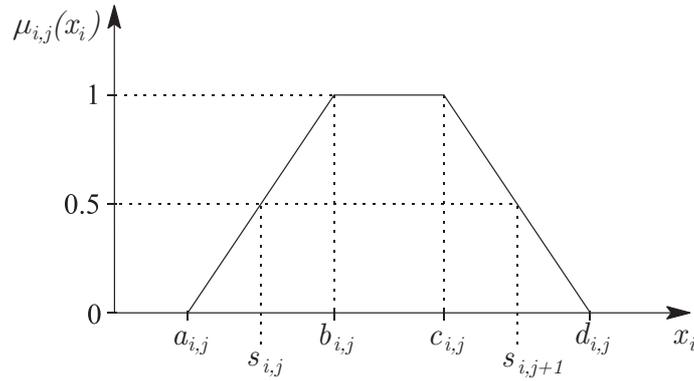


Abbildung 4.4: Trapezförmige Zugehörigkeitsfunktion $\mu_{i,j}(x_i)$ mit den zu bestimmenden Parametern $a_{i,j}$, $b_{i,j}$, $c_{i,j}$ und $d_{i,j}$ im Intervall j , das durch die Stützstellen $s_{i,j}$ und $s_{i,j+1}$ spezifiziert wird.

mit

$$\gamma = (1 - \alpha) \cdot \frac{ab_{i,min}}{2}. \quad (4.51)$$

Dabei ist $ab_{i,min}$ die Spannweite des kleinsten Intervalls der Einflussgröße X_i (Abbildung 4.5). Mit dem Faktor α , $0 \leq \alpha \leq 1$ ist die Steilheit der Flanken skalierbar. Für $\alpha = 1$ erhält man überlappungsfreie Rechtecke. Mit abnehmenden Werten für α wird die Steilheit der Flanken zunehmend erniedrigt, so dass der Kern der Zugehörigkeitsfunktion innerhalb des Intervalls vermindert und gleichzeitig der Support erweitert wird. Letztendlich liefert $\alpha = 0$ im kleinsten Teilintervall eine dreieckförmige Zugehörigkeitsfunktion.

Ferner werden an den linken bzw. rechten Randintervallen des Wertebereichs von X_i die Parameter $a_{i,1} = b_{i,1}$ bzw. $c_{i,k-1} = d_{i,k-1}$ so gewählt, dass die Zugehörigkeitsfunktionen jeweils um die halbe Intervallbreite rechteckförmig nach außen gezogen werden.

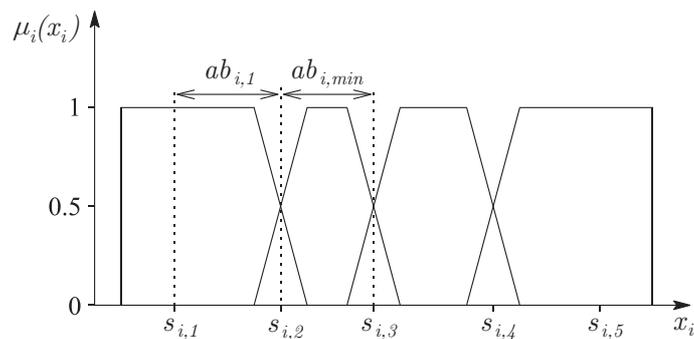


Abbildung 4.5: Beispiel einer Standard-Fuzzy-Partition mit 5 Stützstellen.

4.9.2 Ausgangsseitige Zugehörigkeitsfunktionen

Wie bereits in Abschnitt 4.7 ausgeführt, induziert die ermittelte optimierte Granularisierung für die q Einflussgrößen aus \mathcal{R}_q eine Zerlegung des Eingangsraumes in $(k-1)^q$ Hyperquader. Die Definition der ausgangsseitigen Zugehörigkeitsfunktion $\mu(y)$ erfolgt für Approximationsprobleme auf Grundlage der Verteilungen der Ausgangsgrößenwerte von \mathcal{Y}_δ , die jeweils in einem mit Datenpunkten besetzten Hyperquader H_δ eingeschlossen sind. Für jeden Hyperquader H_δ — und damit für jede Teilmenge \mathcal{Y}_δ — wird eine schiefe, d.h. nicht notwendigerweise zum Medianwert symmetrische Partialverteilungsfunktion $\tilde{f}_\delta(y)$ zur Approximation der Verteilung der darin enthaltenen Ausgangsgrößenwerte bestimmt (Abbildung 4.6).

Die maximale Amplitude der Partialverteilungsfunktion $\tilde{f}_\delta(y)$ wird an der Stelle des Medians $y_{\delta,med}$ durch den partiellen Gewichtungsfaktor

$$\omega_\delta = c_\delta \cdot \left(1 - \frac{y_{\delta,sp}}{y_{sp}}\right) \quad (4.52)$$

bestimmt. Hierdurch wird jedem Hyperquader H_δ bei der Generierung der ausgangsseitigen Zugehörigkeitsfunktionen ein Gewicht ω_δ proportional zur Anzahl c_δ der in H_δ eingeschlossenen Datenpunkte zugewiesen. Mit der gleichen Zielsetzung wird die partielle Spannweite $y_{\delta,sp}$ des Wertebereichs der c_δ enthaltenen Datenpunkte berücksichtigt

$$y_{\delta,sp} = y_{\delta,max} - y_{\delta,min} , \quad (4.53)$$

wobei $y_{\delta,max}$ und $y_{\delta,min}$ den maximalen bzw. minimalen Ausgangsgrößenwert y_s der Datenpunkte in H_δ bezeichnen. Hiermit werden solche Hyperquader, für die der Wertebereich der Ausgangsgrößenwerte der darin eingeschlossenen Datenpunkte klein ist, ebenfalls höher gewichtet. Dazu wird die partielle Spannweite $y_{\delta,sp}$ ins Verhältnis zum gesamten Wertebereich $y_{sp} = y_{s,max} - y_{s,min}$ aller m Ausgangsgrößenwerte y_s aus \mathcal{Y} gesetzt (4.52).

Ausgehend von den Beziehungen (4.52) und (4.53) wird die schiefe Partialverteilungsfunktion $\tilde{f}_\delta(y)$ durch

$$\tilde{f}_\delta(y) = \begin{cases} \omega_\delta \cdot \exp\left(-\frac{(y-y_{\delta,med})^2}{2 \cdot \sigma_{\delta,1}^2}\right) & \text{für } y \leq y_{\delta,med} \\ \omega_\delta \cdot \exp\left(-\frac{(y-y_{\delta,med})^2}{2 \cdot \sigma_{\delta,2}^2}\right) & \text{für } y > y_{\delta,med} \end{cases} \quad (4.54)$$

definiert. Darin werden die beiden Parameter $\sigma_{\delta,1}$ und $\sigma_{\delta,2}$ so gewählt, dass die Funktionswerte $\tilde{f}_\delta(y_{\delta,min})$ und $\tilde{f}_\delta(y_{\delta,max})$ an den Wertebereichsgrenzen 5% des

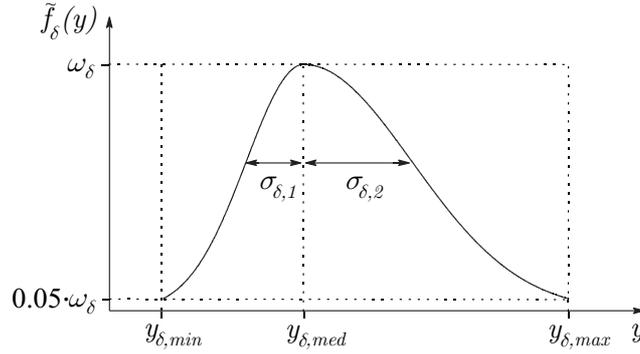


Abbildung 4.6: Schiefe Partialverteilungsfunktion $\tilde{f}_\delta(y)$ zur Approximation der Verteilung der Ausgangsgrößenwerte der Datenpunkte innerhalb eines Hyperquaders.

Maximalwertes ω_δ erreichen:

$$\sigma_{\delta,1} = (\sqrt{2} \cdot \sqrt{-\ln 0.05})^{-1} \cdot (y_{\delta,med} - y_{\delta,min}) \quad (4.55)$$

$$\sigma_{\delta,2} = (\sqrt{2} \cdot \sqrt{-\ln 0.05})^{-1} \cdot (y_{\delta,max} - y_{\delta,med}) . \quad (4.56)$$

Für den Fall, dass alle Ausgangsgrößenwerte von \mathcal{Y}_δ innerhalb eines Hyperquaders H_δ den gleichen Wert aufweisen ($y_{\delta,sp} = 0$), wird für diesen Hyperquader H_δ ein mit der Anzahl c_δ der darin eingeschlossenen Punkte gewichteter Singleton definiert:

$$\tilde{f}_\delta(y) = c_\delta . \quad (4.57)$$

Durch additive Überlagerung aller zu den m Ausgangsgrößenwerten y_s aus \mathcal{Y} gehörigen Partialverteilungsfunktionen $\tilde{f}_\delta(y)$ entsteht die Gesamtverteilungsfunktion

$$F(y) = \sum_{\delta} \tilde{f}_\delta(y) . \quad (4.58)$$

Hierauf aufsetzend, werden dreieckförmige Zugehörigkeitsfunktionen wie folgt generiert: Zunächst werden alle Punkte e_i ermittelt, in denen die Funktion $F(y)$ relative Maxima aufweist. Diese werden in einer Menge $\mathcal{E} = \{e_1, \dots, e_j\}$ zusammengefasst. Unter Verwendung zweier Reduktionsalgorithmen wird diese Menge \mathcal{E} wie folgt reduziert. Zunächst wird von zwei dicht benachbarten Punkten e_i und e_j , die bezogen auf ihren Abstand $|e_i - e_j|$ eine hohe Funktionswertedifferenz $|F(e_i) - F(e_j)|$ aufweisen, der Punkt mit dem geringeren Funktionswert eliminiert. Dies leistet der nachfolgende Algorithmus 4.2.

Die resultierende Menge \mathcal{E}_{r1} wird anschließend durch Agglomeration weiter reduziert. Hierzu werden zwei dicht benachbarte Punkte e_i und e_j zum Punkt $e_0 = (e_i + e_j)/2$ zusammengefasst. Zu diesem Zweck wird der nachfolgende Algorithmus 4.3 verwendet.

Algorithmus 4.2 Reduktion der Menge \mathcal{E} durch Elimination benachbarter Punkte auf Grundlage einer gewichteten Funktionswertedifferenz

Gegeben: Menge \mathcal{E} von Punkten e_i , Schwellwert v_1

Ausgabe: Reduzierte Menge \mathcal{E}_{r1} von Punkten e_i

```

1:  $\mathcal{U} \leftarrow \mathcal{E}, \quad \mathcal{E}_{r1} \leftarrow \mathcal{E}$ 
2: while  $\mathcal{U} \neq \emptyset$  do
3:    $u_0 = \{u_i \mid F(u_i) = \max\{F(u_i) \mid \forall u_i \in \mathcal{U}\}\}$ 
4:    $\mathcal{U} \leftarrow \mathcal{U} \setminus \{u_0\}$ 
5:   for all  $u_i \in \mathcal{U}$  do
6:      $b_{i,breite} = |u_i - u_0|, \quad b_{i,höhe} = |F(u_i) - F(u_0)|$ 
7:   end for
8:    $\beta = \max\{b_{i,höhe} \mid \forall u_i \in \mathcal{U}\} / \max\{b_{i,breite} \mid \forall u_i \in \mathcal{U}\}$ 
9:   for all  $u_i \in \mathcal{U}$  do
10:    if  $\beta \cdot \frac{b_{i,breite}}{b_{i,höhe}} \leq v_1$  then
11:       $\mathcal{U} \leftarrow \mathcal{U} \setminus \{u_i\}, \quad \mathcal{E}_{r1} \leftarrow \mathcal{E}_{r1} \setminus \{u_i\}$ 
12:    end if
13:  end for
14: end while

```

Algorithmus 4.3 Reduktion der Menge \mathcal{E}_{r1} durch Agglomeration dicht benachbarter Punkte

Gegeben: Menge \mathcal{E}_{r1} von Punkten e_i , Schwellwert v_2

Ausgabe: Reduzierte Menge \mathcal{E}_{r2} von Punkten e_i

```

1:  $\mathcal{U} \leftarrow \mathcal{E}_{r1}, \quad \mathcal{E}_{r2} \leftarrow \mathcal{E}_{r1}$ 
2: while  $\mathcal{U} \neq \emptyset$  do
3:    $u_0 = \min\{u_i \mid \forall u_i \in \mathcal{U}\}$ 
4:    $\mathcal{U} \leftarrow \mathcal{U} \setminus \{u_0\}$ 
5:   for all  $u_i \in \mathcal{U}$  do
6:      $b_{i,breite} = |u_i - u_0|$ 
7:     if  $\frac{b_{i,breite}}{y_{sp}} < v_2$  then
8:        $\mathcal{U} \leftarrow \mathcal{U} \setminus \{u_i\}, \quad \mathcal{E}_{r2} \leftarrow \mathcal{E}_{r2} \setminus \{u_i, u_0\}$ 
9:        $e_0 = \frac{u_i + u_0}{2}$ 
10:       $\mathcal{E}_{r2} \leftarrow \mathcal{E}_{r2} \cup \{e_0\}$ 
11:       $\mathcal{U} \leftarrow \mathcal{U} \cup \{e_0\}$ 
12:    end if
13:  end for
14: end while

```

Basierend darauf, wird die ausgangsseitige Zugehörigkeitsfunktion $\mu(y)$ stückweise generiert, indem für jeden Punkt e_i aus \mathcal{E}_{r2} , $i = \{1, 2, \dots, \varepsilon\}$ eine Dreiecksfunktion $\mu_i(y)$ so festgelegt wird, dass die Dreiecksspitze im Punkt e_i und die dazugehörigen Fußpunkte in den Nachbarpunkten e_{i-1} und e_{i+1} liegen (Abbildung 4.7).

Für die Randpunkte e_1 und e_ε werden die nicht vorhandenen Nachbarpunkte e_0 bzw. $e_{\varepsilon+1}$ durch die Wertebereichsgrenzen ersetzt. Für den Fall, dass einer der Punkte e_1 oder e_ε mit einer Wertebereichsgrenze identisch ist, erfolgt die Definition von e_0 bzw. $e_{\varepsilon+1}$ durch Punktspiegelung von e_2 an e_1 bzw. von $e_{\varepsilon-1}$ an e_ε .

Die beschriebene Generierung der ausgangsseitigen Zugehörigkeitsfunktion wird für ein synthetisches Beispiel durch Abbildung 4.7 verdeutlicht.

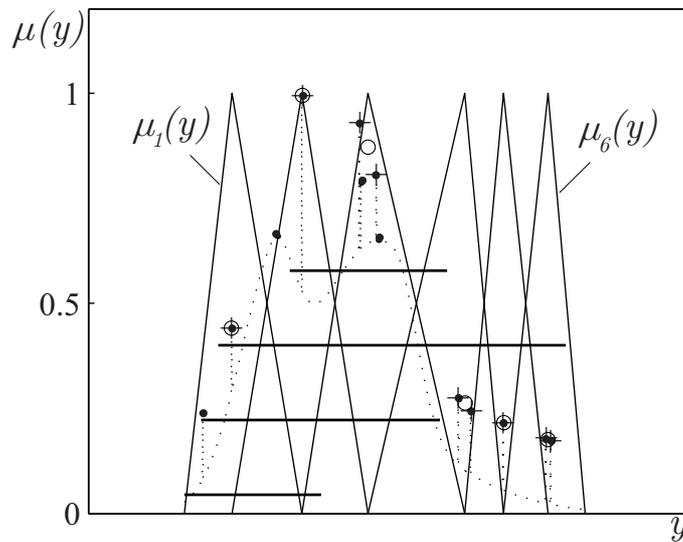


Abbildung 4.7: Beispiel zur Generierung der ausgangsseitigen Zugehörigkeitsfunktionen $\mu_i(y)$.

Hierzu wurden insgesamt 15 mit Datenpunkten besetzte Hyperquader H_δ generiert, von denen vier Hyperquader unterschiedliche Ausgangsgrößenwerte y_s einschließen. Die dazugehörigen partiellen Wertebereichsspannen $y_{\delta,sp}$ sind durch die vier horizontalen Linien dargestellt. Die von der Gesamtverteilungsfunktion $F(y)$ (gepunktete Kurve) ermittelten Extremwertstellen e_i aus \mathcal{E} sind mit “•” gekennzeichnet. Die nach der ersten Reduktion (Algorithmus 4.2) verbliebenen Extremwertstellen e_i aus \mathcal{E}_{r1} sind mit “+” markiert. Die nach der zweiten Reduktion (Algorithmus 4.3) verbliebenen Extremwertstellen e_i aus \mathcal{E}_{r2} sind mit “o” gekennzeichnet. Diese bilden die Grundlage für die Generierung der sechs dreieckförmigen Zugehörigkeitsfunktionen $\mu_i(y)$.

Kapitel 5

Strategien zur Diskretisierung der Evolutionsstrategie

Im direkt folgenden Abschnitt 5.1 wird zunächst motiviert, weshalb eine Modifikation herkömmlicher Evolutionsstrategien durch Diskretisierung des Suchraums sinnvoll ist. In Abschnitt 5.2 werden dann entsprechende diskretisierte Evolutionsstrategien konzipiert. Die Ergebnisse dieser implementierten Strategie belegen die Aufrechterhaltung einer ausreichend großen Diversität der Population während des gesamten Optimierungsprozesses. Hierdurch wird für eine ausgiebige Breitensuche gesorgt und die Neigung zur Terminierung in lokalen Minima verringert.

5.1 Motivation diskretisierter Suchräume

5.1.1 Vorzeitige Konvergenz in lokalen Optima

Jede Evolutionsstrategie steht vor dem prinzipiellen Problem die beiden gegenläufigen Ziele Exploration und Exploitation in sinnvoller Weise zu unterstützen. Tendiert das Suchverhalten nur zu einer Breitensuche, werden zwar Gebiete mit globalen Optima gefunden, die globalen Optima an sich können aber nur schwer entdeckt werden. Andererseits konvergiert ein Optimierungsverfahren mit rein exploitativem Suchverhalten schnell gegen ein Optimum. Meist handelt es sich jedoch um ein lokales Optimum, das aufgrund der Tiefensuche auch nicht mehr verlassen wird. Oft ist in konventionellen Evolutionsstrategien durch die Verwendung der selbstadaptiven Schrittweitenanpassung nach SCHWEFEL [Sch77] eine Kombination zwischen beiden Suchcharakteristiken umgesetzt. Der Anteil beider Charakteristiken am Gesamtsuchverhalten ändert sich im Laufe der Optimierung. Nach einer Phase mit dem Hauptgewicht auf der Breitensuche geht dieser Anteil später gegen Null und das Ver-

halten ändert sich zu einer Tiefensuche. Der Anteil der Tiefensuche ist jedoch auch am Anfang nicht gleich Null. Dies führt dazu, dass das globale Optimum in einer „freundlichen“ (z. B. muldenförmigen) Fitnesslandschaft zwar schnell gefunden, aber bei einer Fitnesslandschaft mit vielen lokalen Optima nur selten oder gar nicht entdeckt wird.

5.1.2 Bedeutung der Diversität einer Population

Ein weiterer Nachteil herkömmlicher Evolutionsstrategien besteht meistens darin, dass die Diversität der Population im Laufe der Optimierung verloren geht. Dabei hat der Begriff der Diversität mehrere Facetten: Diversität kann beispielsweise bedeuten, dass die Fitnesswerte der Individuen einer Population möglichst unterschiedlich sind. Eine andere – auch in dieser Arbeit zu Grunde gelegte – Auffassung der Diversität basiert auf der Verteilung der Individuen im Parameterraum, charakterisiert durch Aufsummierung aller paarweisen Abstände aller Individuen. Hiernach ist die Diversität umso größer, je gleichmäßiger die Individuen im gesamten Parameterraum verteilt sind.

Eine hohe Diversität verhindert eine schnelle ungewollte Konvergenz [Wei02]. Deshalb ist es günstig die Diversität durch geeignete Maßnahmen auf einem hohen Niveau zu halten. Beginnt man die evolutionäre Optimierung mit einer zufällig initialisierten Startpopulation, kann man von einer hohen Diversität zu diesem Zeitpunkt ausgehen. Doch bereits nach wenigen Generation kann sich die Diversität stark verringert haben. Verantwortlich hierfür ist in der Regel der verwendete Mechanismus zur Selektion. Ein hoher Selektionsdruck verringert die Diversität stärker als ein niedriger Selektionsdruck, weil in letzterem Fall auch schlechtere Individuen mit in die Folgepopulation übernommen werden. Diese liegen in der Regel nicht in unmittelbarer Nachbarschaft von guten Individuen.

Die Bedeutung der Diversität für den Erfolg der Optimierung kann man vereinfacht wie folgt beschreiben: Je größer die vorgesehenen Mutationen sind, desto größer kann – abhängig von der Selektion – die Diversität in der Population werden. Ist die Diversität groß, kann die Rekombination ebenfalls einen explorativen Charakter bekommen. Dies bedeutet, dass ein erforschendes Verhalten der Mutation ein ebensolches der Rekombination begünstigt, falls die Selektion nicht zu kontrahierend wirkt. Umgekehrt bewirkt eine geringe Mutation die Verringerung der Diversität, die im Extremfall dazu führt, dass die Rekombination nahezu nutzlos wird. Ein für konventionelle Evolutionsstrategien bekanntes Phänomen ist, dass sich die Individuen im Laufe der Optimierung immer weiter einander annähern, d. h. die Diversität immer mehr verloren geht. Sollte es sich

bei dem gefundenen Optimum nicht um das globale Optimum, sondern nur um ein lokales handeln, wird dies aufgrund der fehlenden Diversität dann nur noch schwer oder gar nicht mehr verlassen.

5.1.3 Motivation der Diskretisierung von Suchräumen

Während [Wei02] den besten Weg zur aktiven Erhaltung der Diversität in der Verwendung eines geeigneten Selektionsmechanismus sieht, wird in dieser Arbeit ein anderer Weg vorgeschlagen und ausgebaut. Dieser basiert auf der Intention, eine Mehrfachbesetzung sowie eine erneute Besetzung von Punkten im Parameterraum durch Individuen auszuschließen. Diese Intention ist allerdings nicht mit vertretbarem Aufwand im reellwertigen Parameterraum realisierbar. Daher wird hier eine Diskretisierung des Suchraums eingeführt. Hierzu wird entlang jeder Koordinatenachse eine Rasterung vorgenommen, so dass der Suchraum letztendlich aus Gitterpunkten besteht. Eine solche Diskretisierung erlaubt sowohl die Mehrfachbesetzung eines Punktes im Suchraum als auch die Wiederbesetzung durch ein Individuum einer späteren Generation auszuschließen. Zudem bleibt die Anzahl der Individuen, die sich in einem günstigen Gebiet versammeln können, in jeder Generation beschränkt. Damit wird zugleich auch die Möglichkeit eröffnet, lokale Optima wieder zu verlassen.

Zusätzlich erlaubt der Diskretisierung das Ergebnis der Optimierung zu beurteilen. Im Gegensatz zur konventionellen Evolutionsstrategie, bei der in der Regel am Ende alle Individuen den gleichen Punkt besetzen, kann jetzt die erhaltene Endpopulation dahingehend untersucht werden, inwieweit größere Suchraumgebiete von Individuen mit annähernd gleich guten Fitnesswerten aufgefunden wurden. Dies ist in vielen praktischen Anwendungsfällen von großem Interesse, denn hier besteht das Suchbestreben nicht darin, nur ein sehr begrenztes Gebiet mit guten Fitnesswerten zu finden. Vielmehr ist ein ausgedehntes Gebiet mit guten Fitnesswerten im Sinne der Robustheit einer Lösung deutlich zu bevorzugen: Bei Kenntnis eines solchen ausgedehnten Gebietes mit guten Fitnesswerten wird man die Lösung so auswählen, dass sie in der Mitte dieses Gebietes liegt. Handelt es sich bei der Lösung beispielsweise um einen Regler, so gewährleistet diese Wahl, dass das Regelungsverhalten robust gegenüber gewissen Parametervariationen des Regelungssystems ist.

5.2 Statische Diskretisierung des Suchraums

Wie vorab beschrieben, liegen die Vorteile eines diskretisierten Suchraums in einer Betonung des explorativen Charakters des Suchverhaltens. Zudem eröffnet die Diskretisierung den erfolgreichen Einsatz von zusätzlichen Maßnahmen zur Aufrechterhaltung einer günstigen Diversität, so dass lokale Optima wieder verlassen werden. Im Folgenden werden zunächst die Änderungen beschrieben, die zur Einführung einer statischen Diskretisierung an einer konventionellen Evolutionsstrategie vorgenommen werden müssen. Daran anschließend werden zusätzliche Strategieelemente zur adaptiven Diskretisierung des Suchraums vorgestellt.

5.2.1 Evolutionäre Suche in statisch diskretisierten Suchräumen

Die Änderungen zur Einführung einer statischen Diskretisierung werden durch Abbildung 5.1 im Überblick veranschaulicht.

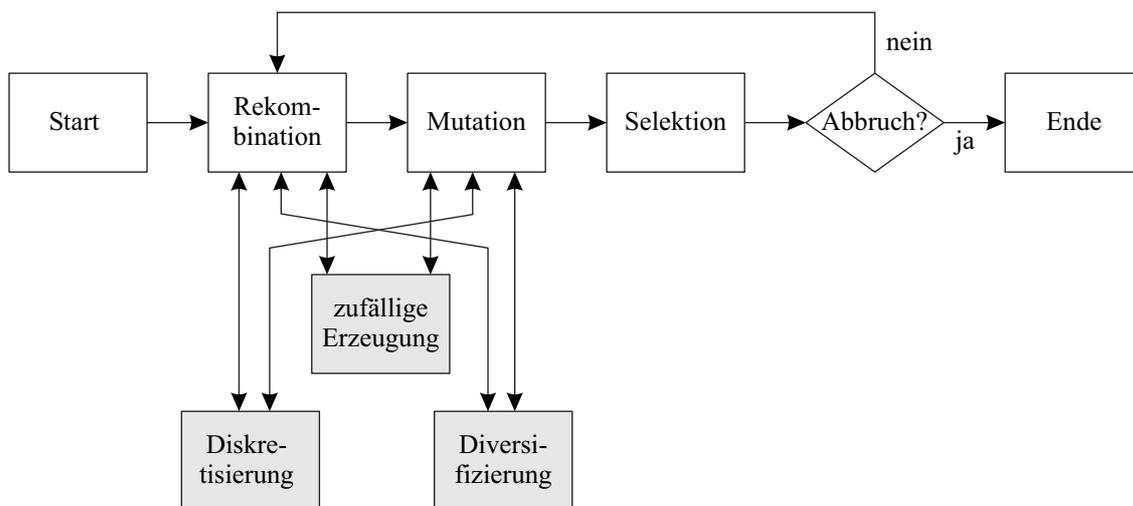


Abbildung 5.1: Modifizierter Ablauf der Evolutionsstrategie in diskretisierten Suchräumen.

Zur Diskretisierung des Parameterraums wird eine Rasterung entlang der Koordinatenachsen vorgenommen. Prinzipiell könnte die Rasterung in jeder Dimension des Parameterraums unterschiedlich gewählt werden. Aus Aufwandsgründen wird aber im Folgenden die Rasterung in jeder Dimension gleich gewählt. Die eigentliche Diskretisierung wird durch Runden auf bestimmte Werte erreicht. Gemäß einer vom Anwender vorgegebenen Rasterweite können die Individuen nur noch bestimmte Punkte in jeder Dimension besetzen. So entsteht aus dem gesamten Parameterraum ein äquidistantes Gitter. Die Individuen können nur noch die Gitterpunkte besetzen.

Die Rasterweite wird in normierter Form durch den Anwender eingestellt. Sie bezieht sich auf die Wertebereichsbreite in der jeweiligen Dimension. Eine eingestellte Rasterweite von $r = 0,01$ führt bei einer Wertebereichsbreite von 1000 zu einer realen Rasterweite von $r = 10$. Dies bedeutet, dass Individuen im Parameterraum nur noch Gitterpunkte mit einem Abstand von 10 entlang einer Koordinatenachse besetzen können.

Ein wichtiger Aspekt ist die Wahl einer geeigneten Rasterweite r . Wird eine große Rasterweite ausgewählt, birgt sie zwar den Vorteil einer Parameterraumabdeckung mit wenigen Schritten, aber möglicherweise können gewisse Details der Fitnesslandschaft nicht erkannt werden. Dieser Fall ist in Abbildung 5.2 dargestellt. Im Gegenzug sollte die Rasterweite aber auch nicht zu klein gewählt werden, da sonst die gleichen Nachteile auftreten, wie sie herkömmlichen kontinuierlich arbeitenden Verfahren eigen sind. Der Vorteil einer kleinen Rasterweite ist allerdings ein kleiner Diskretisierungsfehler, den wir hier als Differenz zwischen dem betrachteten Minimum und dem nächstgelegenen Rasterpunkt verstehen. Besser als mit diesem Fehler kann das Minimum nicht gefunden werden.

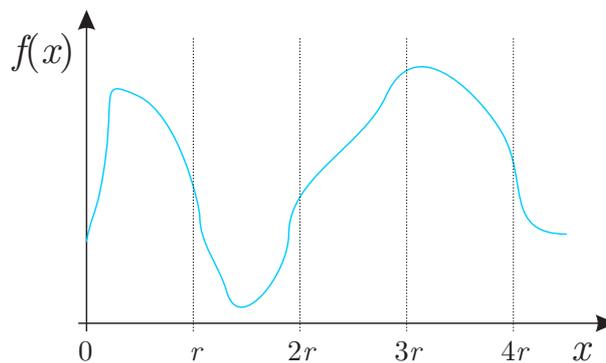


Abbildung 5.2: Beispiel für eine ungünstig gewählte Rasterweite r .

Durch empirische Ermittlung optimaler Rasterweiten für die Testbeispiele aus Abschnitt 5.4 wurden die Werte von $r = 0,1$ und $r = 0,0001$ als kleinster bzw. größter Wert ermittelt. Als bester Kompromiss für die Wahl einer festen Rasterweite für alle betrachteten Testbeispiele stellte sich der Wert $r = 0,01$ heraus.

Ferner ist ein Abbruchmechanismus vorgesehen, der den Optimierungsvorgang abbricht, bevor die Anzahl der Funktionsaufrufe die Anzahl der Gitterpunkte erreicht: Die Anzahl g der Gitterpunkte für ein n -dimensionales Optimierungsproblem mit der normierten Rasterweite r ist durch

$$g = \left(\frac{1}{r} + 1 \right)^n . \quad (5.1)$$

bestimmt. Ohne Abbruchmechanismus würde der Algorithmus bei Überschreitung der Anzahl vorhandener Gitterpunkt in eine Endlosschleife übergehen, weil es aufgrund der ausgeschlossenen Mehrfach- oder Wiederbesetzung nicht möglich ist, weitere Punkte zu besetzen. Daher ist ein Abbruchmechanismus implementiert worden, der die Optimierung beendet, falls ein bestimmter Prozentsatz, typischerweise 25%, aller Rasterpunkte besetzt ist bzw. besetzt war.

5.2.2 Diversifizierung der Population

Entscheidend für den Erfolg der diskretisierten Evolutionsstrategie sind geeignete Maßnahmen zur Populationsdiversifizierung. Wie einführend erläutert, ist der Erhalt der Diversität für den Erfolg der evolutionären Optimierung wesentlich. Im Folgenden wird eine Strategie zur Populationsdiversifizierung beschrieben, die auch im Rahmen dieser Arbeit eingesetzt wird. Diese ist prinzipiell für kontinuierliche, als auch für diskretisierte Suchräume einsetzbar, verspricht aber nur bei diskretisierten Suchräumen wirkliche Vorteile.

Die Maßnahme basiert auf einem einfachen Prinzip: Jeder besuchte Punkt im diskretisierten Suchraum wird in einer entsprechenden Liste markiert. Anschaulich entspricht das Verfahren einer Kennzeichnung besuchter Punkte in einer Karte des Suchraums (*Map*). Jeder dort verzeichnete Punkt kann nicht wieder besetzt werden. Somit wird sowohl eine Mehrfachbesetzung von Punkten im Suchraum, als auch das Wiederbesetzen von Punkten in zukünftigen Generationen verhindert.

Technisch wird die *Map* durch den Aufbau einer Tabelle mit den entsprechenden Einträgen realisiert. Wird während der Rekombination oder der Mutation ein neues Individuum erzeugt, wird zunächst ein Abgleich mit der *Map* durchgeführt. Ist der Punkt, an dem sich das Individuum im Suchraum befindet, bereits in der *Map* vorhanden, wird das Individuum verworfen und ein neues erzeugt. Nach einer bestimmten Anzahl von vergeblichen Versuchen ein neues Individuum zu erzeugen, das noch nicht in der *Map* verzeichnet ist, wird mittels eines Zufallsgenerators ein neues Individuum erzeugt. Der Zufallsgenerator erzeugt dazu Individuen gleichverteilt über den gesamten Suchraum. Die so erzeugten Individuen werden ebenfalls mit der *Map* abgeglichen und gegebenenfalls erneut gebildet.

Um Performanceverlusten durch Abgleiche mit einer sehr großen *Map* entgegenzuwirken, ist es vorgesehen, die *Map* als Speicher nach dem FIFO-Prinzip (First-In-First-Out) zu verwenden. In diesem Fall wirkt sie nur als Aufzeichnung der letzten k Generationen. Bei Erreichen der Generation $k+1$ werden die

Daten der ersten Generation gelöscht. Durch eine geeignete Wahl von k kann die Performance bei Optimierungsproblemen, für die eine große Anzahl von Funktionsaufrufe zu erwarten ist, verbessert werden, ohne dass durch Löschen der Daten die Diversität verloren geht. Durch regelmäßiges Sichern der *Map*-Daten in Dateien kann nach Ende der Optimierung der Inhalt der kompletten *Map* wiederhergestellt werden. Die Informationen können dann beispielsweise als Grundlage für die später beschriebenen Bewertungen genutzt werden.

5.2.3 Rekombination bei diskreten Suchräumen

Dieser Abschnitt erläutert die Änderungen, die an dem ursprünglichen Rekombinationsmechanismus vorgenommen werden müssen, wenn die Diskretisierung eingeführt wird. Die in dieser Arbeit verwendete Standard-Evolutionsstrategie verfügt über zwei verschiedene Rekombinationsmechanismen (vergl. Kapitel 3.5). Im Fall der diskreten Rekombination ist keine Anpassung erforderlich. Wenn vorausgesetzt wird, dass sich die Eltern auf einem Gitterpunkt befinden, müssen sich die Nachkommen ebenfalls auf Gitterpunkten befinden. Eine andere Situation ergibt sich, falls die Rekombination intermediär durchgeführt wird. In diesem Fall wird der Nachkomme analog zur kontinuierlichen Variante erzeugt. Danach wird die Position des Nachkommen auf den nächstliegenden Gitterpunkt gerundet. Wird eine Rasterweite r angenommen, die den Abstand zwischen zwei Gitterpunkten beschreibt, dann kann die durch Rundung entstehende Abweichung a durch

$$0 \leq |a| \leq \frac{r}{2} \quad (5.2)$$

abgeschätzt werden. Durch den Einsatz der *Map* zur Diversifizierung der Population besteht die Möglichkeit, dass erzeugte Individuen als ungültig ausgeschieden werden. In diesem Fall wird der entsprechende Rekombinationsvorgang wiederholt. Um unter Umständen eine Endlosschleife zu vermeiden, wird gleichzeitig ein Zählmechanismus initialisiert. Sollte dieser Zähler einen oberen Grenzwert erreichen, weil während der wiederholten Rekombination stets ungültige Individuen erzeugt wurden, wird das neue Individuum zufällig erzeugt. Hierzu wird zufällig eine durch die *Map* nicht verbotene Position im Parameterraum ausgewählt.

5.2.4 Mutation bei diskreten Suchräumen

Die Mutation der diskretisierten Evolutionsstrategie weicht ebenfalls von der kontinuierlichen Version ab. Der Unterschied besteht hier in der Bedeutung

der Schrittweite. Im kontinuierlichen Fall wird ein zufälliger Schritt auf Basis der mutierten Schrittweite ausgeführt. Technisch wird dies durch Addition eines Zufallsvektors zum Positionsvektor des Individuums realisiert. Die diskretisierte Evolutionsstrategie interpretiert den zufälligen Schritt bezogen auf Rasterpunkte. Realisiert wird dies durch eine Multiplikation des Zufallsvektors mit der Rasterweite und der dann folgenden Addition zum Positionsvektor.

Die Betrachtung der Schrittlänge, bezogen auf die Rasterweite, bewirkt bei unterschiedlichen Rasterweiten einen deutlichen Unterschied im Suchverhalten. Große Rasterweiten führen zu einem wesentlich schnelleren Durchschreiten des Suchraums. Da sich das Individuum vor der Mutation auf einem Gitterpunkt befunden hat, wird es sich auch nach der Mutation auf einem Gitterpunkt befinden, da durch die Mutation nur ganzzahlige Rasterschritte durchgeführt werden. Auch bei der Mutation können ungültige Individuen entstehen. Um dennoch gültige Mutationen durchführen zu können, wird der gleiche Mechanismus wie bei der Rekombination eingesetzt.

5.2.5 Selektion bei diskreten Suchräumen

Der Selektionsmechanismus wird nur insofern geändert, als dass die Möglichkeit vorgesehen wird, die Populationsgröße während der Optimierung situationsabhängig zu ändern.

In der Literatur finden sich verschiedene Arbeiten, die sich mit dynamischen Populationsgrößen befassen. Beispielsweise empfiehlt [CTR99] für den Fall, dass keine weiteren Informationen über das Optimierungsproblem vorhanden sind, die Verwendung einer zufallsgesteuerten Variation der Populationsgröße gegenüber einer willkürlich festgelegten festen Populationsgröße. Die in dieser Arbeit verwendete Art der Populationsgrößenvariation bezieht sich in erster Linie auf eine *Vergrößerung* der Population.

Im Folgenden wird eine Strategie zur Erhöhung der Populationsgröße bei Vorliegen nur einer Fitnessfunktion beschrieben. Eine konventionelle Fitnessselektion wählt aus einer Menge von $\mu + \lambda$ Individuen die besten λ Individuen als Eltern für die nächste Generation aus. Dabei gilt üblicherweise $\mu \leq \lambda$. Technisch wird die Selektion durch eine Sortierung der Individuen nach ihren Fitnesswerten und Auswahl der ersten μ Individuen umgesetzt. Im Rahmen des Populationsvergrößerungsmechanismus wird die Fitness des Individuums \mathbf{Ind}_μ mit der Fitness der Individuen $\mathbf{Ind}_{\mu+1}$ bis $\mathbf{Ind}_{\mu+\lambda}$ verglichen. An den Fitnesswert des Individuums \mathbf{Ind}_μ wird ein Toleranzbereich der Breite $+\varepsilon$ angelegt. Alle Individuen, deren Fitnesswerte innerhalb dieses Toleranzbereichs

liegen, werden zusätzlich in die Population übernommen. Abbildung 5.3 zeigt exemplarisch eine Erhöhung der Populationsgröße um ein weiteres Individuum.

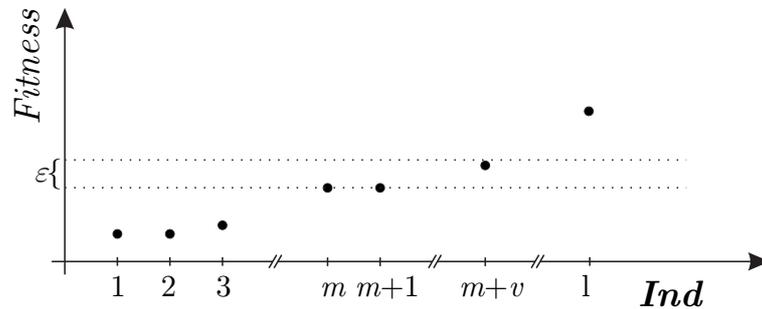


Abbildung 5.3: Prinzip der Erhöhung der Populationsgröße.

Bei skalaren Optimierungsproblemen bewirkt die Populationsvergrößerung, dass Plateaus der Fitnesslandschaft, also Suchraumgebiete mit gleichen Fitnesswerten, *vollständig* besetzt werden. *Vollständig* bedeutet in diesem Zusammenhang, dass alle Gitterpunkte, die das Plateau einschließt, besetzt sind. Die Verwendung der Populationsvergrößerung für ein Optimierungsproblem mit mehr als einer Fitnessbewertungsfunktion wirkt sich erfahrungsgemäß stärker aus als bei einem skalaren Optimierungsproblem. Das entscheidende Kriterium zur Vergrößerung der Population ist dabei die Anzahl der Pareto-optimalen Lösungen. Ist diese größer als die ursprünglich gewünschte Elternzahl, wird die Population entsprechend vergrößert. Während das Kontinuum Pareto-optimaler Lösungen bei einer kontinuierlichen Evolutionsstrategie zu einer unendlich großen Population führt, wird durch die Diskretisierung des Suchraums die mögliche Anzahl Pareto-optimaler Lösungen begrenzt und somit eine endliche Populationsgröße garantiert.

Unabhängig von der Anzahl der Fitnessfunktionen des Optimierungsproblems wird für die Populationsvergrößerung die Möglichkeit vorgesehen, den Selektionsdruck auf den ursprünglichen Wert zurückzusetzen. Hierzu wird die Anzahl der Nachkommen pro Generation auf die geänderte Elternzahl angepasst. Unter Umständen ist es jedoch wünschenswert, den Selektionsdruck nicht konstant zu halten, sondern die Erhöhung des Selektionsdrucks durch die Populationsvergrößerung gezielt dafür einzusetzen, dass das Suchverhalten einen explorativeren Charakter bekommt. Der allgemein bekannte Eigenschaft kontinuierlicher Evolutionsstrategien im Laufe des Optimierungsprozesses unter Umständen zu schnell von einer Breitensuche zur einer Tiefensuche überzugehen, wird hierdurch entgegengewirkt.

5.3 Adaptive Diskretisierung des Suchraums

Die bisherige statische Diskretisierung des Suchraums basierend auf einer festen Rasterweite wird im Folgenden durch einen Mechanismus zur Adaption der Rasterweite erweitert. Hierzu werden nach einer einleitenden Motivation zur Einführung adaptiv diskretisierter Suchräume entsprechende Strategien und Maßnahmen zur Implementierung beschrieben.

5.3.1 Motivation adaptiv diskretisierter Suchräume

Ein zentrales Problem bei der bisher beschriebenen diskretisierten Evolutionsstrategie ist eine geeignete Wahl der Rasterweite. Dies ist ähnlich gelagert wie das Problem der Wahl einer günstigen Schrittweite bis SCHWEFEL [Sch77] die adaptive Schrittweitenanpassung entwickelte. Eine konstante Rasterweite während des gesamten Optimierungsdurchlaufs bringt Nachteile bezüglich der Performance (Schnelligkeit des Optimierungsvortschritts) oder bezüglich der Genauigkeit der gefundenen Lösung. Eine sehr ungünstig gewählte Rasterweite kann auch bei beiden Kriterien zu deutlich schlechteren Ergebnissen führen. Sowohl grobe als auch sehr feine Rasterweiten lassen sich jeweils nur bezüglich eines der Kriterien rechtfertigen.

Daher wird im Folgenden ein Mechanismus eingeführt, der die Rasterweite vergrößert und verfeinert. Dies führt zu einem Gewinn an Performance und der Anwender wird von der Wahl einer geeigneten festen Rasterweite befreit. Stattdessen wählt er nur eine obere und untere Schwelle, zwischen denen der Adaptionsmechanismus die Rasterweite variiert. Der obere Schwellwert der Rasterweite wird abhängig von der Dimensionszahl des Optimierungsproblems gewählt. In einem zweidimensionalen Testproblem hat sich ein Wert von etwa $1/10$ bewährt. Die untere Schwellwert ist in Abhängigkeit der gewünschten Genauigkeit bzw. der erwarteten Robustheit zu wählen.

Man kann die Rasteradaption als ein Analogon zur Methode der Finiten Elemente auffassen. Dort wird das zu berechnende Objekt z. B. für eine Feldberechnung mit einem Gitterpunktenetz überzogen und innerhalb der Netzmaschen das Feld bestimmt. Ausschließlich in den Gebieten mit einer starken Feldänderung wird dieses Gitter sukzessive verfeinert. Würde das Gitter auch in Gebieten mit geringer Feldänderung verfeinert, stiege der Rechenaufwand, ohne dass sich die Genauigkeit des Ergebnisses dort verbessern würde.

Eine Übertragung dieses Mechanismus zur Schaffung einer adaptiv diskretisierten Evolutionsstrategie in völlig anderer Form, dass nämlich das Raster nur an

bestimmten Stellen verfeinert wird, bringt allerdings keinen Vorteil in Bezug auf die Rechenleistung. Eine Rasterverfeinerung bei diskretisierten Suchräumen bedeutet ja nur, dass die Individuen in einem Gebiet mehr Möglichkeiten zur Besetzung eines Gitterpunktes erhalten. Es muss jedoch nicht ein Individuum an all diesen Positionen entstehen, so dass dadurch auch keine zusätzliche Fitnessauswertung und damit einhergehender zusätzlicher Rechenaufwand entsteht. Zudem ist das Erstellen und Verwalten eines Gitters mit verschiedenen Rasterweiten in verschiedenen Bereichen sehr aufwändig. Deshalb wird die Rasteradaptation in anderer Weise eingeführt.

5.3.2 Adaption der Suchraumdiskretisierung durch Gitterebenen

Der andersgeartete Ansatz für die adaptive Diskretisierung von Evolutionsstrategien basiert auf der Einführung einer Schar mehrerer parallel existierender Raster unterschiedlicher Teilung, den Gitterebenen. Ein Individuum befindet sich immer auf einer dieser Ebenen. Durch die Rekombination oder die Mutation wird ein Individuum erzeugt, das sich auf einer anderen Ebene befinden kann als das Ursprungsindividuum. Durch die verschiedenen Rasterweiten der Ebenen resultiert der Schritt eines Individuums mit einer bestimmten Schrittweite in einer vollkommen anderen Position je nachdem, auf welcher Ebene das Individuum sich befindet. Im Folgenden wird zuerst das Prinzip der Gitterebenen dargestellt. Danach folgt die Beschreibung, wie die Individuen andere Ebenen erreichen.

Die Diskretisierung des Suchraums entlang der Koordinatenachsen führt zu einem Gitter, das den Parameterraum ausfüllt. Pro Koordinatenachse können verschiedene Rasterungen vorgenommen werden. Betrachtet man alle unterschiedlichen Rasterungen entlang der Koordinatenachsen parallel, entstehen Gitterebenen. Es wird vorausgesetzt, dass für alle Koordinatenachsen jeweils die gleiche Anzahl verschiedener Rasterungen existiert. Die Anzahl der Rasterungen sei b . In einem n -dimensionalen Suchraum existieren folglich b Gitterebenen. In Abbildung 5.4 ist ein Ausschnitt der Gitterebenen für ein zweidimensionales vereinfachtes Beispiel dargestellt. Die Vereinfachung besteht darin, dass die Rasterung entlang aller Koordinatenachsen gleich angenommen wird, also symmetrisch ist. In diesem Fall existieren nur b Gitterebenen parallel. Jeder Schnittpunkt der Gitter- bzw. Randlinien stellt einen Gitterpunkt dar. Die verschiedenen Rasterungen der Koordinatenachsen sollten so gewählt werden, dass die daraus resultierenden Gitterebenen bezüglich der Gitterpunktlage zueinander kompatibel sind. Seien die Gitterebenen untereinander derart sortiert, dass der Ebene i eine Ebene $i + 1$ folgt, deren Rasterung entlang aller Koordi-

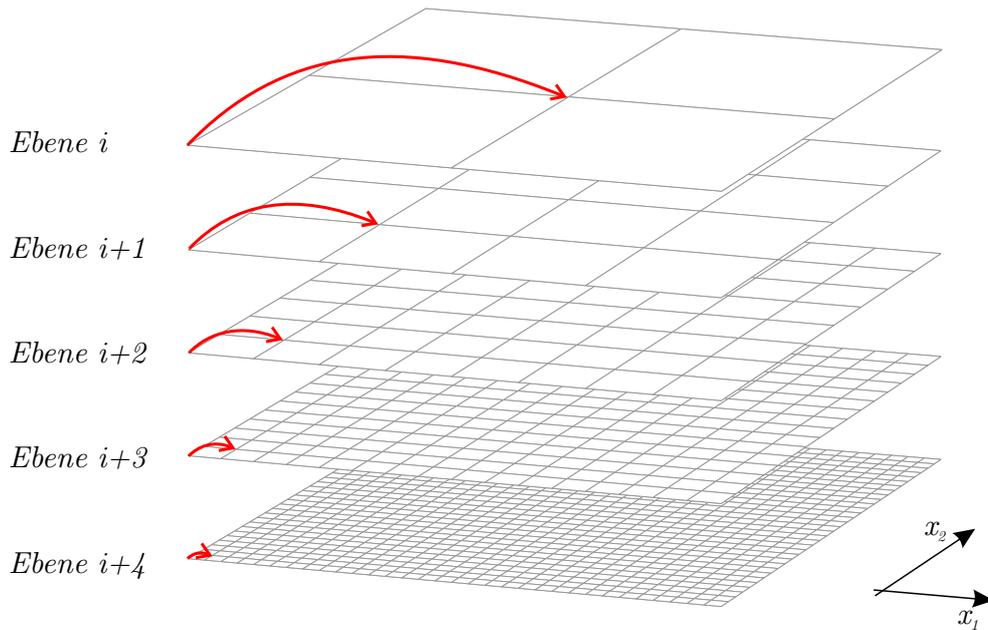


Abbildung 5.4: Gitterebenen bei symmetrischer Rasterung in allen Dimensionen.

natenachsen gleich oder feiner als in der Ebene i ist und sei G_i die Menge der Gitterpunkte der Ebene i , so gilt:

$$G_i \subseteq G_j \quad \forall j \in \{i+1, \dots, b^n\}. \quad (5.3)$$

Ein Ansatz zur Wahl der verschiedenen Rasterweiten, der die obige Forderung erfüllt, ist im Folgenden dargestellt. Wird beim Übergang von der größten Rasterung zur nächstfeineren Rasterung die Rasterweite t mal geteilt ($t \in \mathbb{N}$), dann muss der Wert t auch bei allen übrigen Übergängen verwendet werden. Für die Rasterweite gilt dann:

$$r_{i+1} = \frac{r_i}{t+1}. \quad (5.4)$$

In Rahmen dieser Arbeit wird immer $t = 1$ gewählt. Dies führt also bei absolut gleichmäßiger Abdeckung des Suchraums zu einer Halbierung der Rasterweite innerhalb der Rasterungen.

Die Bedeutung der verschiedenen Gitterebenen wird am folgenden Beispiel deutlich: Abbildung 5.4 zeigt, wie sich ein Schritt bei der Mutation eines Individuums auswirkt, wenn verschiedene Rasterebenen zugrunde gelegt werden. Angenommen werden ein zweidimensionaler Parameterraum und eine Auswahl von fünf Gitterebenen mit symmetrischer Rasterung. Die Rasterweiten halbieren sich von einer Ebene zur nächsten. Wird nun ein Schritt um einen Gitterpunkt in jeder Dimension des Parameterraums ausgeführt, ist leicht ersichtlich, dass je nach Ebene vollkommen andere Positionen im Parameterraum erreicht werden.

5.3.3 Evolutionäre Suche in adaptiv diskretisierten Suchräumen

Für die Implementierung der durch Gitterebenen adaptiv diskretisierten Suchräume nach Abbildung 5.5 werden die zugehörigen Rasterweiten ω_i , die ein Individuum aktuell verwendet, im genetischen Material des Individuums abgelegt.

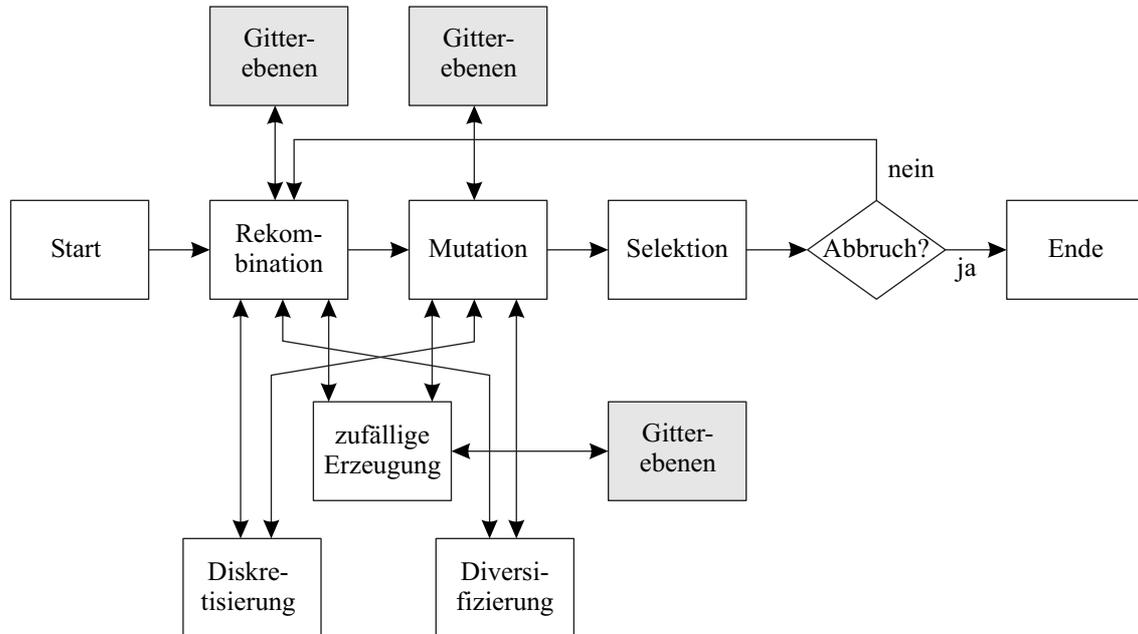


Abbildung 5.5: Ablaufschema der adaptiv diskretisierten Evolutionsstrategie.

Dieser Vektor wird mit $\bar{\Omega}$ bezeichnet. Die einzelnen Elemente dieses Vektors ω_1 bis ω_n enthalten jedoch nicht direkt einen Wert für die Rasterweite, sondern einen Index, der sich auf den Vektor \bar{R} bezieht, der die realen Rasterweiten r_1 bis r_b der b verschiedenen Rasterungen entlang der Koordinatenachsen enthält. Die Elemente ω_1 bis ω_n werden als reellwertige Zahlen angegeben. Die Auswertung als Index auf den Vektor \bar{R} erfolgt erst nach Runden auf einen ganzzahligen Wert. Der gesamte reellwertige Ausdruck spielt erst bei der Adaption des Vektors $\bar{\Omega}$ eine Rolle. Die Elemente des Vektors können Werte im Bereich $1 \leq \omega_i < b + 1$ annehmen.

Sobald ein neues Individuum durch den Mutationsmechanismus erzeugt wird, stellt sich die Frage, welche Rasterweite in jeder Dimension zugrunde gelegt werden soll. Im Konzept der Gitterebenen soll diese Eigenschaft der Individuen ebenfalls durch den Mechanismus der Evolution selbst eingestellt werden, so dass die Information über die Rasterweite genauso behandelt wird, wie die Position des Individuums im Suchraum. Durch die Mutation in Abhängigkeit von einer bestimmten Schrittweite wird die Position des Individuums neu festgelegt. Dieser Mechanismus gilt analog für die Rasterweite. Hierzu wird das genetische

Material des Individuums um einen zweiten Schrittweizensatz erweitert — die Rastermutationsschrittweiten $\sigma_{r,i}$. Zur Einstellung einer geeigneten Rastermutationsschrittweite wird die von SCHWEFEL [Sch77] entwickelte selbstadaptive Schrittweitenanpassung verwendet (Abschnitt 3.5.2).

Abbildung 5.6 zeigt das erweiterte genetische Material eines Individuums, das neben den konventionell vorhandenen Zielvariablen s_i und den Strategieparametern σ_i , welche die Position und die Schrittweiten im Suchraum darstellen, die weiteren Strategieparametersätze der Rasterweiten ω_i und der Rastermutationsschrittweiten $\sigma_{r,i}$ beinhaltet.

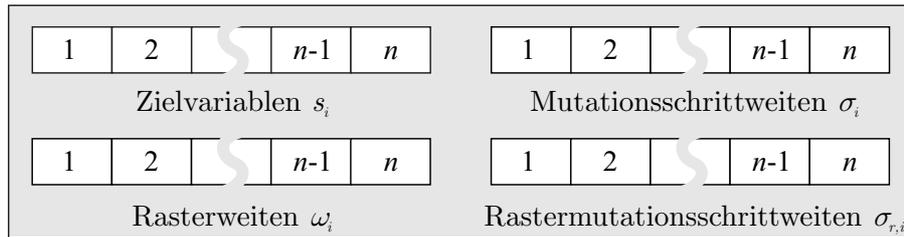


Abbildung 5.6: Individuum mit erweitertem genetischem Material.

Das genetische Material des Individuums besteht nun aus vier Parametersätzen mit jeweils n Elementen. Die hohe Zahl an Freiheitsgraden verbessert die Anpassungsfähigkeit der adaptiv diskretisierten Evolutionsstrategie an komplexe Optimierungsprobleme, führt aber auch zu einer Erhöhung des Suchaufwandes. Die Rastermutationsschrittweiten werden als reellwertige Zahlen angegeben. Wenn b die Anzahl der Rasterungen entlang jeder Koordinatenachse bezeichnet, liegen die Rastermutationsschrittweiten $\sigma_{r,i}$ im Bereich

$$\frac{b}{1000} \leq \sigma_{r,i} \leq \frac{b}{10} \quad \text{für } i = 1, \dots, n. \quad (5.5)$$

Die o. a. oberen und unteren Schwellwerte wurden entsprechend zum Verfahren der selbstadaptiven Schrittweitenanpassung angenommen. Die Mutation des Rastervektors $\bar{\Omega}$ läuft analog zur Mutation der Problemparameter ab. Äquivalent zum Ansatz zur selbstadaptiven Schrittweitenanpassung nach SCHWEFEL aus Gleichung 3.19 werden die Rastermutationsschrittweiten des Vektors σ_r des zu mutierenden Individuums gemäß

$$\sigma_{r,i}^{mut} = \sigma_{r,i} \cdot e^{N(0,\tau_3) + N_i(0,\tau_4)} \quad \text{für } i = 1, \dots, n \quad (5.6)$$

mutiert. Die so erhaltene neue Rastermutationsschrittweite wird als Standardabweichung für die eigentliche Mutation der Rasterebenen verwendet. Die darauf folgenden weiteren Schritte werden analog Kapitel 3.5.2 durchgeführt, wobei

$$\tau_3 = \tau_1 = \frac{c}{\sqrt{2 \cdot n}}, \quad \tau_4 = \tau_2 = \frac{c}{\sqrt{2 \cdot \sqrt{n}}} \quad \text{und} \quad c = 1 \quad (5.7)$$

gesetzt werden.

5.3.4 Steuerung der adaptiven Suchraumdiskretisierung

In simulativen Untersuchungen zum beschriebenen Konzept zur adaptiven Suchraumdiskretisierung zeigte sich, dass die Rasterschrittweiten nach der Schrittweitenanpassung analog SCHWEFEL [Sch87] unter Umständen zu schnell zu kleine Werte annehmen. Bei der Übertragung der selbstadaptiven Schrittweitenanpassung auf die Rastermutationsschrittweiten sollte dieser Effekt im Interesse einer ausgewogenen Auswahl der Rasterweiten vermieden werden.

Zur Schaffung einer Eingriffsmöglichkeit zur Ausnutzung aller Gitterebenen wird zusätzlich eine Wahrscheinlichkeit p_i für jede einzelne der b Gitterebenen eingeführt, die die Wahrscheinlichkeit der Verwendung dieser Gitterebene für ein Individuum angibt. Für den Fall dass jede Gitterebene die Wahrscheinlichkeit $p_i = \frac{1}{b}$ erhält, ist die Verwendung aller Gitterebenen gleichwahrscheinlich.

Der Wahrscheinlichkeitsvektor $\mathbf{p} = (p_1, p_2, \dots, p_b)$ erlaubt eine Steuerung der Rastermutation, hebt aber den Mechanismus der selbstadaptiven Anpassung der Rastermutationsschrittweite nicht aus. Im Interesse einer verstärkten Breitensuche wird zu Beginn des evolutionären Optimierungsprozesses zuerst die Suche auf gröberen Gitterebenen präferiert. Mit zunehmendem Optimierungsfortschritt wird die Suche dann auf feineren Ebenen fortgesetzt. Mithilfe des eingeführten Vektors \mathbf{p} , der die Auswahlwahrscheinlichkeiten für jede der b Rasterungen pro Koordinatenachse enthält, wird eine derart wirkende Steuerung implementiert.

Ein weiterer Nutzen des Wahrscheinlichkeitsvektors \mathbf{p} liegt in der Steuermöglichkeit der Abdeckung in den einzelnen Gitterebenen. Für den Fall, dass bezüglich der Maßnahme zur Diversifizierung der Population eine zunehmend höhere Abdeckung in einer Gitterebene erzielt wird, wird die Wahrscheinlichkeit für die entsprechende Rasterung entsprechend herabgesetzt. Die gröberen Gitterebenen werden ebenfalls derart mit niedrigeren Wahrscheinlichkeitswerten versehen werden, da gröbere Gitterebenen immer eine Teilmenge der Gitterpunkte der betrachteten Ebene enthalten. Details hierzu sind in [Sch02] zu finden.

5.4 Effizienzvergleich anhand analytischer Testfunktionen

Dieser Abschnitt beschreibt fünf ausgewählte Testprobleme, die in der Literatur üblicherweise verwendet werden. Mit Ausnahme der Six–Hump–Funktion sind die Testfunktionen aller Probleme n -dimensional definiert, also als Optimierungsprobleme mit einstellbarer Komplexität verwendbar. Die Testprobleme werden in dieser Arbeit zunächst zweidimensional untersucht, um entsprechende Vorteile und Wirkungsweisen graphisch zu illustrieren. Zu einigen ausgewählten Testproblemen werden darüber hinaus hochdimensionale Varianten analysiert. Bei den ersten vier verwendeten Testfunktionen handelt es sich um skalare Optimierungsprobleme, das fünfte hingegen ist ein multikriterielles Optimierungsproblem mit zwei unabhängig voneinander zu minimierenden Bewertungsfunktionen. Eine große Auswahl multikriterieller Optimierungsprobleme, die in der Literatur oft verwendet werden, findet sich in [Veld99]. Dort ist zu jedem Optimierungsproblem eine Approximation der Pareto–Front angegeben, so dass eine Einordnung der erzielten Ergebnisse leicht vorgenommen werden kann.

5.4.1 Testfunktionen

Six–Hump–Camel–Back–Funktion [DS78]

Diese Testfunktion wurde von Dixon und Szego vorgeschlagen. Innerhalb der Bereichsgrenzen befinden sich sechs lokale Minima, von denen zwei auch globale Minima sind. Die Funktion ist zweidimensional definiert. Die globalen Minima befinden sich in den Punkten $\mathbf{x}_1 = (-0,0898, 0,7126)$ und $\mathbf{x}_2 = (0,0898, -0,7126)$ mit $Q(\mathbf{x}_1) = Q(\mathbf{x}_2) = -1,0316$. Um einen Funktionswert von Null im globalen Minimum der Funktion zu erreichen, wird die Funktion durch Addition des konstanten Wertes 1,0316 angepasst:

$$Q(\mathbf{x}) = 4 \cdot x_1^2 - 2,1 \cdot x_1^4 + \frac{1}{3} \cdot x_1^6 + x_1 \cdot x_2 - 4 \cdot x_2^2 + 4 \cdot x_2^4 + 1,0316. \quad (5.8)$$

Bei Verwendung dieser Testfunktion werden die Zielvariablen im folgenden Suchraum optimiert:

$$-2,5 \leq x_1 \leq 2,5 \quad -1,5 \leq x_2 \leq 1,5.$$

Als Fitness-Schwellwert für den Abbruch des Optimierungsprozesses wird bei dieser Funktion ein Wert von $\rho = 10^{-5}$ verwendet. Aufgrund der Genauigkeit von 10^{-4} des additiven Terms 1,0316 können bei dieser Funktion auch

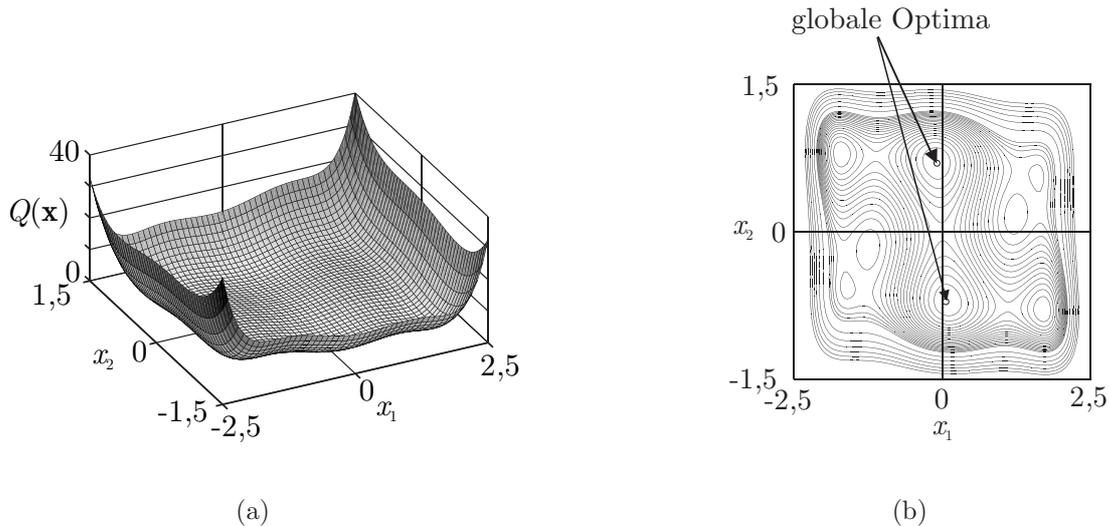


Abbildung 5.7: Testfunktion Six-Hump-Camel-Back, (a): $Q(x_1, x_2)$, (b): Höhenlinien, d. h. Linien gleicher Funktionswerte $Q(x_1, x_2)$.

negative Funktionswerte erreicht werden, die ebenfalls zum Abbruch des Optimierungsprozesses führen. Im Allgemeinen wird in der Literatur als Ziel des Optimierungsprozesses angegeben, mindestens eines dieser beiden globalen Optima mit der angegebenen Genauigkeit ρ zu ermitteln. Für die im Rahmen dieser Arbeit vorgestellte diskretisierte Evolutionsstrategie wird die Anforderung deutlich verschärft. Der Optimierungsprozess soll beide globalen Minima ermitteln. Die Funktion ist in Abbildung 5.7 dargestellt.

Rastrigin-Funktion [SHF94]

Rastrigin's Testfunktion basiert auf einer n -dimensional definierten einfachen quadratischen Funktion, die um einen Cosinus-Modulationsterm erweitert wurde. Die Funktion ist multimodal. Die Modulation führt zu einer Vielzahl lokaler Minima. Das globale Minimum liegt, unabhängig von der gewählten Dimension n , immer im Ursprung:

$$Q(\mathbf{x}) = n \cdot A + \sum_{i=1}^n (x_i^2 - A \cdot \cos(2 \cdot \pi \cdot x_i)) \quad \text{mit } A = 10. \quad (5.9)$$

Für $n = 2$ ergibt sich:

$$Q(\mathbf{x}) = 20 + (x_1^2 - 10 \cdot \cos(2 \cdot \pi \cdot x_1)) + (x_2^2 - 10 \cdot \cos(2 \cdot \pi \cdot x_2)). \quad (5.10)$$

Der Suchraum wird dabei in folgendem Bereich festgelegt:

$$-5,12 \leq x_i \leq 5,12 \quad \text{für } i = 1, \dots, n.$$

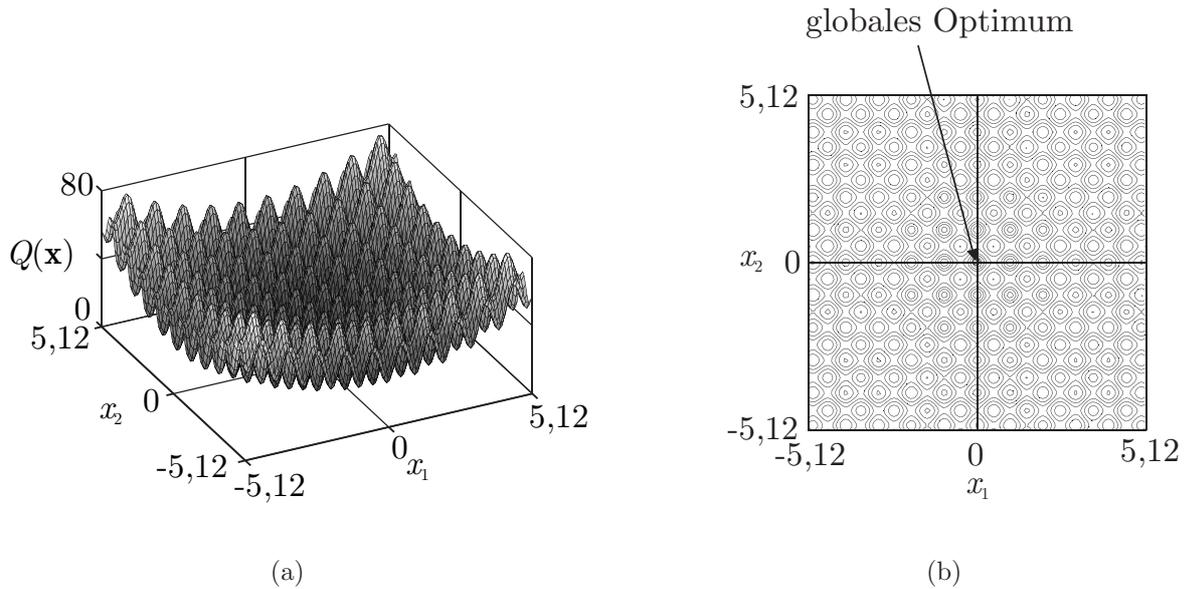


Abbildung 5.8: Testfunktion Rastrigin bei Verwendung zweier Eingangsgrößen x_1 und x_2 , (a): $Q(x_1, x_2)$, (b): Höhenlinien, d. h. Linien gleicher Funktionswerte $Q(x_1, x_2)$.

Als Schwellwert für den Abbruch des Optimierungsprozesses wird bei dieser Funktion ebenfalls ein Wert von $\rho = 10^{-5}$ verwendet. Abbildung 5.8 illustriert die Schwierigkeit des Optimierungsproblems für $n = 2$.

Rosenbrock–Funktion [Poh00]

Die Rosenbrock–Funktion [Ros60] ist in der Literatur sehr weit verbreitet. Aufgrund ihrer optischen Eigenschaften wird sie dort oft als *banana function* bezeichnet. Es handelt sich um eine unimodale bi–quadratische Funktion, die n -dimensional definiert ist. Obwohl diese Funktion keine lokalen Minima hat, ist es wesentlich schwieriger, das globale Minimum, zu finden als bei der Testfunktion von Rastrigin oder der Six–Hump–Funktion. Dies liegt an einem schmalen parabelförmigen Tal mit steil ansteigenden Rändern, das einen sehr flachen Boden hat.

Während das Auffinden des Tales an sich trivial ist, gestaltet sich die Konvergenz gegen das globale Minimum sehr schwierig, da der Gradient innerhalb des Tales extrem schwach abfällt, welches zudem auch noch diagonal zu den Koordinatenachsen verläuft:

$$Q(\mathbf{x}) = \sum_{i=1}^{n-1} (100 \cdot (x_i^2 - x_{i+1})^2 + (1 - x_i)^2). \quad (5.11)$$

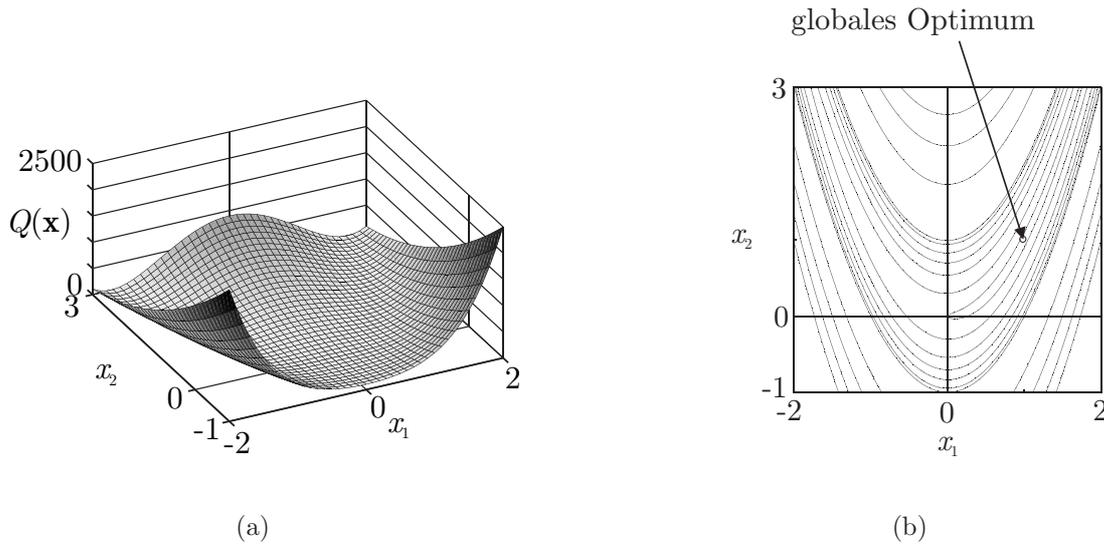


Abbildung 5.9: Testfunktion Rosenbrock bei Verwendung zweier Eingangsgrößen x_1 und x_2 , (a): $Q(x_1, x_2)$, (b): Höhenlinien, d. h. Linien gleicher Funktionswerte $Q(x_1, x_2)$.

Der Suchraum wird dabei in folgendem Bereich festgelegt:

$$-2.048 \leq x_i \leq 2.048 \quad \text{für } i = 1, \dots, n .$$

Als Schwellwert für den Abbruch des Optimierungsprozesses wird auch bei dieser Funktion ein Wert von $\rho = 10^{-5}$ verwendet. Die Funktion ist in Abbildung 5.9 für $n = 2$ dargestellt.

Schwefel-Funktion [SHF94]

SCHWEFEL hat die schwierigste Testfunktion im Rahmen der hier betrachteten Probleme vorgeschlagen. Seine Funktion ist ebenfalls n -dimensional definiert. Sie ist multimodal und besitzt sehr viele lokale Minima sowie ein globales Minimum für $\check{x}_i = 420,9687$ für $i = 1, \dots, n$ mit $Q(\check{\mathbf{x}}) = -837,9658 \cdot \frac{n}{2}$:

$$Q(\mathbf{x}) = \sum_{i=1}^n (-x_i \cdot \sin(\sqrt{|x_i|})) . \quad (5.12)$$

Um einen minimalen Funktionswert bei Null zu erhalten, wird die Funktion in dieser Arbeit durch Addition des konstanten Wertes $-837,9658$ für die Abbruchbedingung 10^{-6} angepasst. Für $n = 2$ ergibt sich entsprechend

$$Q(\mathbf{x}) = -837,9658 - x_1 \cdot \sin(\sqrt{|x_1|}) - x_2 \cdot \sin(\sqrt{|x_2|}) . \quad (5.13)$$

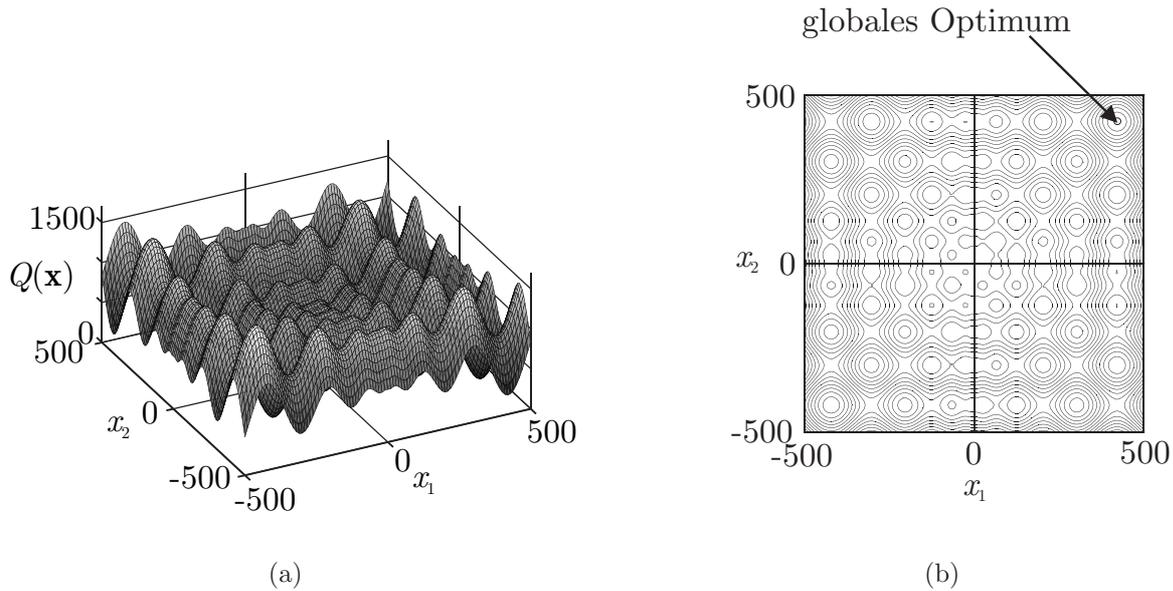


Abbildung 5.10: Testfunktion nach Schwefel bei Verwendung zweier Eingangsgrößen x_1 und x_2 , (a): $Q(x_1, x_2)$, (b): Höhenlinien, d. h. Linien gleicher Funktionswerte $Q(x_1, x_2)$.

Die einzustellenden Zielvariablen werden in folgendem Suchbereich variiert:

$$-500 \leq x_i \leq 500 \quad \text{für } i = 1, \dots, n .$$

Diese Testfunktion gehört zu einer Gruppe von „boshaften“ Testfunktionen, die ausschließlich dazu konzipiert sind, Schwächen von Optimierungsverfahren zu offenbaren. Viele Optimierungsverfahren konvergieren daher in der Regel vorzeitig gegen ein lokales Minimum dieser Funktion.

Kursawe–Funktion

Diese Testfunktion beinhaltet als multikriterielles Optimierungsproblem zwei zu minimierende Fitnessfunktionen:

$$Q_1(\mathbf{x}) = \sum_{i=1}^{n-1} -10 \cdot e^{-0,2 \cdot \sqrt{x_i^2 + w_{i+1}^2}} \quad (5.14)$$

und

$$Q_2(\mathbf{x}) = \sum_{i=1}^n |x_i|^{0,8} + 5 \cdot \sin^3(x_i) . \quad (5.15)$$

Die einzustellenden Zielvariablen werden in folgendem Suchbereich variiert:

$$-5 \leq x_i \leq 5 \quad \text{für } i = 1, \dots, n .$$

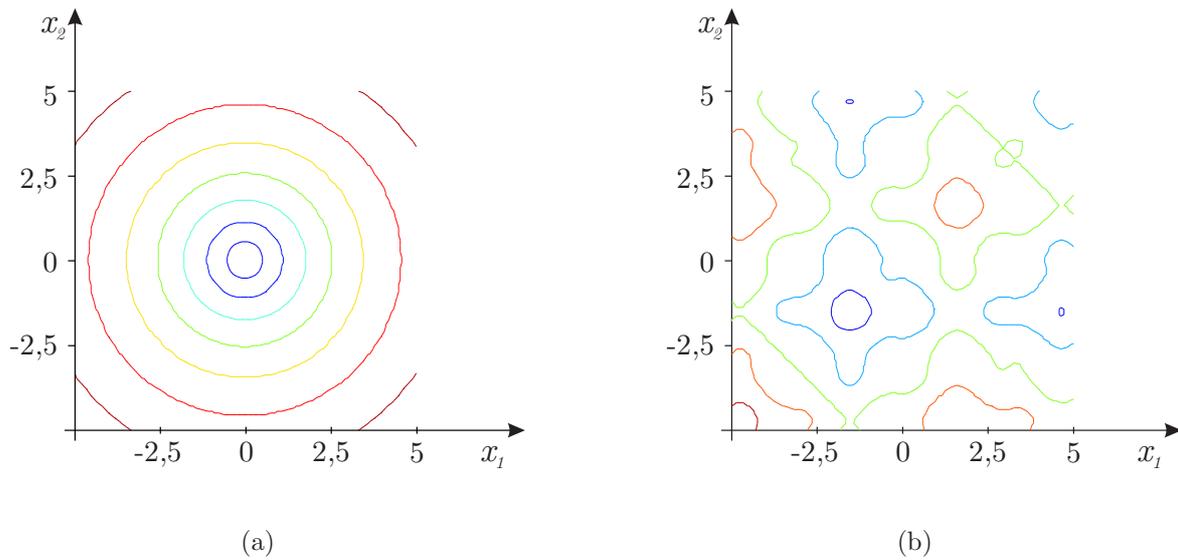


Abbildung 5.11: Höhenlinien der Testfunktion nach Kursawe bei Verwendung zweier Eingangsgrößen x_1 und x_2 , (a): $Q_1(x_1, x_2)$, (b): $Q_2(x_1, x_2)$.

Die Funktion Q_1 hat ihr globales Minimum bei $\check{x}_i = 0$ für $i = 1, \dots, n$ mit $Q_1(\check{\mathbf{x}}) = -10$. Das globale Minimum der Funktion Q_2 liegt bei $\check{x}_i = -1,5216$ für $i = 1, \dots, n$ mit $Q_2(\check{\mathbf{x}}) = -7,1656$. In Abbildung 5.11 sind die Höhenlinien beider Funktionen dargestellt.

Aufgrund der unterschiedlichen Lage der minimalen Funktionswerte für Q_1 und Q_2 im Parameterraum gibt es keine einzelne optimale Lösung, sondern es entsteht eine pareto-optimale Lösungsmenge. In Abbildung 5.12 ist eine aus der Literatur bekannte Approximation der Pareto-Menge dargestellt.

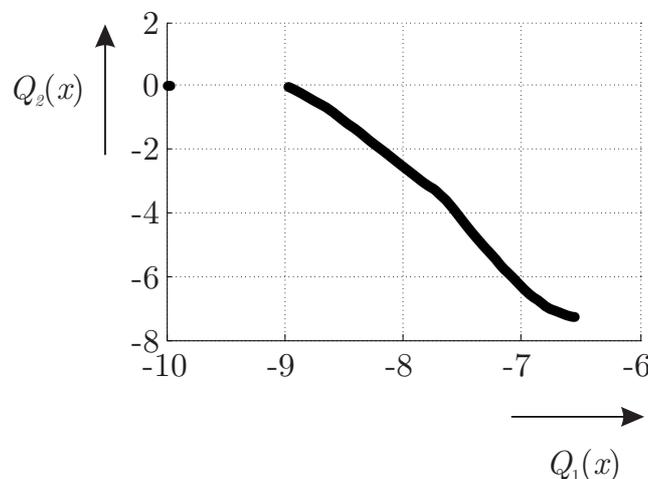


Abbildung 5.12: Approximation der Pareto-Front für Kursawes Testfunktion.

5.4.2 Bewertungskriterien

Im Folgenden werden wesentliche Kriterien beschrieben, die zur Bewertung der Effizienz der verschiedenen getesteten Evolutionsstrategien herangezogen werden können.

Anzahl der Fitnessfunktionsauswertungen

Ein sehr wichtiges und zugleich sehr einfaches Bewertungskriterium ist das Zählen der Fitnessfunktionsaufrufe. Um ein vergleichbares Maß über die Dauer der Suche eines Optimierungsdurchlaufs zu erhalten, ist es nicht sinnvoll, die Generationen zu zählen, nach denen das Optimum gefunden wurde. Diese Zahl wird stark beeinflusst von der Größe der Population, der Zahl der Nachkommen, der Zahl der Mutanten pro Nachkommen und weiteren Faktoren. Eine verlässlichere Aussage kann über das Zählen der Fitnessauswertungen getroffen werden. Bei der Optimierung eines technischen Problems ist die Fitnessfunktion in der Regel nicht bekannt. Die Fitness wird durch das Verhalten des technischen Systems auf die optimierten Parameter bestimmt. Dieser Vorgang dauert in der Regel länger als die Auswertung einer explizit gegebenen Funktion. Daher ist die Anzahl von Fitnessauswertungen, die zum Finden des Optimums notwendig sind, eine wichtige Größe bei der Bewertung von Optimierungsverfahren. Üblicherweise zieht man sie als wichtigstes Bewertungskriterium heran, was sinnvoll ist, wie folgendes Beispiel zeigt: Hat ein Optimierungsverfahren den Suchraum sehr gleichmäßig und breit abgesucht, ist dies nur dann als positiv zu werten, falls dafür nicht um ein Vielfaches mehr Fitnessfunktionsaufrufe durchgeführt wurden als bei einem anderen Optimierungsverfahren.

Entfernungen zu allen lokalen Minima

Dieses Bewertungskriterium bewertet den Abstand der im Suchraum „erforschten“ Punkte zu den lokalen und globalen Minima. Dieses Maß ist sehr gut zur Effizienzbewertung der Suchraumabdeckung durch das Optimierungsverfahren geeignet. Allerdings kann es nur bei Optimierungsproblemen verwendet werden, bei denen die Lage aller Minima bekannt ist. Dies ist nur bei den hier vorgestellten analytischen Testproblemen gegeben. Doch trotz dieser Einschränkung der Allgemeinheit ist das Maß ein geeignetes Mittel zur Beurteilung des Suchverhaltens von Optimierungsverfahren. Im Folgenden werden die Einzelheiten der Bewertung erläutert. Aus der Menge aller aufgezeichneten Fitnessfunktionsaufrufe eines vollständigen Optimierungsverlaufs wird zu jedem Minimum

derjenige ausgewählt, der minimalen Abstand zu diesem besitzt. Die Abstandswerte d_i werden gemäß

$$d'_i = e^{\lambda \cdot d_i} \quad (5.16)$$

nichtlinear umskaliert, wobei λ problemabhängig gewählt wird. Zusätzlich wird eine Gewichtung nach der Fitness der Minima vorgenommen. Das globale Minimum erhält das höchste Gewicht, alle anderen Minima erhalten linearverteilt schwächere Gewichte.

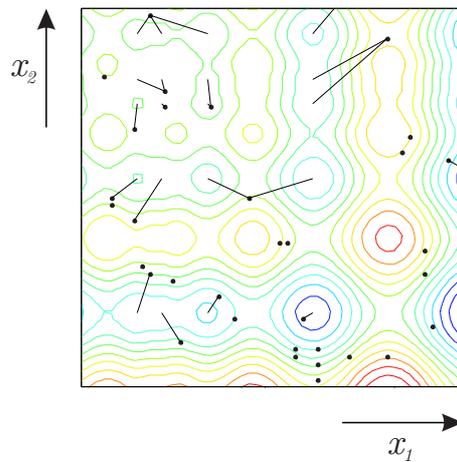


Abbildung 5.13: Abstandsbestimmung an einer Beispielfunktion.

Schließlich wird das Verhältnis zwischen der Summe der gewichteten umskalierten Abstandswerte und der Summe der Werte für einen Fitnessfunktionsaufruf in jedem Minimum gebildet. In der folgenden Gleichung wird g_i als umskalierter Fitnesswert des i -ten Minimums und d'_i als kleinster gewichteter Abstand eines Fitnessfunktionsaufrufs zu diesem Minimum gewählt.

Für n Minima gilt:

$$C_e = \frac{\sum_{i=1}^n g_i \cdot d'_i}{\sum_{i=1}^n g_i} \quad (5.17)$$

Das Ergebnis der Bewertung C_e wird den Wert 1 annehmen, wenn genau in jedem Minimum eine Fitnessfunktionsauswertung durchgeführt wurde. Abbildung 5.13 zeigt ein Beispiel im zweidimensionalen Raum. Die Orte der Fitnessfunktionsauswertungen sind durch Punkte gekennzeichnet. Die Abstände der ausgewählten Punkte zu den Minima sind durch Verbindungslinien dargestellt.

5.4.3 Ergebnisse

Dieser Abschnitt präsentiert die Ergebnisse, die bei simulativen Untersuchungen an den Testproblemen zu verschiedenen Fragestellungen erzielt wurden. Abschließend erfolgt eine Interpretation der vorgestellten Ergebnisse.

Effizienz zweidimensional

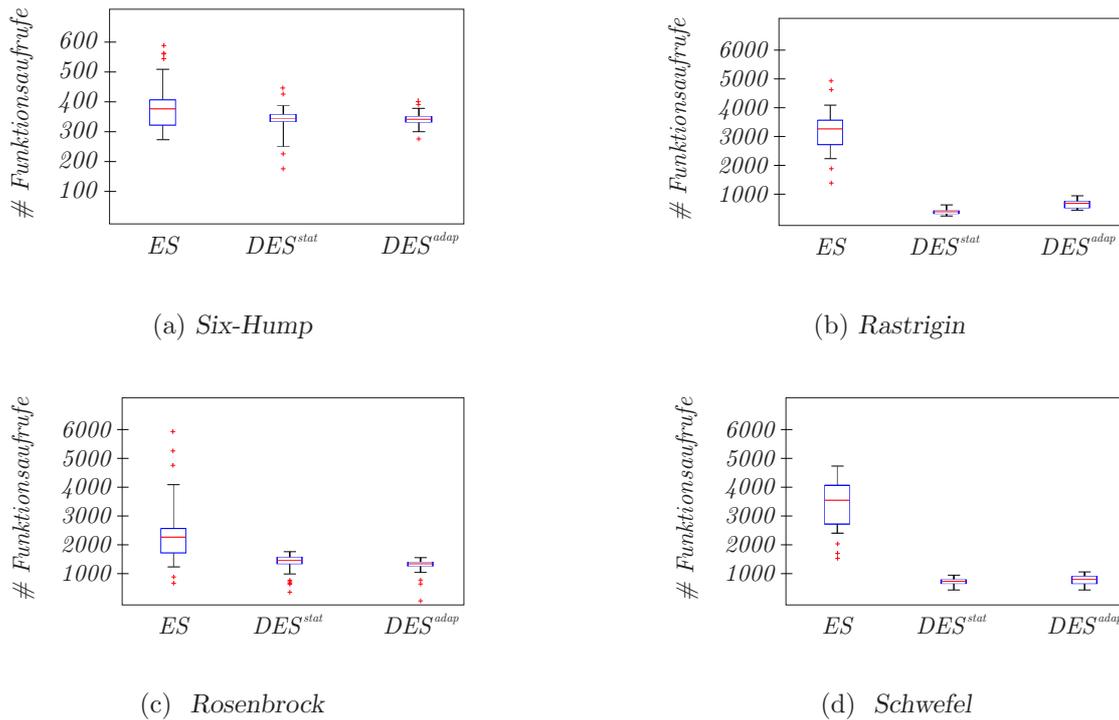


Abbildung 5.14: Vergleich der Effizienz der Varianten ES , DES^{stat} und DES^{adap} anhand zweidimensionaler Testfunktionen.

Die Ergebnisse dieser Untersuchung untermauern die Erwartungen in Bezug auf die Effizienzsteigerung. In Abbildung 5.14 ist die Anzahl der Fitnessfunktionsaufrufe bis zum Erreichen des Minimums der vorgestellten skalaren Testfunktionen mit aktivierter Populationsdiversifizierung für die Standard-Evolutionsstrategie ES , der statisch diskretisierten Variante DES^{stat} sowie der adaptiv diskretisierten Variante DES^{adap} dargestellt. Dabei sind die jeweils benötigte Anzahl von Funktionsaufrufen zur Ermittlung des Optimums im zweidimensionalen Suchraum von 1000 Optimierungsdurchläufen der Testfunktionen in Boxplots für die drei Varianten dargestellt. Die Anzahl der Elternindividuen wurde dabei jeweils auf 15 und die der Nachkommen auf 100 eingestellt. In jeweils einem der 1000 Durchläufe wurde zur Erhaltung der Vergleichbarkeit die gleiche Startpopulation für die getesteten Varianten der Evolutionsstrategie

verwendet. Die Startpopulation ist immer im vollständigen Suchraum zufällig verteilt erzeugt worden. Die statisch diskretisierte Evolutionsstrategie DES^{stat} wurde mit einer Rasterweite von 0,001 durchgeführt. Die Grenzwerte der Rasterweiten im Fall der adaptiv diskretisierten Evolutionsstrategie DES^{adap} wurden auf 0,0008 und 0,01 festgelegt.

Effizienz 24-dimensional

In analoger Weise sind in Anlehnung an ein industrielles Anwendungsproblem [NKBS01] die entsprechend ausgelegten analytischen Testfunktionen im 24-dimensionalen Suchraum analysiert worden.

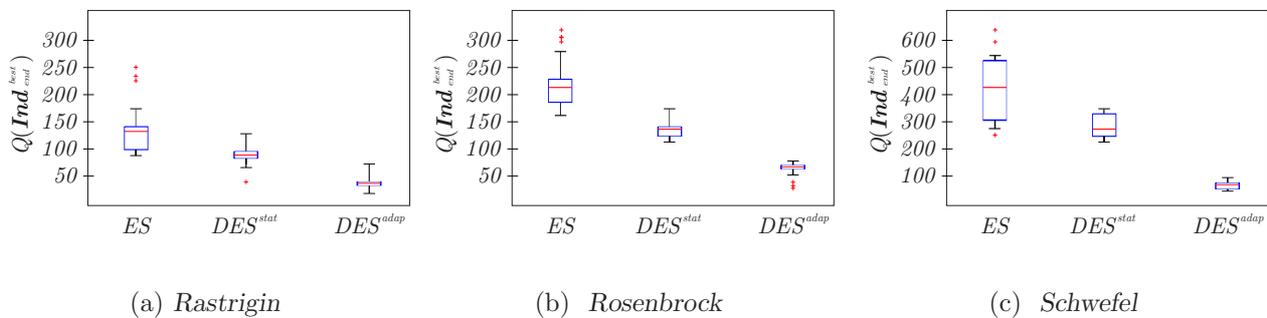


Abbildung 5.15: Vergleich der Effizienz der Varianten ES , DES^{stat} und DES^{adap} anhand 24-dimensionaler Testfunktionen.

Jeder einzelne der 1000 Testläufe ist zur Erhaltung der Vergleichbarkeit wie vorher mit gleicher Startpopulation für die getesteten Varianten ES , DES^{stat} und DES^{adap} durchgeführt worden. In Abbildung 5.15 sind in Boxplots die Fitnesswerte $Q(\mathbf{Ind}_{end}^{best})$ des jeweils besten Individuums eines Optimierungsdurchlaufs dargestellt, wobei jeder Optimierungsdurchlauf nach 10000 Fitnessauswertungen abgebrochen wurde.

Effizienz multikriteriell

In diesem Abschnitt wird ein Vergleich der adaptiv diskretisierten Evolutionsstrategie DES^{adap} und der kontinuierlichen Evolutionsstrategie ES anhand des multikriteriellen Testproblems von Kursawe im dreidimensionalen Suchraum vorgenommen. Die Effizienz der Varianten zeigt sich in der Approximationsgüte, mit der die Pareto-Front des Testproblems nach 2000 Fitnessfunktionsaufrufen in zwei exemplarischen Optimierungsdurchläufen dargestellt wird.

Für diesen Fall wird ein qualitativer Vergleich gemäß Abbildung 5.16 vorgenommen, worin die erzielte Approximation der Pareto-Front durch die adaptiv

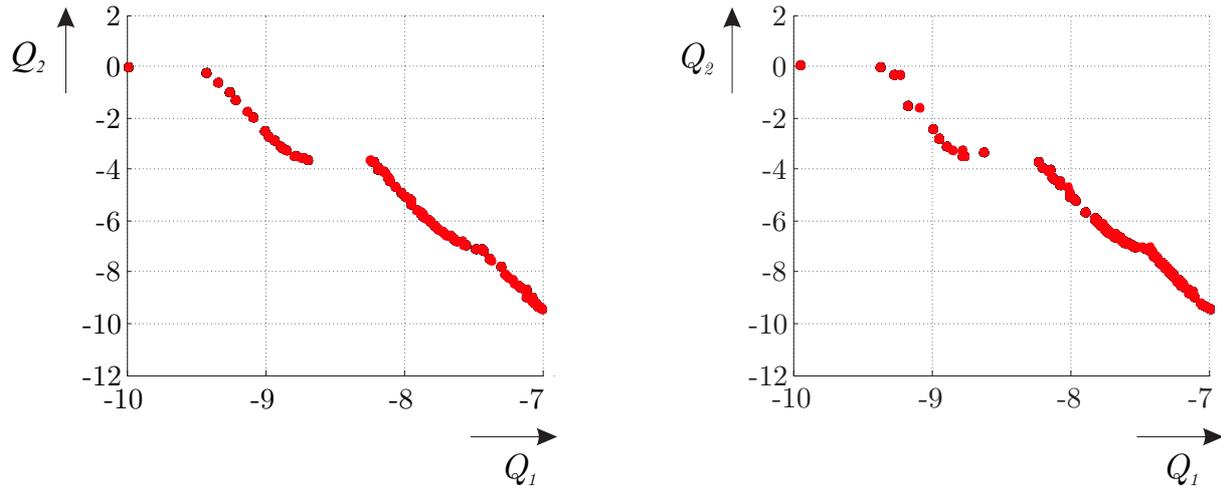


Abbildung 5.16: Gegenüberstellung der erzielten Approximationen der Pareto-Front.

diskretisierte Evolutionsstrategie DES^{adap} im linken Diagramm der Approximation durch die kontinuierliche Variante ES im rechten Diagramm gegenübergestellt wird. Es ist leicht ersichtlich, dass die Approximation durch die adaptiv diskretisierte Evolutionsstrategie Vorteile sowohl in der Gleichmäßigkeit der Verteilung als auch in den Fitnesswerten der einzelnen Pareto-Punkte bietet.

Gleichmäßigkeit der Abdeckung des Suchraums

Die Ergebnisse dieser Untersuchung zeigen die Gleichmäßigkeit mit der die Varianten ES und DES^{adap} den Suchraum abgesucht haben. Abbildung 5.17 zeigt die Orte der Fitnessauswertungen für einen exemplarischen Optimierungsdurchlauf mit der gleichen Startpopulation in beiden Fällen. Es ist leicht ersichtlich, dass die Fitnessfunktionsauswertungen der adaptiv diskretisierten Evolutionsstrategie DES^{adap} wesentlich gleichmäßiger über den Parameterraum verteilt sind. Durch die bessere Verteilung im Parameterraum erreicht die adaptiv diskretisierte Variante auch bei der Bewertung der Entfernung der zu den lokalen Minima jeweils nächsten durchgeführten Fitnessauswertungen einen besseren Wert.

Schlussfolgerungen

Die vorgestellten Untersuchungsergebnisse bestätigen, dass die diskretisierten Evolutionsstrategien der kontinuierlichen Variante in Bezug auf die Effizienz sowohl in nieder- als auch in hochdimensionalen Suchräumen überlegen sind. Dies bedeutet, dass sie entweder das gleiche Ziel mit weniger Aufwand oder mit gleichem Aufwand einen besseren Fitnessendwert erreichen.

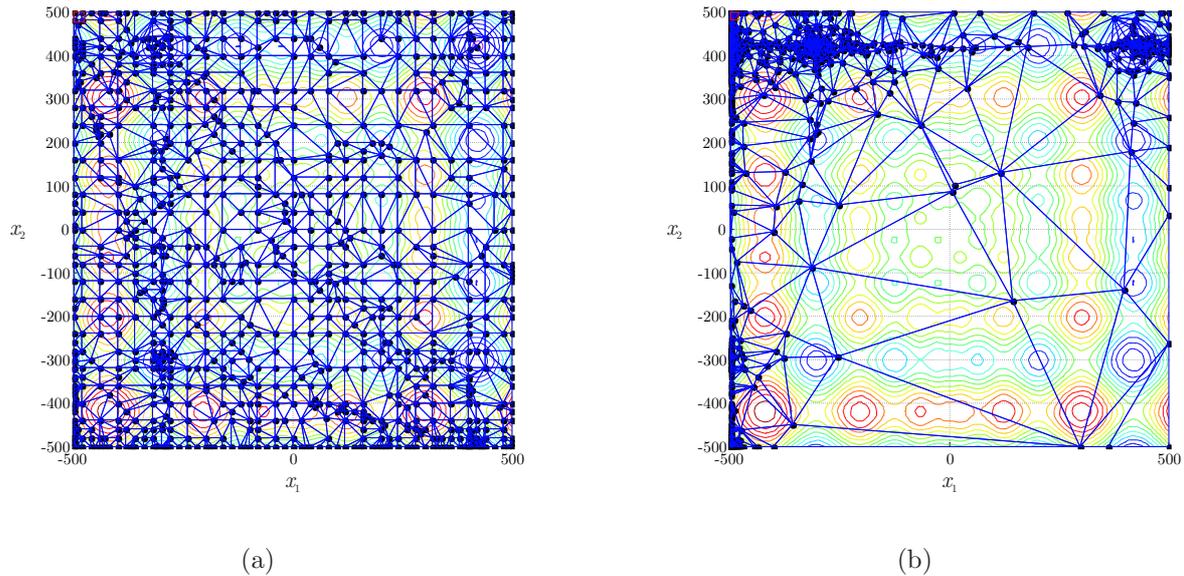


Abbildung 5.17: Illustration der erzielten Abdeckung im zweidimensionalen Suchraum, (a): DES^{adap} erreicht Optimum nach 975 Fitnessauswertungen, (b): ES erreicht Optimum nach 2361 Fitnessauswertungen.

Im Fall der multikriteriellen Optimierungsprobleme zeigt sich zudem, dass bei gleichem Aufwand in Form der Anzahl von Fitnessfunktionsaufrufen durch die adaptiv diskretisierte Evolutionsstrategie eine bessere Approximation der Pareto-Front erreicht wird.

Die Gleichmäßigkeit der Abdeckung des Suchraums ist wesentlich höher, wenn eine diskretisierte Evolutionsstrategie verwendet wird. Damit steigt die Wahrscheinlichkeit, weitere und unter Umständen bessere Optima der Fitnessfunktion zu finden. Darüber hinaus ergibt sich, dass der Aufwand für die Adaption der Diskretisierung einen Vorteil der adaptiv diskretisierten Evolutionsstrategie gegenüber der statisch diskretisierten Variante bei niederdimensionalen Optimierungsproblemen verhindert. Ein plausibler Grund dafür ist, dass die feste Rasterweite bei der statisch diskretisierten Evolutionsstrategie im Laufe der Zeit empirisch auf einen sehr günstigen Wert eingestellt wurde. Da ein großer Aufwand für das Finden einer geeigneten Rasterweite in der Regel nicht betrieben wird, relativiert sich die vorherige Aussage. Bei höherdimensionalen Problemen ergibt sich zudem ohnehin ein deutlich erkennbarer Effizienzvorteil für die adaptiv diskretisierte Variante.

Kapitel 6

Komplexe Anwendungen

Im Folgenden wird zunächst ein Überblick über die Verbesserungen, die im Vorfeld mit dem Benchmarkproblemen erhalten worden sind, gegeben. Anschließend wird auf die Leistungssteigerungen bei industriellen Anwendungen eingegangen, die im Rahmen dieser Arbeit durchgeführt worden sind. Die wesentliche Zielsetzung besteht dabei darin, die Leistungssteigerungen aufgrund der neuentwickelten Strategieelemente hinsichtlich der Komplexitätsreduktion und Steigerung der Interpretierbarkeit sowie der erreichbaren Modellierungsgüte aufzuzeigen.

6.1 Benchmarkprobleme

Die im Rahmen der Arbeit von SLAWINSKI erstellte systematische Vorgehensweise bei der Modellierung mit dem Fuzzy–ROSA–Verfahren ist auf verschiedenen Benchmarkproblemen getestet worden. Die damit erzielten Ergebnisse FRM_{sys} sind dabei in Relation zu dem jemals besten mit dem Fuzzy–ROSA–Verfahren erzielten Ergebnis FRM_{best} gestellt worden. Durch den vorgelagerten Einsatz der im Rahmen dieser Arbeit konzipierten evolutionären Merkmalsselektion sind teilweise deutlich verbesserte Ergebnisse FRM_{EFS} erzielt worden, die im Folgenden entsprechend der aus der Literatur Bekannten eingeordnet werden.

6.1.1 Klassifikation von Schwertlilien (IRIS)

Dieses Klassifikationsproblem betrifft den als Benchmarkproblem in der Literatur häufig verwendeten IRIS–Datensatz [And35, PPB97]. Die Aufgabe besteht darin, auf Grundlage von vier Einflussgrößen (*Kelchblattlänge*, *Kelchblattbreite*,

Blütenblattlänge und *Blütenblattbreite*) die zugehörige Schwertlilienart (*Setosa*, *Versicolor* oder *Virginica*) zu bestimmen.

Die vorhandenen 150 Datenpunkte wurden zufällig in zwei Datensätze gleicher Größe für eine Kreuzvalidierung geteilt. Mit einer gewählten Anzahl von $k = 5$ Stützstellen wurde für jede der vier Einflussgrößen eine Einteilung in vier Intervalle vorgegeben. Der iterative Selektionsprozess terminierte mit $\eta_3 = 0$ und lieferte als relevanten Satz $\mathcal{M}_3 = \{\textit{Blütenblattlänge}, \textit{Blütenblattbreite}, \textit{Kelchblattbreite}\}$.

Nach einer automatischen Generierung der Fuzzy-Zugehörigkeitsfunktionen konnte mit dem Fuzzy-ROSA-Verfahren ein Klassifikator mit einer Regelbasis von neun Regeln generiert werden, der sowohl auf Lerndaten als auch auf Validierungsdaten in jedem Fall korrekt klassifiziert ($\epsilon_{\textit{vali}} = \text{RCE} = 0$). Der erzielte relative Klassifikationsfehler auf Validierungsdaten $\epsilon_{\textit{vali}}$ wird in Abbildung 6.1 mit den Ansätzen aus [Hol98] verglichen. Zum Vergleich sei angemerkt, dass die besten bisher bekannten publizierten Ergebnisse¹ alle vier Einflussgrößen verwenden und dennoch auf Validierungsdaten relative Klassifikationsfehler im Bereich von $\text{RCE} > 0,02$ aufweisen. Mit einer Regelanzahl von $R = 9$ ist zudem die Interpretierbarkeit des mit $\text{FRM}_{\textit{EFS}}$ erhaltenen Fuzzy-Modells in einem sehr hohen Maße gewährleistet.



Abbildung 6.1: Relativer Klassifikationsfehler $\epsilon_{\textit{vali}}$ auf Validierungsdaten in Prozent für IRIS bei folgenden Ansätzen: \diamond $\text{FRM}_{\textit{EFS}}$, ∇ $\text{FRM}_{\textit{sys}}$, \circ $\text{FRM}_{\textit{best}}$ und \times Literatur.

6.1.2 Bestimmung der Herkunft von Weinen (WINE)

Der WINE-Datensatz [CS94] betrifft die chemische Analyse von Weinen, die aus der gleichen Region Italiens stammen, aber von unterschiedlichen Weinbauern angebaut wurden. Das Ergebnis der chemischen Analysen ist jeweils die gefundene Menge von 13 Inhaltsstoffen, anhand derer die Herkunft der Weine klassifiziert werden soll.

Der verfügbare Datensatz von 178 Datenpunkten wurde zufällig in zwei Datensätze von je 89 Datenpunkten für eine Kreuzvalidierung geteilt. Mit einer vorgegebenen Anzahl von $k = 6$ Stützstellen für jede Einflussgröße terminierte

¹UCI Repository of Machine Learning Databases:
<http://www.ics.uci.edu/~mlearn/MLRepository.html>

in vier Intervalle je Einflussgröße vorgegeben. Wie bei allen Klassifikationsproblemen wurde als Relevanzmaß der relative Klassifikationsfehler (4.26) benutzt. Die Definition der Partialfunktionen erfolgte durch Ermittlung der am häufigsten vertretenen Klasse innerhalb eines mit Datenpunkten besetzten Hyperquaders nach Gleichung 4.21 aus Abschnitt 4.5.2. Der iterative Selektionsprozess lieferte $\mathcal{M}_{11} = \{X_{29}, X_{30}, X_{32}, X_{31}, X_{28}, X_{35}, X_{21}, X_{33}, X_{44}, X_{22}, X_3\}$ als Satz relevanter Einflussgrößen mit $\eta_{11} = 0,007$. Bemerkenswert dabei ist, dass die zuerst ermittelten — und damit relevantesten — fünf Einflussgrößen sich auf Elemente um die Mitte der Nukleotidsequenz beziehen ($X_{29}, X_{30}, X_{32}, X_{31}, X_{28}$). Am wenigsten relevant dagegen sind die Einflussgrößen, die sich auf die äußeren Bereiche der Sequenz beziehen.



Abbildung 6.3: Relativer Klassifikationsfehler ϵ_{vali} auf Validierungsdaten in Prozent für GENE bei folgenden Ansätzen: \diamond FRM_{EFS} , ∇ FRM_{sys} , \circ FRM_{best} und \times Literatur.

Für die nachgeschaltete Modellierung mit dem Fuzzy–ROSA–Verfahren wurden entsprechend der kodierten symbolischen Werte für jede ermittelte Einflussgröße aus \mathcal{M}_{11} vier und für die drei Klassen der Ausgangsgröße drei Singletons definiert. Die vorhandenen 3175 Datenpunkte wurden in einen Lerndatensatz mit 1588 Datenpunkten und einen Validierungsdatensatz mit 1587 Datenpunkten zufällig aufgeteilt. Als Ergebnis konnte ein Fuzzy–Klassifikator mit 141 Regeln generiert werden, der auf Lerndaten einen Klassifikationsfehler von $RCE = 0,033$ und für die Validierungsdaten einen Klassifikationsfehler von $\epsilon_{vali} = RCE = 0,058$ aufweist. Der hier erzielte relative Klassifikationsfehler auf Validierungsdaten ϵ_{vali} wird in Abbildung 6.3 mit den Ansätzen aus [MST94] verglichen.

Im Vergleich zu 33 anderen publizierten Ansätzen [Kro99, MST94] wird hiermit bezüglich des Klassifikationsfehlers auf Validierungsdaten das viertbeste Ergebnis erzielt. Die Dimensionalität des Modellierungsproblems wurde zudem von 60 auf 11 zu berücksichtigende Einflussgrößen reduziert. Damit konnte zusätzlich im Vergleich zu anderen mit dem Fuzzy–ROSA–Verfahren erzielten Ergebnissen eine deutlich kompaktere Regelbasis generiert werden. Zusätzlich ist hier anzumerken, dass der Referenzregelsatz [Kro99] aus einem aufwändigen und nicht deterministischen Regelgenerierungsprozess resultiert, der kaum ohne Expertenwissen vorgenommen werden kann.

6.1.4 Klassifikation der Erdoberfläche in Satellitenbildern (SAT)

Der Originaldatensatz „*Landsat Satellite Image*“ kommt von der NASA und enthält die Intensität von vier Spektralbändern für 82×100 Pixel. Ein Pixel entspricht dabei einer Fläche von ungefähr 80×80 Meter. Die Originalbezeichnungen für die Bodentypen an der Erdoberfläche sind: *Red Soil*, *Cotton Crop*, *Gray Soil*, *Damp Gray Soil*, *Soil with Vegetation Stubble*, *very Damp Gray Soil* und *Mixture*. Die Aufgabe³ besteht darin, den Bodentyp eines Pixels in Satellitenbildern zu klassifizieren. Als potenzielle Einflussgrößen stehen Intensitäten der Spektralbänder des Pixels selbst und die seiner acht Nachbarpixel zur Verfügung.

Der verfügbare Datensatz von 6435 Datenpunkten wurde zufällig in zwei gleichgroße Datensätze für eine Kreuzvalidierung geteilt. Mit einer vorgegebenen Anzahl von $k = 4$ Stützstellen für jede Einflussgröße lieferte der iterative Selektionsprozess $\mathcal{M}_7 = \{X_{19}, X_{18}, X_{21}, X_{11}, X_{16}, X_{26}, X_{35}\}$ als Satz relevanter Einflussgrößen mit $\eta_7 = 9, 2$.

Darauf aufsetzend wurden unter Verwendung automatisch generierter Fuzzy-Zugehörigkeitsfunktionen mit dem Fuzzy-ROSA-Verfahren für beide Datensätze Klassifikatoren mit 149 bzw. 161 Regeln generiert. Die resultierenden Klassifikationsfehler beider Klassifikatoren ergeben sich auf Lerndaten im Mittel zu $RCE = 11, 1$ sowie zu $\epsilon_{vali} = RCE = 15, 1$ auf Validierungsdaten. Der relative Klassifikationsfehler auf Validierungsdaten ϵ_{vali} wird in Abbildung 6.4 mit den Ansätzen aus [MST94] verglichen.

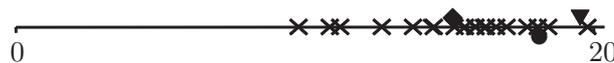


Abbildung 6.4: Relativer Klassifikationsfehler ϵ_{vali} auf Validierungsdaten in Prozent für SAT bei folgenden Ansätzen: \diamond FRM_{EFS} , ∇ FRM_{sys} , \circ FRM_{best} und \times Literatur.

Mit dem bisherigen besten Ergebnis FRM_{best} und dem Ergebnis bei systematischem Ansatz FRM_{sys} werden in diesem Anwendungsbeispiel nur hintere Plätze bezüglich der Modellierungsgüte belegt.

6.1.5 Vorhersage der Mackey–Glass–Zeitfolge (MACKEY)

Die chaotische Mackey–Glass–Zeitfolge [MG77] wird in der Literatur häufig als Benchmarkproblem für Approximationsverfahren genutzt. Die diskrete Version

³UCI Repository of Machine Learning Databases: <http://www.ics.uci.edu/~mllearn/MLRepository.html>

der Mackey–Glass–Zeitfolge lässt sich beschreiben durch

$$x(t+1) = (1-a)x(t) + \frac{bx(t-\tau)}{1+x^{10}(t-\tau)}. \quad (6.1)$$

Als Parameter werden $a = 0,1$, $b = 0,2$ und $\tau = 17$, mit $x(t) = 0$ für $t < 0$ und $x(0) = 1,2$ gewählt. Die Aufgabe besteht darin, auf Basis der Werte $x(t-18)$, $x(t-12)$, $x(t-6)$ und $x(t)$ den Wert $x(t+6)$ vorausszusagen. Für das hier vorgestellte Verfahren wurden 1000 Datenpunkte zufällig in zwei Hälften zu je 500 Datenpunkten für eine Kreuzvalidierung geteilt. Für die evolutionäre Granularisierung wurde mit $k = 6$ Stützstellen für jede der vier Einflussgrößen eine Einteilung in fünf Intervalle vorgegeben. Als Relevanzmaß wurde der mittlere quadratische Fehler (4.29) benutzt. In Hinblick auf eine nachgeschaltete Modellierung von TSK–Fuzzy–Systemen wurden die Partialfunktionen durch lineare Approximation nach Gleichung 4.23 aus Abschnitt 4.5.2 definiert. Der iterative Selektionsprozess ermittelte den Satz $\mathcal{M}_3 = \{x(t-18), x(t-12), x(t-6)\}$.

Auf Grundlage dieses Satzes von Einflussgrößen wurden unter Berücksichtigung der optimierten Granularisierungen — zur Generierung der Fuzzy–Zugehörigkeitsfunktionen — TSK–Fuzzy–Systeme mit dem Fuzzy–ROSA–Verfahren entworfen. Diese enthalten je etwa 120 Regeln und liefern im Mittel einen quadratischen Fehler von $\epsilon_{vali} = \text{MSE} = 0,58 \cdot 10^{-4}$ sowie einen mittleren absoluten Fehler von $\text{MAE} = 54 \cdot 10^{-4}$ auf Validierungsdaten. Der mittlere quadratische Modellierungsfehler auf Validierungsdaten ϵ_{vali} wird in Abbildung 6.5 mit den Ansätzen aus [Kra00] verglichen.



Abbildung 6.5: Mittlerer quadratischer Fehler ϵ_{vali} auf Validierungsdaten für MACKEY bei folgenden Ansätzen: \blacktriangledown FRM_{EFS} , \blacktriangledown FRM_{sys} , \circ FRM_{best} und \times Literatur.

Das bisher beste bekannte Ergebnis [Kra00] ($\text{MSE} = 1,4 \cdot 10^{-4}$) kann damit um 59% verbessert werden. Zusätzlich wird die Komplexität des Modellierungsproblems um eine Dimension reduziert.

6.1.6 Modellierung von Wohnungspreisen in Boston (BOSTON)

Das Boston Housing Problem [HR78] hat zwei Zielsetzungen: Zum einen soll ein Modell für den *Wohnungspreis* in der Gegend von Boston anhand einiger ausgewählter Merkmale erstellt, zum anderen ein Modell für die Schadstoffbelastung anhand der gleichen Merkmale bestimmt werden. Hier wird nur die

6.2 Industrielle Anwendungen

6.2.1 Kurzfristige Lastprognose für die Kraftwerksführung

Die zentralen Aufgaben der Energieversorgung bestehen in der *Energieerzeugung* und der *Netzbetriebsführung* für eine ausreichende und wirtschaftliche Bereitstellung sowie den sicheren Transport von Energie in einem Versorgungsgebiet. Vor der Liberalisierung des Energiemarktes konnten diese beiden Aufgabenstellungen durch eine zentrale Kraftwerksführung erfüllt werden. Für die Kraftwerksführung wurde dazu ein sogenanntes Kraftwerksführungssystem eingesetzt, das die Aufteilung der Energieerzeugung auf die momentan im Einsatz befindlichen Kraftwerke vornahm (vgl. [Alb97, AAS96, AKAS96]). Dabei sollten Nebenbedingungen wie schonender und kostengünstiger Einsatz der Kraftwerke und eine möglichst gute Übereinstimmung zwischen der erzeugten und der tatsächlich nachgefragten Leistung eingehalten werden. Abbildung 6.7 zeigt schematisch die Wirkungskette eines Kraftwerksführungssystems innerhalb eines Energieversorgungsnetzes.

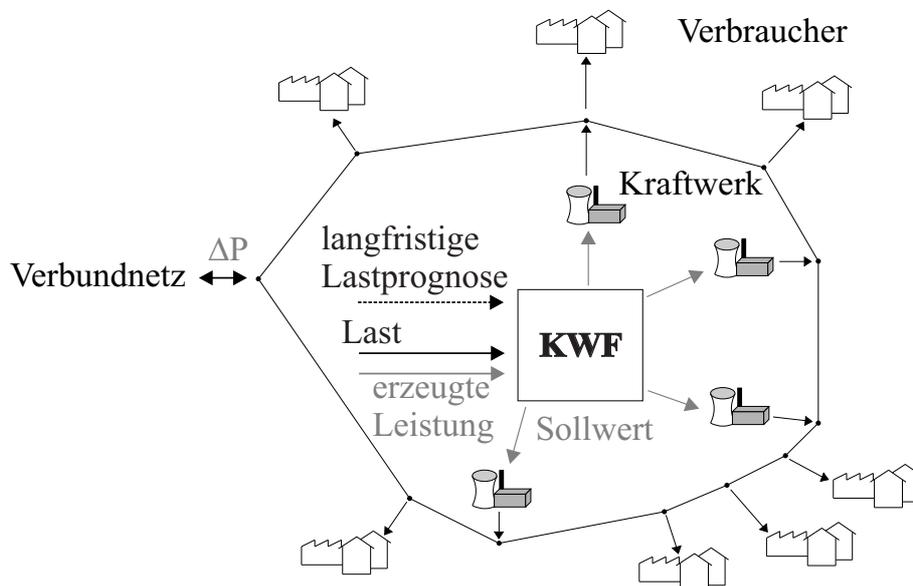


Abbildung 6.7: Kraftwerksführungssystem (KWF) zur Steuerung eines Kraftwerkparkes innerhalb eines Energieversorgungsnetzes.

Da elektrische Energie nicht in ausreichendem Umfang speicherbar ist, muss zur Erhaltung eines kontinuierlichen Leistungsgleichgewichts die Erzeugung der Energie an die jeweils aktuell geforderte Leistung angepasst werden. Damit eine hohe Versorgungssicherheit in Europa garantiert werden kann, waren und

sind die Netze der verschiedenen Energieversorgungsunternehmen miteinander gekoppelt. Für einen möglichst wirtschaftlichen Betrieb des Energieversorgungsprozesses bestand die Notwendigkeit, die Austauschleistung zwischen dem eigenen Netz und den Netzen anderer Unternehmen möglichst gering zu halten. Dieses Ziel war die zentrale Aufgabenstellung für das Kraftwerksführungssystem.

Die wesentlichen Einflussgrößen, auf denen die Steuerungsstrategien des Kraftwerksführungssystems basieren, sind zum einen die *erzeugte* und tatsächlich *verbrauchte Leistung* und zum anderen eine *Lastprognose* für verschiedene Zeithorizonte. Die Lastprognose ist dabei ein sehr wesentlicher Bestandteil für die Entscheidungsfindung und stützt sich auf historisch aufgezeichnete Daten. Zur Verbesserung einer in einem Energieversorgungsunternehmen eingesetzten Lastprognose sind verschiedene Ansätze zur Entwicklung eines Prognosemodells gemacht worden [Pra99, Jes00a, Jes00b].

In der Arbeit [Jes00a] ist die Aufgabenstellung der Verbesserung der Lastprognose intensiv bearbeitet worden. Neben der Entwicklung von Methoden zur verbesserten datenbasierten Generierung von Fuzzy-Modellen bei verrauschten Lerndaten, wie im Falle der Lastprognose, ist dabei auch der Einsatz von ARMA-Modellen⁵ für die Lastprognose untersucht worden. Die in Tabelle 6.2 aufgeführten Ergebnisse zeigen jedoch, dass nur Modellierungsfehler oberhalb des Wertes des Referenzsystems und somit kein zufriedenstellendes ARMA-Modell erzielt werden konnte. In weiteren Arbeiten innerhalb des *Sonderforschungsbereichs 531* wurde zum Vergleich auch ein neuronales Netz für die Lastprognose trainiert. Die Fehlerwerte liegen nur knapp unter denen des Referenzsystems und es konnte auch hier keine angemessene Verbesserung erzielt werden. Durch den Einsatz des Fuzzy-ROSA-Verfahrens mit dem mittelwertbasierten Bewertungsindex (Definition B.5) konnte ein Prognosemodell erzeugt werden, das einen deutlich niedrigeren mittleren Fehler als das Referenzsystem aufweist. Die dazu teilweise aufwändigen notwendigen Schritte sind in [Jes00a] beschrieben. Der wesentliche Aufwand bestand dabei in der Definition einer geeignet kaskadierten Fuzzy-Modell-Struktur in Verbindung mit zeitintensiven empirischen Untersuchungen zur Selektion geeigneter Merkmalsätze von potenziell relevanten Einflussgrößen.

Für den Einsatz der im Rahmen dieser Arbeit konzipierten Verfahren für das hier betrachtete Modellierungsproblem stehen beispielhaft Daten der tatsächlich verbrauchten Leistung des Jahres 1995 in 15-Minuten-Schritten gerastert zur Verfügung (35040 Datenpunkte). Durch eine entsprechende Daten-

⁵ARMA-Modelle stellen eine Kombination von Autoregressiven Modellen (AR) und Moving-Average-Modellen (MA) dar.

aufbereitung weiterer archivierter Daten ist ein Satz von 32 potenziellen Einflussgrößen, wie beispielsweise *Tageszeit*, *aktuelle Temperatur*, *Bewölkung*, *aktueller Lastverbrauch*, *Jahreszeit* und *Tagestyp* (Werktag, Wochenende) erstellt worden. Die Modellausgangsgröße besteht aus der zu prognostizierenden Änderung der tatsächlich verbrauchten Leistung (*Laständerung*) für den zukünftigen Datenpunkt, also in 15 Minuten. Dabei sind als Lerndaten die Datenpunkte der ersten Hälfte des Jahres und als Validierungsdaten die der zweiten Hälfte verwendet worden. Aus den vielen verschiedenen möglichen Merkmalsätzen liefert der iterative Selektionsprozess sowohl mit dem Ansatz zur Generierung eines Mamdani-Fuzzy-Modells als auch für den Ansatz zur Generierung eines TSK-Fuzzy-Modells den Satz $\mathcal{M}_3 = \{Tageszeit, Tagestyp, Jahreszeit\}$ zur Prädiktion der zu erwartenden Laständerung innerhalb der nächsten 15 Minuten. Die bei einer darauf aufsetzenden Modellierung mit dem Fuzzy-ROSA-Verfahren jeweils erzielbaren Ergebnisse sind in Tabelle 6.2 dargelegt.

Tabelle 6.2: Mittlerer absoluter Prognosefehler (MAE) auf Lerndaten (ϵ_{learn}) und Validierungsdaten (ϵ_{vali}) für mit unterschiedlichen Verfahren generierte Prognosemodelle sowie für das tatsächliche Referenzsystem.

Modelltyp	Verfahren	MAE in MW	
		ϵ_{learn}	ϵ_{vali}
Eingesetztes Referenzsystem	Statistik	46.3	
Alternative Ansätze	ARMA-Modell	49.2	47.9
	neuronales Netz	44.0	45.1
kaskadiertes Mamdani-Fuzzy-Modell	Empirik, Fuzzy-ROSA-Verfahren	40.9	40.9
Mamdani-Fuzzy-Modell	iterativer Selektionsprozess		
	Fuzzy-ROSA-Verfahren	40.1	40.3
TSK-Fuzzy-Modell	iterativer Selektionsprozess		
	Fuzzy-ROSA-Verfahren	38.7	39.2

Durch den vorgeschalteten Einsatz des Verfahrens zur Merkmalsselektion und der optimiert generierten Fuzzy-Partitionierungen des Eingangsgrößenraumes werden mit dem ansonsten gleich parametrisierten Fuzzy-ROSA-Verfahren deutlich verbesserte Modelle generiert. In Abbildung 6.2.2 werden die mit diesem Ansatz erzielten besten Ergebnisse mit denen des Referenzsystems der VEW AG verglichen. Dabei konnte der mittlere absolute Fehler der Prognose von gut 46 MW auf ca. 39 MW reduziert werden. Zur damaligen Zeit bestand zur Quantifizierung des Kosteneinsparpotenzials folgende Faustregel: Eine Verbesserung der kurzfristigen Lastprognose um im Mittel 1 MW bedeutet eine Kosteneinsparung von etwa 1.000.000 DM.

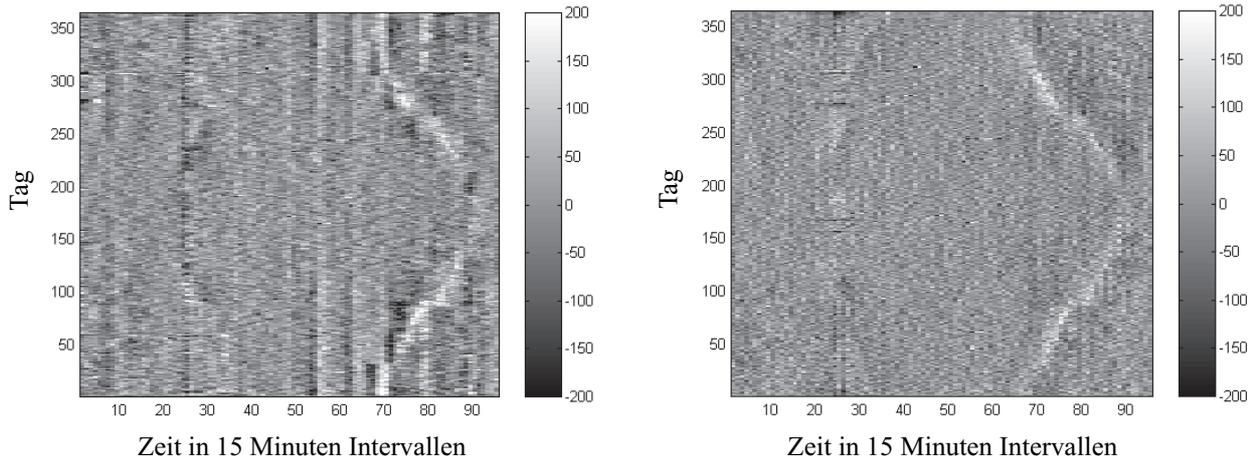


Abbildung 6.8: Prognosefehler in MW des Referenzsystems (links) und des Fuzzy-Prognosesystems (rechts); Abzisse: Mittelwert über 15 Minuten Intervalle für alle 365 Tage des Jahres 1995; Ordinate: Positive/negative Prognosefehler werden durch helle/dunkle Grautöne dargestellt.

6.2.2 Adaption von Strategieparametern der Kraftwerksführung

Innerhalb des im vorigen Abschnitt vorgestellten Kraftwerksführungssystems übernimmt ein *Optimierungsmodul* die Aufgabe, die Sollwerte für die einzelnen Kraftwerke unter Berücksichtigung des prognostizierten Leistungsbedarfs so einzustellen, dass die erforderliche Leistung insgesamt unter Einsatz möglichst geringer Brennstoffkosten erbracht und eine übermäßige Beanspruchung der Kraftwerke durch Vorgabe möglichst ruhiger Sollwertverläufe vermieden wird. Eine besondere Schwierigkeit dieser Prozessführungsaufgabe liegt darin, dass der betrachtete Kraftwerkspark aus unterschiedlichen Kraftwerkstypen besteht, die sich hinsichtlich ihrer spezifischen Erzeugungskosten und dynamischen Eigenschaften stark unterscheiden. Deshalb kann die elektrische Energie durch bevorzugten Einsatz der kostengünstigen Kraftwerke um so wirtschaftlicher erzeugt werden, je frühzeitiger und genauer das Optimierungsmodul über zukünftige Leistungsänderungen anhand der Lastprognose informiert wird. Ebenso wirkt es sich kostensenkend aus, wenn die für die Regelung vorgehaltene Leistungsreserve nicht unnötig groß bemessen wird.

Zur Lösung dieser komplexen Kraftwerksführungsaufgabe wurde in Vorprojekten ein neues Verfahren entwickelt, das gegenüber dem Stand der Technik wesentliche Vorzüge aufweist [AAS96, Alb97, AKM⁺95, KA97]. Ein zentrales Element des Verfahrens besteht darin, dass die für die Regelung vorzuhaltende Reserve nicht anteilmäßig einzelnen Kraftwerken fest zugewiesen wird, sondern stattdessen die Forderung nach einer summarischen Reserve (*Regelband*) erhoben und diese situationsabhängig den beteiligten Kraftwerken zugewiesen wird. Dieses Konzept der summarischen Reserve schafft eine größere Flexibilität, in

der ein wesentliches Kosteneinsparpotenzial liegt. Die eigentliche Optimierung, d. h. die Aufteilung der Prognose auf die einzelnen Kraftwerke unter Freihaltung der Reserve, wird durch ein vorausschauendes Verfahren gelöst, das mit Rücksicht auf die erforderliche Onlinefähigkeit suboptimal arbeitet. Darüber hinaus verfügt das Verfahren über eine Reihe von Strategieparameter (wie beispielsweise die Bemessung der vorzuhaltenden summarischen Reserve), die nach dem damaligen Entwicklungsstand des Verfahrens auf solche Werte fest oder nach einem Fahrplan eingestellt wurden, die sich im Mittel als günstig erwiesen hatten.

Ausgangspunkt zur Adaption von Strategieparametern des Kraftwerksführungssystems war die Überlegung, dass die Prozessexperten vor Ort aufgrund langjähriger Erfahrung die Prozessführungsgüte in unterschiedlichen Prozesssituationen intuitiv bewerten und daraufhin Vorschläge darüber machen können, wie die Werte bestimmter Strategieparameter zur Verbesserung der Prozessführungsgüte zu verändern sind. Dies betraf insbesondere drei Strategieparameter:

Größe des Regelbandes: Vorgabe der mindestens vorzuhaltenden summarischen Leistungsreserve zur Ermöglichung des Eingriffs der *Leistungsfrequenzregelung*. Durch diese Vorgabe wird der Spielraum für die kostenoptimale Lastaufteilung vorgegeben.

Anpassung der Basisprognose: Verbesserung der Genauigkeit der langfristigen Lastprognose durch eine geeignete Offsetverschiebung der Basisprognose.

Zusatzreserve: Vergrößerung der verfügbaren Leistungsreserve durch Vorverlegung bzw. Zurückverlegung von An- und Abfahrtzeiten bestimmter Kraftwerke.

Das letztendlich eingeschlagene Modellierungskonzept bestand darin, ohne den Umweg über eine Prozessführungsgüte direkt Handlungsvorschläge der Prozessexperten hinsichtlich einer situationsabhängigen Adaption dieser drei Strategieparameter zu modellieren. Zur datenbasierten Modellierung hochdimensionaler Systeme ist bei der Datenakquisition durch Expertenbefragung zu berücksichtigen, dass der Prozess durch eine große Zahl von Prozessgrößen, die ein Experte nicht simultan erfassen kann, gekennzeichnet ist. Deshalb muss eine sinnvolle Auswahl von möglichst wenigen Prozessgrößen, die die Prozesssituationen dennoch angemessen beschreiben, getroffen werden. Zudem müssen diese Größen so präsentiert werden, dass alle Situationen als Grundlage für den Entscheidungsprozess überschaubar und bewertbar sind. Beim vorliegenden Kraftwerksführungssystem stand eine Datenbasis zur Verfügung, die Werte

von näherungsweise 150 Prozessgrößen enthielt. Diese Größen waren Soll- und Istwerte der Kraftwerke, innere Systemgrößen des Kraftwerksführungssystems, Trendgrößen bezüglich zeitlich zurückliegender Werte sowie zukünftiger Prognosewerte und kalendarische Größen wie Tageszeit und Tagestyp.

Zur Informationskondensation wurden alle im Minutentakt gemessenen Prozessgrößen soweit möglich reduziert, beispielsweise, indem artverwandte Prozessgrößen wie solche, die zu ähnlichen Kraftwerkstypen gehören, zu einer einzigen Größe zusammengefasst wurden. Ferner wurden Daten durch Mittelwertbildung über Stundenintervalle konzentriert und es wurde für Trendgrößen ein maximaler Zeithorizont von zwei Stunden festgelegt. Auf diese Weise wurde eine reduzierte Datenbasis, die sich nur noch auf 47 Prozessgrößen bezieht, erzeugt. Diese wurde in einer grafischen Benutzeroberfläche, die sich an vertraute Flussdiagramme des vorhandenen Prozessleitsystems anlehnt, präsentiert. Insgesamt wurden 882 Betriebssystemen in dieser Form aufbereitet und vier voneinander unabhängig agierenden Prozessexperten zur Abgabe ihrer Handlungsvorschläge vorgelegt.

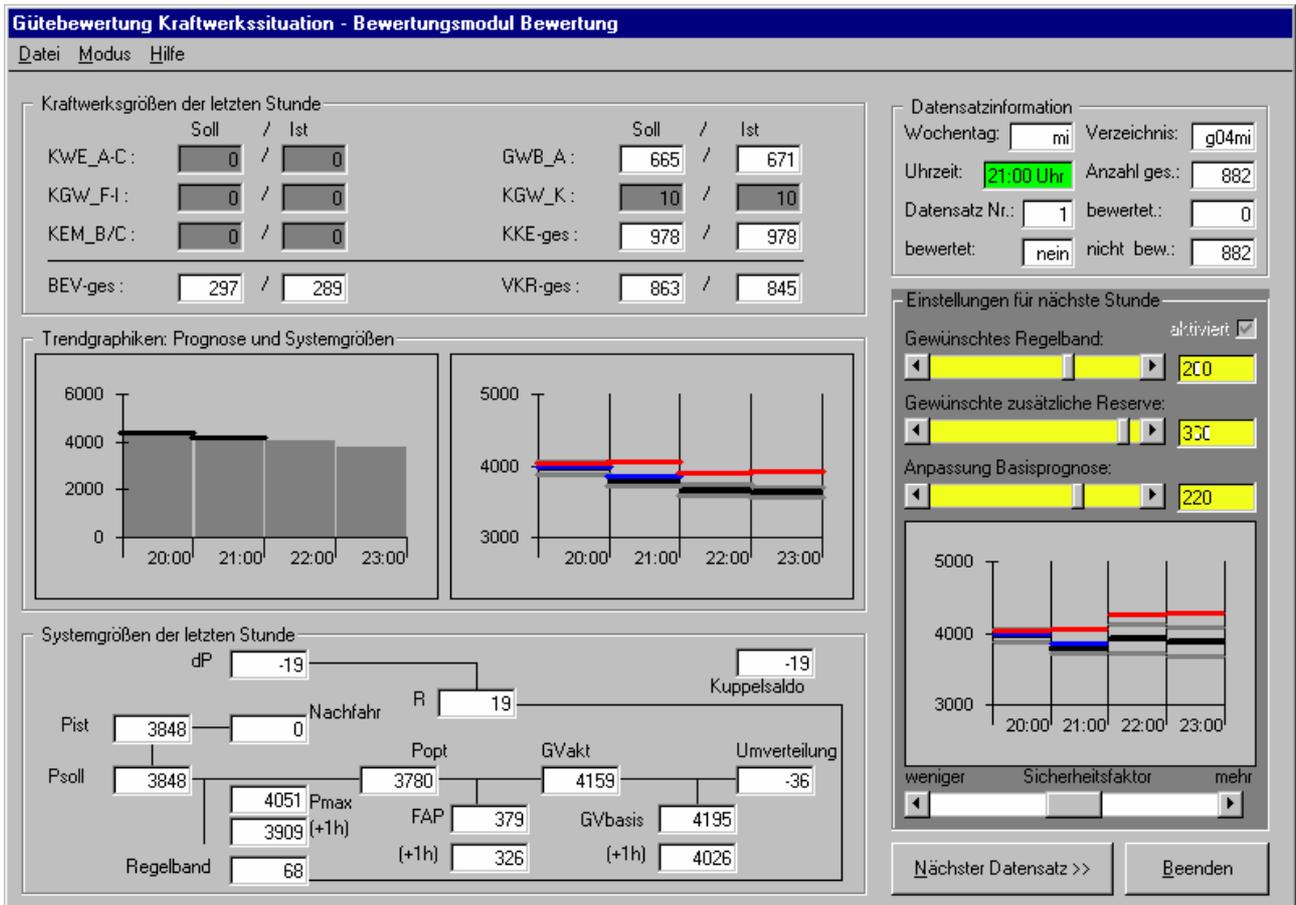


Abbildung 6.9: Graphische Benutzeroberfläche zur Präsentation eines aktuellen Prozesszustandes und Erhebung der drei Handlungsvorschläge im unteren rechten Bereich.

Zur Modellierung der Expertenvorschläge wurde für jede der drei genannten Strategieparameter mit Hilfe des Fuzzy–ROSA–Verfahrens ein separates Teilmodell generiert. Details hierzu sind ausführlich in [SSK01] beschrieben. An dieser Stelle werden beispielhaft die Ergebnisse zur Modellierung der Expertenvorschläge zum Strategieparameter *Anpassung der Basisprognose* detailliert beschrieben. Dabei wurden zum Vergleich unterschiedliche Sätze von Prozessgrößen als Eingangsgrößen verwendet und die so generierten Fuzzy–Module miteinander verglichen. Die Zugehörigkeitsfunktionen für jede Eingangsvariable wurde mit Ausnahme der Vorverarbeitung mit der evolutionären iterativen Merkmalsselektion anhand einer Analyse der Häufigkeitsverteilung, die durch Projektion der Lerndaten auf die zur betreffenden Eingangsvariablen gehörigen Koordinatenachse entsteht, bestimmt. Die Regelgenerierung wurde auf Basis der Lerndaten vorgenommen. Zur Bewertung der Modellgüte wird das Modell anhand der Validierungsdaten getestet, wobei die vorhandene Datenbasis im Verhältnis 1 : 1 zufällig in Lern- und Validierungsdaten aufgeteilt wurde. Als Maß für die Modellgüte wird der Korrelationskoeffizient zwischen tatsächlichem Expertenvorschlag und Modellausgang verwendet. Zur Einordnung der resultierenden Fehler wird der mittlere absolute Fehler **a** (Mittelwert) bzw. **b** (Mittelwert) angegeben, der beschreibt, wie stark die einzelnen Handlungsvorschläge im Mittel vom Mittelwert dieser Handlungsvorschläge abweichen.

Tabelle 6.3: Modellierungsergebnisse zur *Anpassung der Basisprognose* für unterschiedlich ausgewählte Merkmalsätze.

Modell	Validierungsdaten		Lerndaten	
	MAE ϵ_{vali}	Korrelation	MAE ϵ_{learn}	Korrelation
a (Mittelwert)	90,3 MW	-	90,0 MW	-
a (47)	10,1 MW	0,991	11,2 MW	0,994
a (29)	8,8 MW	0,994	8,6 MW	0,994
a (4)	6,2 MW	0,997	6,1 MW	0,995
a (2)	5,7 MW	0,998	5,4 MW	0,999
b (Mittelwert)	90,3 MW	-	90,0 MW	-
b (47)	9,5 MW	0,990	9,8 MW	0,989
b (29)	6,5 MW	0,992	6,4 MW	0,991
b (4)	4,6 MW	0,994	4,4 MW	0,993
b (2)	3,6 MW	0,997	3,1 MW	0,999

Die erzielten Ergebnisse zur Modellierung der Vorschläge der Experten **a** bzw. **b** sind in Tabelle 6.3 beispielhaft für mehrere Modelle angegeben, die für unterschiedliche Merkmalsätze als Eingangsgrößen erhalten wurden. Die mit **a**(47) und **b**(47) gekennzeichneten Modelle verwenden als Merkmalsatz alle ursprünglichen 47 Prozessgrößen, während sich die Modelle **a**(29) und **b**(29) auf den Satz von Eingangsgrößen beziehen, der sich durch aufwändige em-

pirische Reduktion ergeben hatte. Durch die Anwendung der datenbasierten Komplexitätsreduktion nach [PKS99] auf diese 29 Prozessgrößen konnte ein Merkmalsatz von vier Eingangsgrößen ermittelt werden, worauf die Modelle $\mathbf{a}(4)$ und $\mathbf{b}(4)$ generiert wurden. Die Anwendung der evolutionären iterativen Merkmalsselektion erlaubte schließlich, die Modelle $\mathbf{a}(2)$ und $\mathbf{b}(2)$ auf Basis eines Merkmalsatzes mit zwei Eingangsgrößen zu generieren. Obwohl sich diese resultierenden Modelle nur noch auf zwei Eingangsgrößen stützen, weisen sie den geringsten Modellfehler auf. Dies unterstreicht die überragende Bedeutung, die der Auswahl relevanter Eingangsgrößen zukommt. Die sehr gute Qualität der Modelle $\mathbf{a}(2)$ und $\mathbf{b}(2)$, die mit nur noch 24 bzw. 34 Regeln eine deutlich kleinere Regelbasis als alle anderen Modelle aufweisen, wird durch Abbildung 6.10 unterstrichen. Dargestellt sind die Ausgangsgrößenwerte des Modells und die zugehörigen Expertenvorschläge sortiert nach steigenden Werten der Expertenvorschläge.

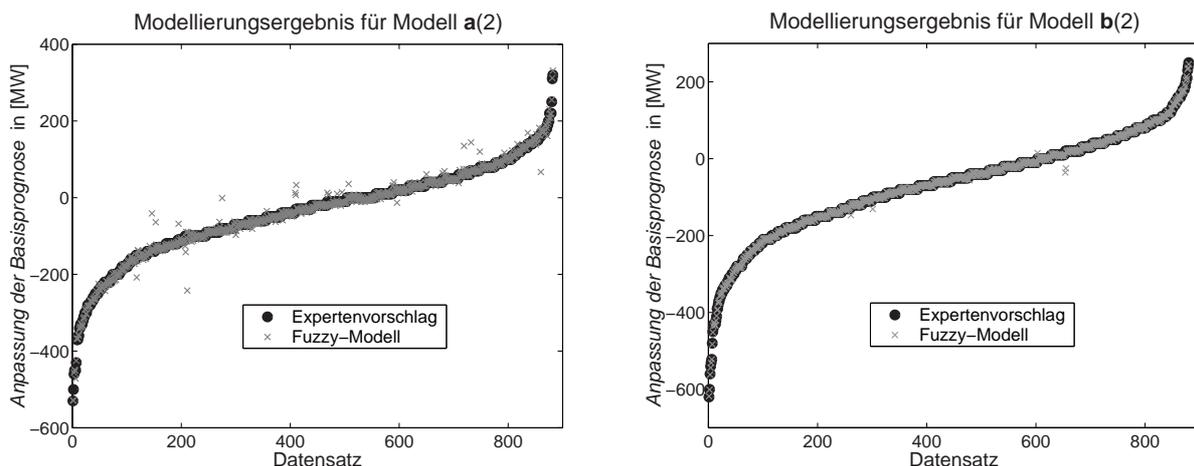


Abbildung 6.10: Modellierungsergebnisse zur *Anpassung der Basisprognose* für die Experten \mathbf{a} und \mathbf{b} (links bzw. rechts).

Ebenso wie Tabelle 6.3 zeigt auch dieses Bild, dass das Modell $\mathbf{b}(2)$ die Vorschläge des Experten \mathbf{b} noch besser wiedergibt als das Modell $\mathbf{a}(2)$ die Vorschläge des Experten \mathbf{a} . Aus dieser Beobachtung lässt sich zudem die Feststellung ableiten, dass der Modellfehler allein zur Bewertung eines Modells nicht immer ausreicht. Vielmehr interessiert neben einer guten Übereinstimmung zwischen Modellausgang und Lern- bzw. Validierungsdaten auch ein ausreichend glattes monotoneres Modellverhalten: Kleine Änderungen der Eingangssituationen sollen nicht zu unplausibel großen Änderungen des Ausgangsgrößenwertes führen. Wenn ein im Hinblick auf den Modellfehler inkorrekt erscheinendes Modell in diesem Sinne glatter als die Lerndaten sind, hat man einen Hinweis darauf, dass das Modell tatsächlich korrekter als die Lerndaten ist.

6.2.3 Qualitätskontrolle von Automatikgetrieben

Ein namhafter deutscher Automobilhersteller ermittelt vor der Endmontage die Fehlerfreiheit und Laufruhe von Automatikgetrieben (siehe Abbildung 6.11) durch eine akustische Prüfung von Prozessexperten. Ziel dabei ist die Steigerung der Produktqualität, die Senkung der Kosten und eine schnellere Detektion von Fehlerquellen.



Abbildung 6.11: Automatikgetriebe, zu Demonstrationszwecken geöffnet.

Die akustische Prüfung erfüllt dabei zwei Aufgaben: Zum einen sollen für den Menschen unangenehm laute Aggregate aussortiert und zum anderen mechanische Defekte durch ihre charakteristischen akustischen Signaturen ermittelt werden. Der Prüferscheid des Experten kann nur „in Ordnung“ (*i.o.*) oder „nicht in Ordnung“ (*n.i.o.*) lauten. Um den Prüferscheid des menschlichen Prozessexperten zu unterstützen und letztendlich zu automatisieren, ist in [Lie99, SK97, Hae99] ein Fuzzy-Klassifikator wissensbasiert entwickelt worden, der auf einen Merkmalsatz von insgesamt 149 aus gemessenen Zeitsignalen ermittelten Prozessmerkmalen aufsetzt. In diesem hochdimensionalen Eingangsgrößenraum ist der Entwurf eines Fuzzy-Klassifikators ein sehr komplexes Problem. Ein wissensbasierter Ansatz ist daher sehr zeitaufwändig und setzt sehr gute Prozesskenntnisse voraus. Im Rahmen einer langjährigen Arbeit war es möglich, von Hand einen Fuzzy-Klassifikator zu entwerfen. Mit über 200 Regeln und einer maximalen Kombinationstiefe von $c_{max} = 16$ linguistischer Aussagen in der Regelprämisse kann dieser mehr als 95% der im Datensatz enthaltenen Getriebe *richtig* klassifizieren, wobei *richtig* die Übereinstimmung von Fuzzy-Klassifizierung mit dem Urteil des menschlichen Prozessexperten meint.

In einem alternativen datenbasierten Ansatz [SSP⁺98, SPS⁺99] wurde zunächst ein Fuzzy-Klassifikator mit dem Fuzzy-ROSA-Verfahren auf der Basis von 1060 Datensätzen (1000 *i.o.*, 60 *n.i.o.*) generiert. Um zu demonstrieren, dass kein

Tabelle 6.4: Mittlerer absoluter Klassifikationsfehler (RCE) auf Lerndaten (ϵ_{learn}) und Validierungsdaten (ϵ_{vali}) für mit dem Fuzzy-ROSA-Verfahren generierte Fuzzy-Klassifikatoren basierend auf unterschiedlichen Merkmalsätzen.

Selektionsverfahren	Anzahl Einflussgrößen	Anzahl Regeln	RCE [%]	
			ϵ_{learn}	ϵ_{vali}
alle verfügbaren Einflussgrößen	149	1276	88,7	67,5
datenbasierte Komplexitätsreduktion	93	230	100	92,3
evolutionäre global operierende Merkmalsselektion	47	254	100	100
evolutionäre iterative Merkmalsselektion	7	134	100	99,9

Prozesswissen in den Entwurf einfließt, wurde jedes der 149 potenziellen Prozessmerkmale auf den vorliegenden Wertebereich normiert und mit fünf äquidistanten Zugehörigkeitsfunktionen partitioniert. Nach einer Rechenzeit von etwa 20 Stunden war es möglich, unter Verwendung aller Datenpunkte als Lerndaten einen Regelsatz mit *mehreren tausend Regeln* zu generieren, der die Getriebe zu 100 % richtig klassifiziert. Die von SLAWINSKI entwickelte systematische Vorgehensweise für die Modellierung führte im besten Ergebnis zu einem Regelsatz von 1276 Regeln, der auf Validierungsdaten 67,5 % der defekten *n.i.o.*-Getriebe richtig klassifiziert.

Mit dem Ziel, die Komplexität des Suchraums zu verringern und damit das Suchverfahren zu entlasten, ist auch das in [PSK98, PKS99, Pra99] entwickelte Verfahren zur datenbasierten Komplexitätsreduktion auf diese Aufgabenstellung angewendet worden. Damit konnten die 149 potenziellen Merkmale auf 93 relevante nicht redundante Merkmale reduziert werden. Mit dieser reduzierten Anzahl von Prozessmerkmalen wurde analog zur im vorherigen Abschnitt beschriebenen Vorgehensweise ein Fuzzy-Klassifikator, bestehend aus 230 Regeln, generiert. Dieser Fuzzy-Klassifikator erzielt auf den Lerndaten eine Klassifizierungsrate von 100 % und auf den Validierungsdaten eine von 92 %.

Von den im Rahmen dieser Arbeit konzipierten Verfahren ist aufgrund der hohen Dimensionalität des Modellierungsproblems an diesem Anwendungsbeispiel zunächst die evolutionäre global operierende Merkmalsselektion angewendet worden. Mit diesem Verfahren ist innerhalb einiger Minuten ein Merkmalsatz \mathcal{M}_{47} von 47 Einflussgrößen ermittelt worden, auf dessen Grundlage ein vollständig richtig entscheidender Fuzzy-Klassifikator mit 254 Regeln generiert werden konnte. Im Vergleich dazu hat eine mehrstündige Anwendung des evolutionären iterativen Verfahrens zur Merkmalsselektion einen extrem verkleinert

ten Merkmalsatz \mathcal{M}_7 von sieben Einflussgrößen ermittelt. Von den gefundenen sieben Einflussgrößen finden sich sechs in dem vorher mittels global operierendem Selektionsverfahren ermittelten Merkmalsatz \mathcal{M}_{47} wieder. Basierend darauf konnte ein Fuzzy-Klassifikator mit 134 Regeln generiert werden, der auf den Lerndaten eine Klassifizierungsrate von 100 % und auf den Validierungsdaten von 99 % erzielt. Bei den falschen Klassifikationsergebnissen handelte es sich um genau vier korrekt gefertigte Getriebe, die fälschlicherweise als *n.i.o.* klassifiziert wurden. Eine übersichtliche Zusammenfassung der Ergebnisse findet sich in Tabelle 6.4.

6.2.4 Strömungskraftschätzung eines Hydraulikventils

Die im Rahmen eines mittlerweile langjährigen Kooperationsprojektes mit der der Bosch Rexroth AG (ehemals Robert Bosch GmbH) durchgeführten Arbeiten dienten der Weiterentwicklung der automatisierten Regleroptimierung für ein industrielles Hydraulikventil am Beispiel eines Regelventils NG6. Ein wesentlicher Teilaspekt der Forschungsarbeiten bestand dabei in der datenbasierten Modellierung der an dem Hydraulikventil auftretenden Strömungskräfte.

Der Einsatzbereich industrieller Hydraulikventile erstreckt sich über weite Bereiche des Maschinen- und Anlagenbaus, wie beispielsweise in der Holz-, Papier- und Verpackungsindustrie. Abbildung 6.12 zeigt schematisch das zur Gruppe der Schieberventile gehörende Regelventil einer Querschnittsdarstellung.

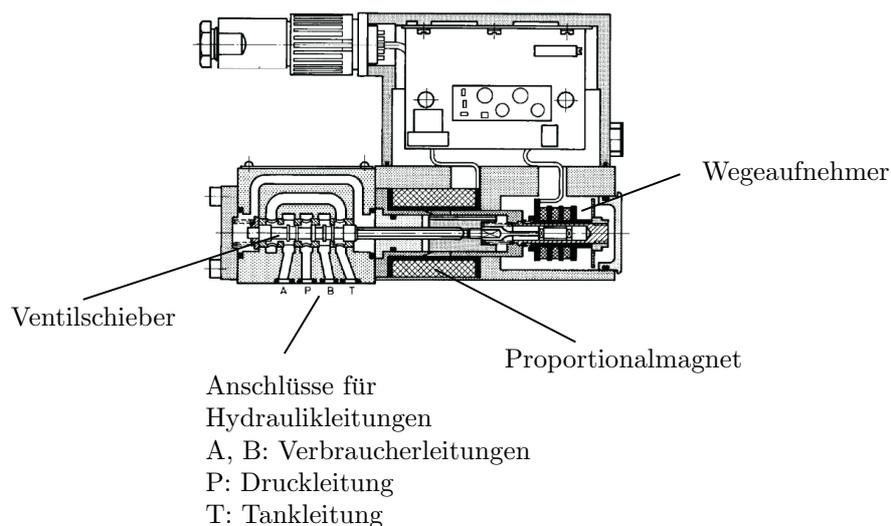


Abbildung 6.12: Querschnittsdarstellung des eingesetzten Hydraulikventils [BLSK00].

Daraus ersichtlich enthält das Ventil mittig in einer Längsbohrung im Gehäuse einen Kolben (Ventilschieber), der durch magnetische Krafteinwirkung in

Längsrichtung verschiebbar ist. In die Längsbohrung sind Ringkanäle eingearbeitet, die jeweils mit einem äußeren Anschluss am Gehäuse verbunden sind (Anschlüsse A, B, P und T). Eine Bewegung des Ventilschiebers bewirkt eine Veränderung der Durchflusswege für eine Hydraulikflüssigkeit durch Schließung und Öffnung der einzelnen Ringkanäle. Die Position $x(t)$ (auch als *Ventilhub* bezeichnet) des Ventilschiebers soll mit Hilfe der eingesetzten *Lageregelung* geregelt werden. Detailliertere Informationen zur Wirkungsweise und zum Einsatz von Hydraulikventilen sind in [Rob99] zu finden.

Die Dynamik hydraulischer Stetigventile wird wesentlich beeinflusst durch Nichtlinearitäten wie Strombegrenzung, Coulombsche Reibung, magnetische Hysterese, Sättigung der Flussdichte sowie durch Störgrößen in Form von Strömungskräften. Die Struktur des implementierten Lagereglers ist aus diesem Grund stark an die Nichtlinearitäten der Strecke angepasst und beinhaltet insgesamt 24 einzustellende Parameter.

Nachdem das Bewertungsverhalten von Prozessexperten bei eigenhändigen Parametrisierungen des Reglers mit einem Fuzzy-Bewertungsmodul nachgebildet worden war, konnte eine automatisierte Optimierung der Reglerparameter mit dem realen Hydraulikventil Hardware-in-the-Loop durchgeführt werden. Als Optimierungsalgorithmus wurde aufgrund der großen Anzahl der Reglerparameter eine Evolutionsstrategie eingesetzt. Abbildung 6.13 zeigt schematisch den Aufbau des Laborstands.

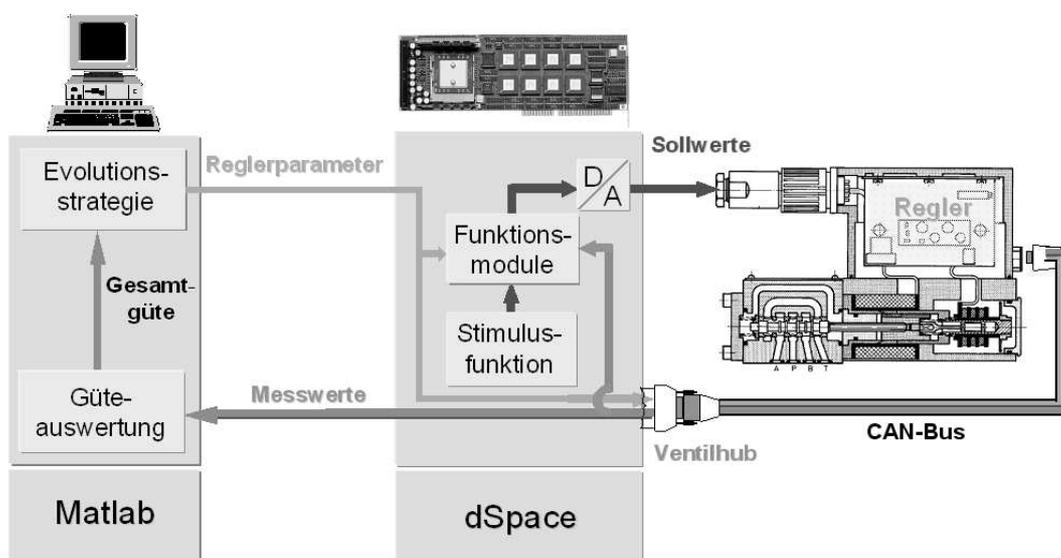


Abbildung 6.13: Aufbau der evolutionären Hardware-in-the-Loop-Optimierung.

Das Auftreten der Strömungskräfte an dem Regelventil in einem geschlossenen Hydrauliksystem wurde bei der Optimierung für eine lange Projektlaufzeit ignoriert. Die Gründe hierfür waren zum einen, dass eine Optimierung unter dem Einfluss einer hydraulischen Last aufgrund des erforderlichen hydraulischen Laborstandes nur vor Ort bei der Bosch Rexroth AG unter einem erheblichen Mehraufwand möglich gewesen wäre. Der andere Grund lag darin, dass es keine geeignete Möglichkeit gab, das Auftreten dieser Kräfte zu simulieren. Die Ergebnisse einer „trockenen Optimierung“ zeigen zwar auch beim Betrieb unter Lasteinwirkung ein akzeptables Verhalten, stellen aber höchstwahrscheinlich nur eine suboptimale Lösung dar. Man kann davon ausgehen, dass eine unter Lasteinwirkung optimierte Parametereinstellung beim praktischen Einsatz ein besseres Resultat liefern würde. Das unterschiedliche Ventilverhalten bei Betrieb mit oder ohne hydraulische Last zeigt sich beispielsweise auch bei einer Hysteresemessung in Abbildung 6.14.

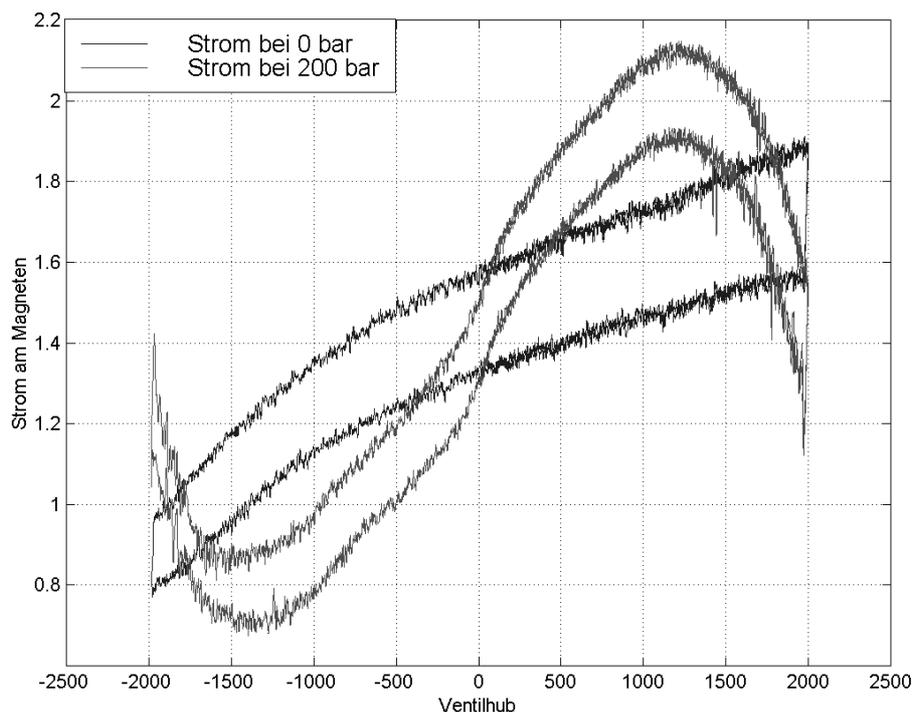


Abbildung 6.14: Stromkennlinie beim Betrieb mit und ohne hydraulische Last.

Für die Aufgabenstellung der Modellierung stellte sich zunächst das Problem, dass eine direkte Messung der Strömungskräfte nicht möglich ist, so dass Standardverfahren für eine datenbasierte Modellierung nicht unmittelbar eingesetzt werden konnten. Eine wissensbasierte Modellierung konnte ebenfalls nicht durchgeführt werden, da keine genaueren Zusammenhänge für die Beschreibung der Strömungskräfte bekannt sind. Durch den Einsatz verschiedener individuell

angepasster aufwändiger Funktionsmodule gelang es, das Verhalten unter Last- einwirkung im trockenen Betrieb größtenteils ausreichend gut zu simulieren. Die Verwendung der Funktionsmodule war allerdings wenig flexibel und nur bedingt transparent. Angestrebt wurde daher eine datenbasierte Modellierung, die sich auf die Messwerte, die mit den Funktionsmodulen aufgezeichnet wurden, stützen sollte. Ein so entworfenes Modell musste allerdings nicht nur von der Modellierungsgüte her gute Ergebnisse liefern, sondern auch in einer geeigneten Form realisierbar sein, um in die „Hardware-in-the-Loop-Optimierung“ eingebunden werden zu können. Es kamen daher sowohl ein Neuronales Netz als auch ein neu entwickeltes Zellenmodell in Frage, wobei sich das Zellenmodell für das hier vorliegende Problem als die wesentlich geeignetere Modellierungs- variante erwies.

In den ersten Modellierungen als Zellenmodell wurde das evolutionäre Verfahren zur Merkmalsselektion nur zur Optimierung der Partitionierungen der vorhandenen Einflussgrößen *Position*, *Geschwindigkeit* und *Beschleunigung*, also ohne ohne Relevanzanalyse verwendet. In weiteren Modellierungsansätzen wurden im Hinblick auf eine Verbesserung der Modellierungsgüte aus den vorhandenen Eingangsgößen neue Größen abgeleitet, mit deren Hilfe eine bessere Trennung der Daten für bestimmte aufgenommene Trajektorien gelingen sollte. Dazu wurden aus den ursprünglichen drei Eingangsgößen mit Hilfe von Verzögerungsgliedern erster Ordnung (PT1-Glieder) mit unterschiedlichen Zeitkonstanten neue abgeleitete Einflussgrößen generiert. Die Zeitkonstanten wurden dabei in Anlehnung an die typischen Sprungantworten gewählt. Zusätzlich wurden unterschiedliche Summen aus den normierten Daten von Position und Geschwindigkeit gebildet. Insgesamt wurden so 15 neue potenzielle Einflussgrößen generiert und mit Hilfe der iterativen evolutionären Merkmalsselektion ausgewertet. Für das Zellenmodell wurden so die Eingangsgößen *Position*, *Geschwindigkeit* und *verzögerte Position* ($T1=25.5ms$) ausgewählt. Der Ansatz zur Approximation mit Hyper- ebenen ermöglichte dabei deutlich die besten Modellierungsergebnisse.

Der Vergleich von Zellenmodell **A** mit Zellenmodell **B** zeigt eindrucksvoll die Steigerung der erreichbaren Modellierungsgüte bei Verwendung der optimierten Partitionierung der selektierten Merkmale. Die Verwendung einer äquidistan- ten Intervalleinteilung für das Zellenmodell **A** erlaubte die Modellierung mit einem mittleren absoluten Fehler von 0,023 (siehe Abbildung 6.15), was im Vergleich zu vorherigen Modellierungen mit Verwendung der originalen vor- handenen Einflussgrößen einer Verminderung des Modellierungsfehlers auf ein Drittel der vorherigen Werte entsprach. Trotzdem trat aber noch immer ein maximaler absoluter Fehler von 1,357 auf. Durch die Optimierung der Parti- tionierungen der gefundenen Merkmale für das Zellenmodell **B** ergab sich ein

mittlerer absoluter Fehler von nur noch 0,005, wobei der maximale absolute Fehler signifikant auf 0,141 gesenkt wurde.

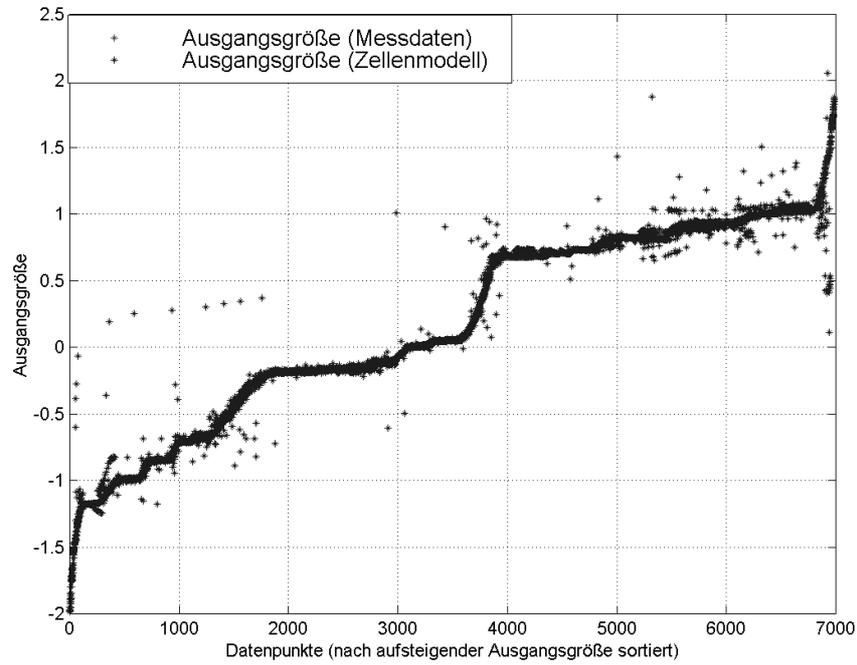


Abbildung 6.15: Modellierungsergebnis Zellenmodell A.

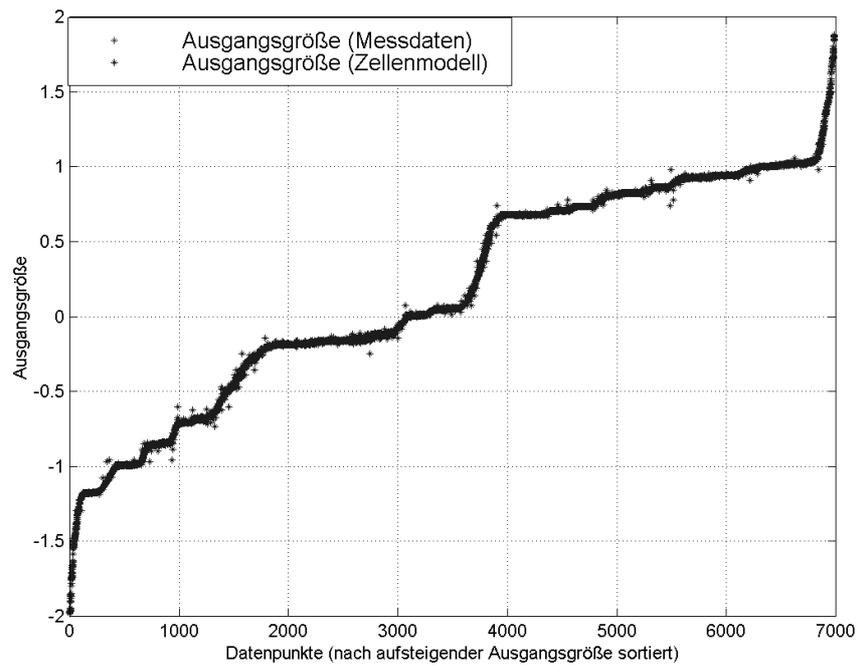


Abbildung 6.16: Modellierungsergebnis Zellenmodell B.

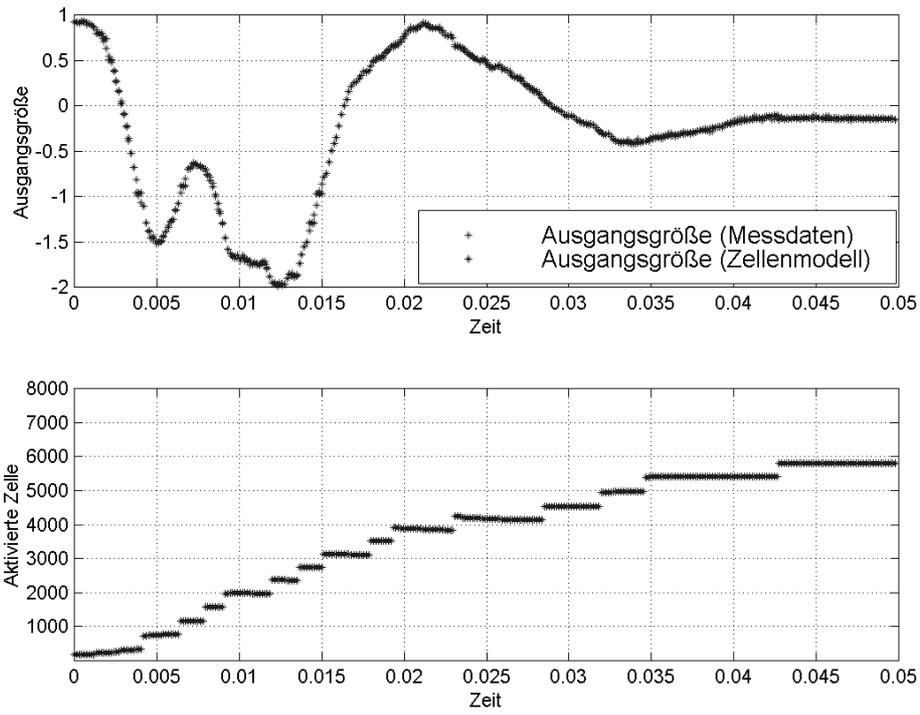


Abbildung 6.17: Modellierungsergebnis Zellenmodell **B** Trajektorie 1.

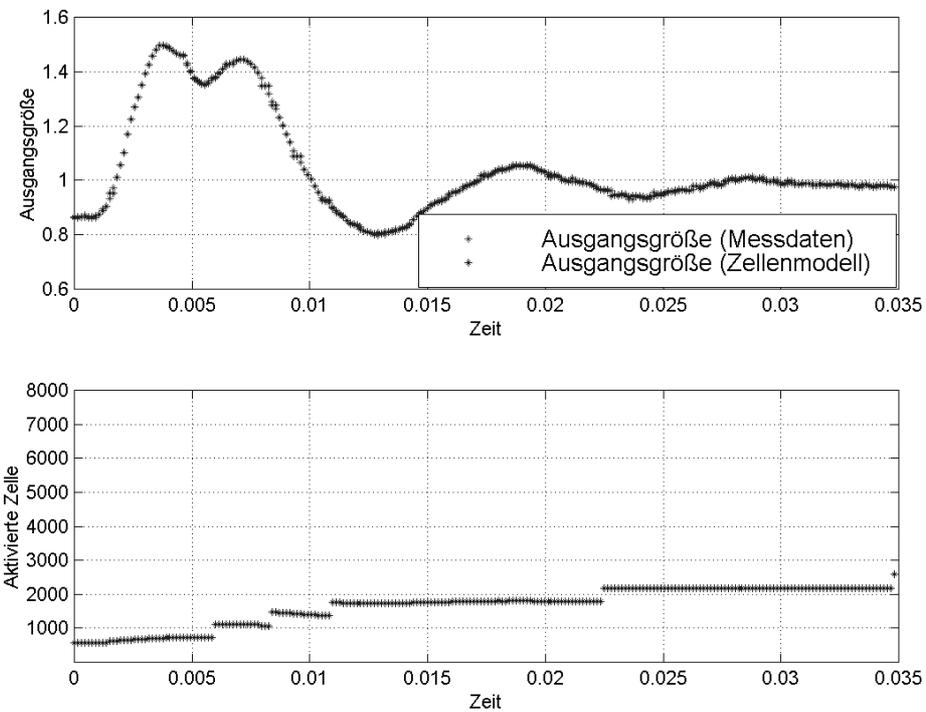


Abbildung 6.18: Modellierungsergebnis Zellenmodell **B** Trajektorie 8.

Die Abbildungen 6.17 und 6.18 zeigen beispielhaft für zwei ausgewählte Sprungantworten die extrem gute Übereinstimmung von aufgenommenen Datenwerten und Ausgangswerten des Zellenmodells **B**. Welches Verbesserungspotenzial sich in Bezug auf die Regleroptimierung durch die Beachtung der Strömungskräfte ergibt, wird aus Abbildung 6.19 deutlich. Darin ist das Führungsverhalten eines NG6-Ventils beim Betrieb unter Last dargestellt, wobei das verwendete Ventil einen größeren Arbeitsbereich aufweist als das üblicherweise betrachtete Ventil. Dieses Ventil wird einmal mit dem Standardregler betrieben und einmal mit einem unter simulierter Last optimierten Regler. Aufgrund des größeren Arbeitsbereichs ist das Strömungskraftmodell nicht optimal an dieses Ventil angepasst. Das führt dazu, dass beispielsweise beim dritten Sprung der dargestellten Sprungfolge eine Verschlechterung auftritt. Es können aber trotzdem bereits deutliche Verbesserungen erzielt werden.

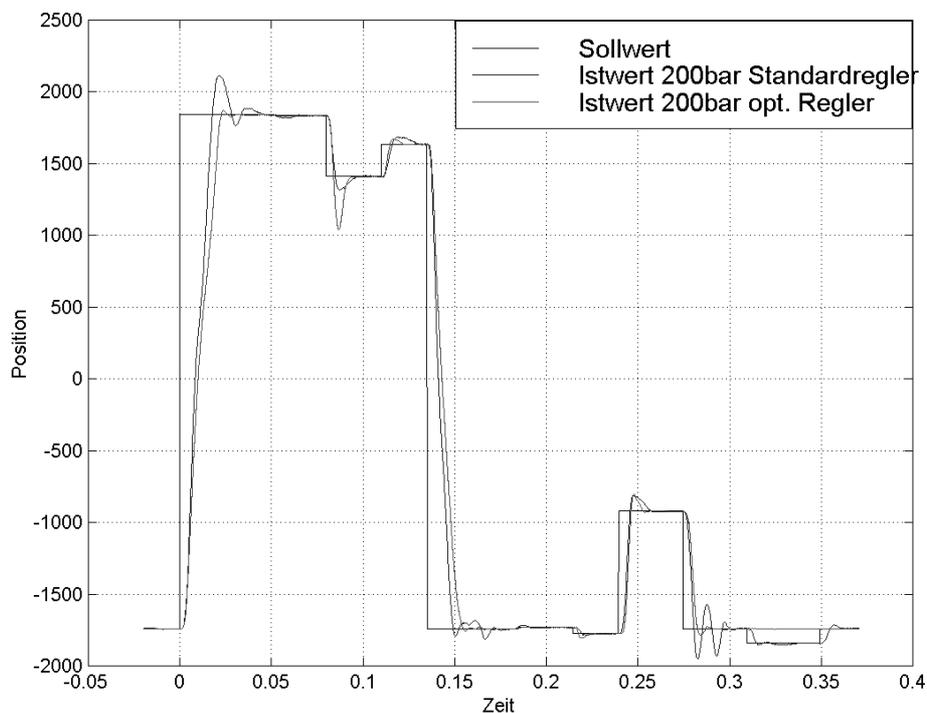


Abbildung 6.19: Vergleich des Standardreglers mit dem unter simulierter Last optimierten Regler.

Der unter simulierter Last optimierte Regler verhindert beim ersten Großsignalsprung ein großes Überschwingen, wie es beim Einsatz des Standardreglers auftritt. Auch bei dem zweiten Großsignalsprung ist die Schwingneigung reduziert. Besonders deutlich tritt der Unterschied auch bei dem letzten Mittelsignalsprung hervor, wo die großen Oszillationen des Standardreglers vollständig beseitigt sind.

6.2.5 Relevanzbasierte strukturelle Evolution von TSK-Fuzzy-Reglern

Ausgehend von den vorherigen sehr guten Modellierungsergebnissen ist die Grundidee der graduellen Verfeinerung beim evolutionären iterativen Selektionsverfahren für die Fuzzy-Modellierung auf einen schrittweisen Entwurf von TSK-Fuzzy-Regler für nichtlineare Systeme übertragen und an einem realen Anwendungsbeispiel untersucht worden [HHS05, HS05]. Da das optimale Kennfeld eines Fuzzy-Reglers im vorhinein nicht bekannt ist, lässt sich beim datenbasierten Reglerentwurf nicht wie bei der Fuzzy-Modellierung ein lokaler Modellfehler bewerten, anhand dessen sich Strukturparameter optimieren ließen. Es lässt sich lediglich das Gesamtverhalten des geschlossenen Regelkreises im Zusammenspiel aller Regeln bestimmen. Darüber hinaus ist analog zur Problematik der Merkmalsselektion auch beim Reglerentwurf nicht bekannt, welche Partitionierung für das Regelungsproblem optimal ist, also entlang welcher Zustandsgrößen der Zustandsraum zu partitionieren ist und in welche Fuzzy-Mengen diese zu unterteilen sind. Zur Lösung dieses komplexen Entwurfsproblems ist das im Rahmen dieser Arbeit entwickelte evolutionäre iterative Verfahren zur Merkmalsselektion dahingehend übertragen worden, dass sowohl die Struktur und die Parameter der Zustandsraumpartitionierung als auch die Verstärkungsfaktoren in der Konklusion der TSK-Fuzzy-Regeln in Bezug auf ein anwendungsspezifisch formuliertes Gütefunktional optimiert werden. Dieses Verfahren ist am Beispiel eines inversen Rotationspendels erprobt worden.

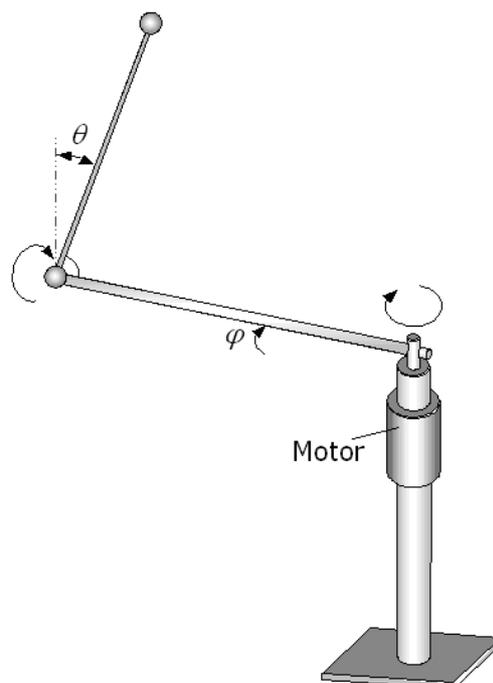


Abbildung 6.20: Inverses Rotationspendel.

Das inverse Rotationspendel in Abbildung 6.20 besteht aus zwei senkrecht zueinander angeordneten Achsen, wobei die erste vertikale Achse von einem Motor angetrieben wird. Das an der zweiten horizontalen Achse befestigte Pendel kann dagegen frei schwingen. Das Regelungsziel besteht im Aufschwingen des Pendels aus der unteren Ruhelage und der anschließenden Stabilisierung des freien Pendels in der oberen instabilen Ruhelage. Das dynamische Verhalten des inversen Rotationspendels ist beschrieben durch ein nichtlineares Differentialgleichungssystem vierter Ordnung

$$\ddot{\varphi} = -c_1 \dot{\varphi} + c_2 u \quad \text{und} \quad (6.2)$$

$$\ddot{\theta} = c_3 \left[-c_4 \ddot{\varphi} \cos(\theta) + c_5 \dot{\varphi}^2 \cos(\theta) \sin(\theta) + c_6 \sin(\theta) - c_7 \dot{\theta} \right], \quad (6.3)$$

wobei θ die Winkelstellung des inversen Pendels, φ die Winkelstellung des horizontalen Armes und τ das vom Motor erzeugte Drehmoment darstellen. Die Koeffizienten c_i lassen sich aus den Systemparametern

$$\begin{aligned} c_1 = \mu_{gear} \quad c_3 = \frac{1}{J_p + m_p \left(\frac{l_p}{2}\right)^2} \quad c_5 = m_p \left(\frac{l_p}{2}\right)^2 \\ c_2 = \eta \quad c_4 = m_p \frac{l_p l_b}{2} \quad c_6 = m_p g \frac{l_p}{2} \quad c_7 = \mu_{bearing} \end{aligned}$$

berechnen, wobei l_p die Länge, m_p die Masse und J_p das Trägheitsmoment des vertikalen Armes, μ_{gear} und $\mu_{bearing}$ die Reibungskoeffizienten des Motorgetriebes und des Pendellagers und η die statische Verstärkung zwischen Motorspannung und Motorgeschwindigkeit beschreiben. Die Regelungsaufgabe wird durch zwei separate Fuzzy-Regler gelöst, zwischen denen über ein Hystereseglied situationsabhängig umgeschaltet wird: Ein Regler soll das Pendel in möglichst kurzer Zeit aufschwingen, wohingegen ein zweiter Regler das Pendel für kleinere Auslenkungen um die obere Ruhelage stabilisiert.

Den Ausgangspunkt zur *relevanzbasierten Evolution eines TSK-Fuzzy-Reglers* stellt dabei ein optimaler linearer Regler dar, welcher aus einem am gewünschten Arbeitspunkt linearisierten Modell hergeleitet und dann durch eine einzelne globale Fuzzy-Regel

$$\text{IF } X \text{ is } A \text{ THEN } u = k_0 + \sum_i k_i x_i, \quad \text{mit } \mu_A(x) = 1, \quad \forall x \in X$$

implementiert wird. Die Verstärkungsfaktoren k_i werden mit Hilfe des LQR-Verfahrens bestimmt. Das ursprünglich lineare Kennfeld wird durch eine Partitionierung des Zustandsraumes in Fuzzy-Mengen schrittweise verfeinert, so dass

der Zustandsraum in verschiedenen Regionen durch Regeln mit unterschiedlichen Verstärkungsfaktoren überdeckt wird. Die Strukturverweiterung erfolgt dabei durch Einfügen zusätzlicher Fuzzy-Mengen entlang jeweils einer Eingangsgröße, einhergehend mit einer Duplizierung der Regeln. Dabei kann die neue Fuzzy-Menge auf zwei unterschiedliche Arten generiert werden und zwar durch Teilung einer Fuzzy-Menge oder durch Einfügen einer neuen Fuzzy-Menge zwischen zwei bestehende Fuzzy-Mengen. Im Fall des inversen Rotationspendels als Anwendungsbeispiel sind alle Fuzzy-Partitionierungen symmetrisch zum Ursprung definiert. Daher wird dort eine Erweiterungsoperation immer simultan auf die entsprechende Fuzzy-Menge im negativen und im positiven Teilraum angewandt.

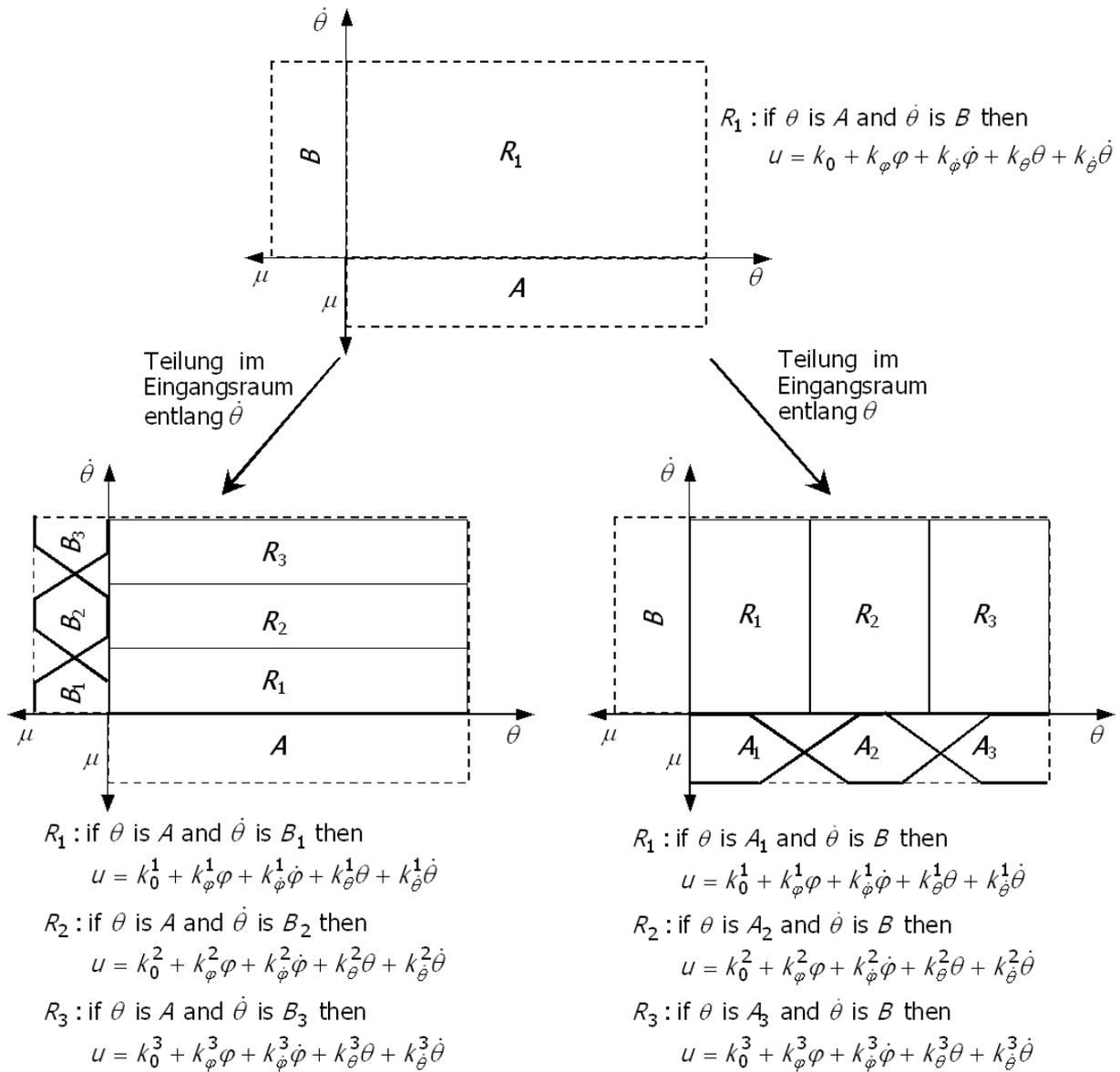


Abbildung 6.21: Zwei mögliche Partitionierungen des Zustandsraums für einen TSK-Fuzzy-Regler.

Abbildung 6.21 illustriert den Erweiterungsschritt für einen Regler am Beispiel von zwei der vier Zustandsgrößen (θ und $\dot{\theta}$). Die Anfangspopulation enthält Individuen, die die Koeffizienten $\{k_0, k_\varphi, k_{\dot{\varphi}}, k_\theta, k_{\dot{\theta}}\}$ einer einzelnen TSK-Regel R_1 (Abbildung 6.21 oben) der Form

$$\begin{aligned} & \text{WENN } \theta \text{ ist } A \text{ UND } \dot{\theta} \text{ ist } B \quad \text{DANN } u = k_0 + k_\varphi\varphi + k_{\dot{\varphi}}\dot{\varphi} + k_\theta\theta + k_{\dot{\theta}}\dot{\theta}, \\ & \text{mit } \mu_A(\varphi) = 1 \text{ und } \mu_B(\dot{\varphi}) = 1 \quad \forall \varphi, \dot{\varphi} \in X \end{aligned}$$

enthalten. Die Koeffizienten werden entweder zufällig initialisiert, oder durch die Anwendung des LQR-Verfahrens ermittelt.

Die Verfeinerung der Zustandsraumpartitionierung kann entlang jeder der vier Zustandsgrößen erfolgen. Abbildung 6.21 zeigt zwei mögliche Unterteilungen entlang der Zustandsgrößen $\dot{\theta}$ (unten links) und θ (unten rechts). Im ersten Fall wird die globale Zugehörigkeitsfunktion B durch drei überlappende trapezförmige Fuzzy-Mengen B_1, B_2, B_3 ersetzt. Die ursprünglich einzige Regel R_1 wird dabei durch drei Regeln R_1, R_2, R_3 mit unterschiedlichen Koeffizienten k_i^1, k_i^2, k_i^3 substituiert. Ein zur Kodierung der k_i ursprünglich aus fünf Zielvariablen bestehendes Individuum wird durch ein 15-parametriges Individuum ersetzt, das zunächst aus drei identischen Kopien der ursprünglichen Koeffizienten k_i besteht. Zusätzliche Parameter kodieren die Stützstellen des Überlappungsbereichs der benachbarten Fuzzy-Mengen A_1/A_2 und A_2/A_3 bzw. B_1/B_2 und B_2/B_3 .

Jede vorhandene Zustandsgröße wird entsprechend behandelt und innerhalb einer isolierten Subpopulation mittels einer verzweigten Evolutionsstrategie optimiert. Alle Lösungen einer Subpopulation partitionieren den Eingangsraum entlang der gleichen Eingangsgröße. In den nachfolgenden Generationen entwickeln sich die Subpopulationen parallel und unabhängig voneinander, wobei die Evolutionsstrategie jeweils die ihrer Partitionierung des Fuzzy-Reglers entsprechenden Parameter optimiert.

Nach einigen Generationen werden die besten Lösungen der Subpopulationen untereinander verglichen. Die Population mit dem größten Fortschritt definiert für die vorgesehene Erweiterung die relevanteste Einflussgröße und überlebt, alle übrigen Subpopulation werden verworfen. Dieser Kreislauf aus struktureller Erweiterung des Fuzzy-Reglers und Optimierung der zusätzlichen Freiheitsgrade wird so lange fortgesetzt, bis eine feinere Partitionierung zu keiner nennenswerten Verbesserung des Regelverhaltens mehr führt. Abbildung 6.22 verdeutlicht graphisch dieses Schema. Der wesentliche Vorteil dieser neuen Methode zum Reglerentwurf besteht darin, dass eine strukturelle Anpassung und Parameter-

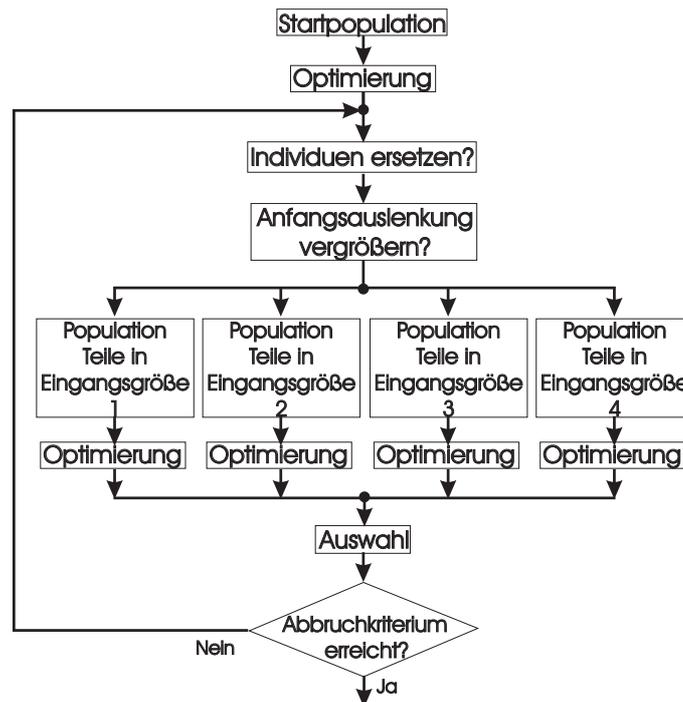


Abbildung 6.22: Evolutionsschema zum iterativen evolutionären Entwurf eines TSK-Fuzzy-Reglers für eine Regelstrecke mit vier Zustandsgrößen.

optimierung kombiniert angewendet wird. Anstatt eine feste komplexe Struktur mit einer Vielzahl von Parametern direkt zu optimieren, wird so die Dimension des Optimierungsraumes schrittweise erhöht und die Struktur flexibel gehalten. Dabei wird die Population der nächsten Stufe durch die zunächst neutrale Erweiterung des Reglers bereits mit einer gut angepassten Lösung der vorherigen Stufe vorinitialisiert.

Die relevanzbasierte strukturelle Entwicklung des Kennfeldes über die zwei Erweiterungsstufen ist in Abbildung 6.23 für das Beispiel des Stabilisierungsreglers für das inverse Rotationspendel dargestellt.

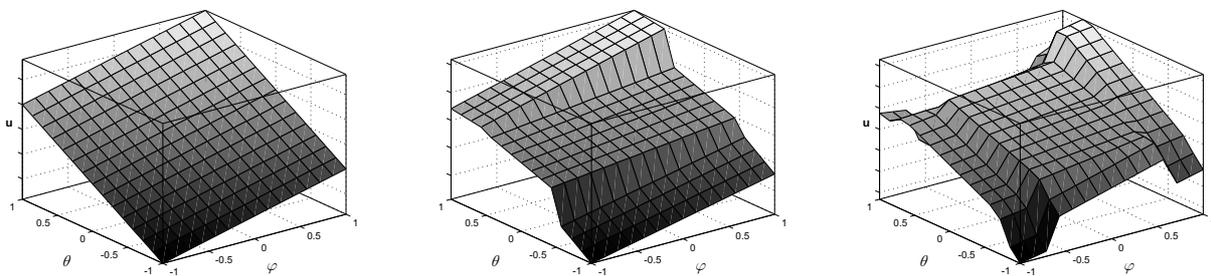


Abbildung 6.23: Entwicklung des Kennfeldes für die ersten zwei Erweiterungsstufen.

Dabei sei angemerkt, dass die Graphiken jeweils die Projektion des vierdimensionalen Kennfeldes auf die Zustandsgrößen φ und θ enthalten. In der linken Abbildung ist das beste lineare Reglergesetz abgebildet. In den nachfolgenden

Erweiterungen wird deutlich, wie die Nichtlinearität des Kennfelds mit wachsender Anzahl von Regeln in den nächsten zwei Erweiterungsschritten zunimmt.

Die damit erzielten Ergebnisse bezüglich des Ausregelverhaltens werden in Abbildung 6.24 verdeutlicht. Darin sind die Verläufe der beiden Winkel θ für das Pendel (oben) und φ für den horizontalen Arm (unten) dargestellt, die während eines Ausregelvorganges bei einer vorgegebenen Störauslenkung von $\theta = 0,6$ auftreten. Beide Abbildung enthalten sechs zeitliche Verläufe, die dem Ergebnis der nach Relevanzgesichtspunkten ausgewählten optimierten Reglern der jeweiligen Erweiterungsstufe entsprechen. Dunklere Linien korrespondieren dabei mit Reglern gesteigerter Komplexität, wobei die ganz dunkle Linie dem finalen Regler nach fünf Erweiterungsschritten mit dem geringsten Zustandsfehler entspricht. Die Ergebnisse bestätigen eine kontinuierliche Verbesserung des Regelverhaltens mit zunehmender Komplexitätsstufe und damit verbundenen zusätzlichen Freiheitsgraden des Reglers.

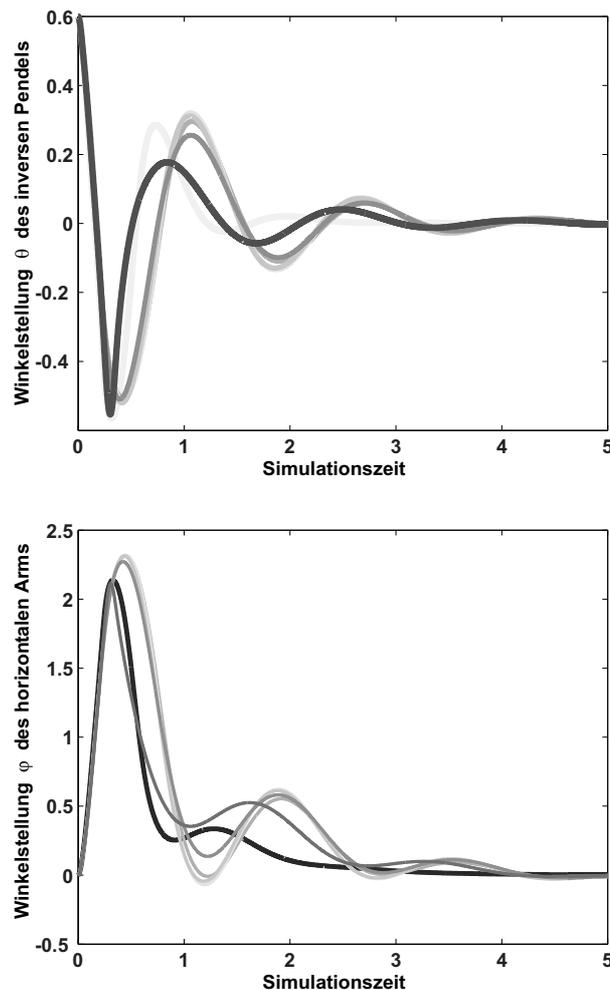


Abbildung 6.24: Entwicklung des Regelverhaltens für verschiedene Strukturereicherungen des TSK-Fuzzy-Reglers.

Die Anwendung des relevanzbasierten evolutionären Entwurfsverfahrens für einen TSK–Aufschwingregler erfolgt analog der vorher beschriebenen Vorgehensweise unter Verwendung eines an die Aufgabenstellung angepassten Gütefunktional. Das erreichte gute Regelungsverhalten des so gewonnenen Gesamtsystems ist in Abbildung 6.25 als gestrichelte Linie dargestellt. Zur Analyse der Robustheit des Regelungssystems werden zwei verschiedene Störgrößen auf den Eingang des Pendelmodells überlagert. Das Störsignal besteht einerseits aus einem dauerhaften gaußverteilten Rauschen mit einer Varianz von $\sigma^2 = 0,0005$. Zusätzlich wird zum Zeitpunkt $t^* = 8s$ dem bereits stabilisierten Pendel ein Rechteckimpuls mit einer Pulsweite von $T = 200ms$ und einer Amplitude von 0,1 aufgeschaltet. Diese Störung ist so groß, dass der stabilisierende Regler diese nicht ausregeln kann, so dass vorübergehend der Aufschwingregler aktiviert wird. Als Ergebnis zeigt sich insgesamt ein sehr robustes Regelungsverhalten, welches nach dem Störimpuls zu einem schnellen Wiederaufschwingen und dauerhaftem Stabilisieren des inversen Rotationspendels führt (durchgezogene Linien in Abbildung 6.25).

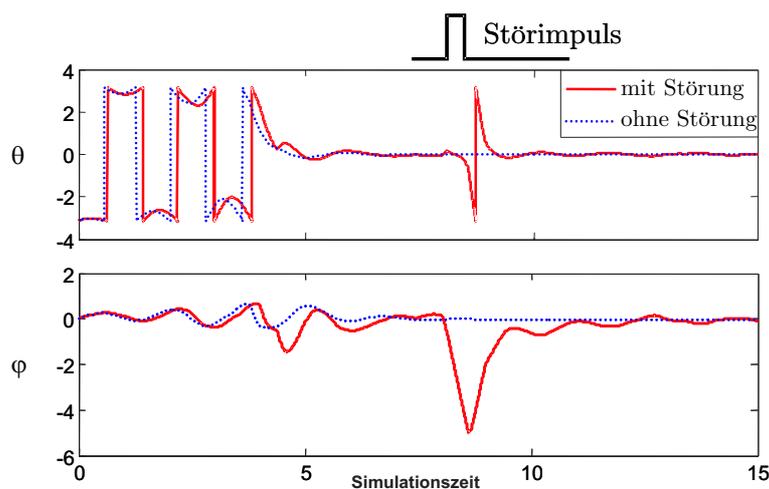


Abbildung 6.25: Ausgangsverhalten des Regelungssystems mit und ohne Störsignale.

Weitere Details hierzu sowie eine weitere Anwendung des Verfahrens zum Entwurf eines TSK–Reglers zur Roboternavigation ist in [HSH07] zu finden.

Kapitel 7

Zusammenfassung und Ausblick

Eine zentraler Aufgabenstellung ingenieurwissenschaftlicher Tätigkeiten besteht in der Konzeption, Analyse, und zielgerichteten Beeinflussung von technischen Systemen. Typischer Weise erfolgen diese Tätigkeiten aus Kostengründen, Effizienzgründen und häufig auch aus sicherheitstechnischen Gründen zunächst auf simulativer Basis, so dass die Verfügbarkeit eines Systemmodells eine wesentliche Voraussetzung hierfür darstellt. In zunehmenden Maße werden für die Modellgenerierung für komplexe Systeme datenbasierte Verfahren aus dem Bereich der Computational Intelligence eingesetzt. Dabei nimmt die datenbasierte Fuzzy-Modellierung eine besondere Stellung ein, weil durch sie interpretierbare Modelle auch bei wenig verfügbarem Systemverständnis generiert werden können.

Das übergeordnete Ziel dieser Arbeit besteht darin, den Anwendungsbereich der datenbasierten Fuzzy-Modellierung in Richtung höherer Problemkomplexität auszudehnen und gleichzeitig die Komplexität bei der Parametrierung des Fuzzy-Modellierungsverfahrens zur Modellgenerierung für den Anwender zu vermindern. Des Weiteren ist aus Anwendersicht von Bedeutung, dass die generierten Fuzzy-Modelle den meist hohen Anforderungen an Modellierungsgüte und Interpretierbarkeit genügen. Darüber hinaus ist es häufig wünschenswert das Systemverständnis über das zu modellierende System durch Auswertung der generierten Regeln zu steigern. Weiterhin hängt die Akzeptanz eines Modellierungsverfahrens in der Praxis auch davon ab, ob es zum einen einfach, d. h. ohne großen Aufwand und spezielle Vorkenntnisse, und zum anderen flexibel und robust eingesetzt werden kann. Hinsichtlich dieser teilweise konkurrierenden Zielsetzungen sind im Rahmen dieser Arbeit neue Konzepte und Lösungsstrategien entwickelt und am Beispiel der Fuzzy-Modellierung mit dem von KIENDL, KRABS, KRONE und SLAWINSKI entwickelten Fuzzy-ROSA-Verfahren [KK89, Kra94, KK95, Kro99, Sla01] getestet worden.

Die dabei erzielten Ergebnisse bei Benchmarkproblemen sowie bei realen An-

wendungsproblemen belegen eindrucksvoll, dass wesentliche Fortschritte durch den Einsatz der im Rahmen dieser Arbeit konzipierten Strategieelemente erzielt werden. Von besonderer Bedeutung ist dabei auch, dass in nahezu allen getesteten Anwendungsproblemen die erzielten deutlichen Verbesserungen bezüglich ein oder mehrerer *Teilziele* ohne signifikante Verschlechterungen hinsichtlich der übrigen *Teilziele* erreicht werden.

Ein übergeordnetes Ziel dieser Arbeit ist es, aus einer gegebenen Menge von potenziellen Einflussgrößen eines gegebenen Modellierungsproblems im Rahmen einer *Relevanzanalyse und Redundanzanalyse potenzieller Einflussgrößen* automatisiert einen relevanten Merkmalsatz zu selektieren. Hierdurch wird eine *Komplexitätsreduktion* für das nachgeschaltete Modellierungsverfahren erzielt und darüber hinaus dem Anwender durch Angabe der prozessrelevanten Einflussgrößen ein transparenter Einblick zur *Erhöhung des Systemverständnisses* gegeben. Um die Handhabung möglichst komplexer Modellierungsprobleme zu ermöglichen, ist hierzu aus rechentechnischen Gesichtspunkten ein eigenständiges evolutionäres Verfahren zur Selektion relevanter und nichtredundanter Merkmalsätze entwickelt worden (Kapitel 4).

Ein zentrales Strategiekonzept des als Filteransatz realisierten Verfahrens besteht in der Berücksichtigung des Zusammenspiels betrachteter Einflussgrößen und einer *gesamtheitlichen Relevanzbewertung*, wodurch Redundanzen innerhalb des Merkmalsatzes aufgespürt und eliminiert werden. Dieser Aspekt ist ebenfalls von enorm großer Bedeutung, da redundante Einflussgrößen bei der Modellgenerierung zu einer Vervielfachung der Regeln führen, ohne dass dabei eine Erhöhung der Modellgenauigkeit, wohl aber ein deutlicher Verlust an Transparenz und Interpretierbarkeit der generierten Modelle erzielt wird.

Ein weiteres herausragendes Merkmal des konzipierten evolutionären Selektionsverfahrens besteht in der *Berücksichtigung und Optimierung der Partitionierung des Merkmalsraumes*. Wie in Abschnitt 2.1.4 ausführlich dargelegt, ist für eine datenbasierte Generierung von Fuzzy-Modellen mit hoher Modellierungsgüte die Bereitstellung einer den Datenstrukturen gut angepassten Fuzzy-Partitionierung des Merkmalsraumes erforderlich. Daraus ergibt sich die logische Konsequenz, auch bei der Relevanzbewertung eines potenziellen Merkmalsatzes eine Partitionierung des Merkmalsraumes zu berücksichtigen. Statt einer einfachen Berücksichtigung einer vorgegebenen Partitionierung ist in dem Verfahren zur Abschätzung der maximal möglichen Relevanz eines potenziellen Merkmalsatzes eine Optimierung der Partitionierung implementiert worden. Erst hierdurch ist erstmalig die Möglichkeit zu einer einheitlichen vergleichenden Beurteilung verschiedener Merkmalsätze mit individuell partitionierten Merkmalsräumen geschaffen worden.

Aufgrund des starken nichtlinearen Einflusses der Partitionierung und der damit verbundenen Multimodalität der Relevanzbewertungsfunktion werden höchste Anforderungen an das eingesetzte Optimierungsverfahren — der Evolutionsstrategie — gestellt. Aufgrund der Komplexität des formulierten Optimierungsproblems sind einige Strategieelemente zur *Effizienzsteigerung von Evolutionsstrategien* motiviert und umgesetzt worden (Kapitel 5). Der wesentliche Kern besteht dabei in der Einführung adaptiv diskretisierter Suchräume.

Die daraus resultierende neue Variante einer adaptiv diskretisierten Evolutionsstrategie ist prinzipiell für alle Optimierungsprobleme geeignet, wo in hochgradig multimodalen Fitnesslandschaften ein möglichst sicheres und schnell konvergierendes Auffinden *robuster Lösungen* im Sinne eines guten Kompromisses zwischen den Gesichtspunkten *Rechenaufwand* und *resultierendes Optimierungsergebnis* hergestellt werden soll. Die *Robustheit* einer gefundenen Lösung wird dabei durch die detektierbare Größe desjenigen umliegenden Gebietes gekennzeichnet, welches ausschließlich Punkte mit annähernd gleichwertigen Fitnesswerten beinhaltet, so dass kleine Variationen im Parameterraum nicht zu nennenswert schlechteren Lösungen führen.

Der deutlich Effizienzgewinn für einen Anwender beim Einsatz sämtlicher Strategieelemente für komplexe Modellierungsprobleme wird in Kapitel 6 verdeutlicht. Durch die der Modellierung vorgelagerte evolutionäre Selektion relevanter und nichtredundanter Merkmalsätze ist in allen Anwendungsbeispielen eine deutliche Reduzierung der bei der Fuzzy-Modellierung berücksichtigten Einflussgrößen (*Reduktion des Suchraumes*) erzielt worden. Dies führt in nahezu allen Modellierungsergebnissen zu einer Verschlinkung der generierten Regelbasis (*gesteigerte Interpretierbarkeit* und *reduzierter Aufwand bei der Modellauswertung*) bei gleichzeitiger Erhöhung der erzielten Modellqualität (*hohe Modellierungsgüte*). Zusätzlich erhält erst hierdurch der Anwender durch die *Angabe der relevanten Einflussgrößen* einen transparenten Systemeinblick dahingehend, in welchem Grad die beobachteten Einflussgrößen einen prozessbestimmenden Einfluss ausüben.

Auch in Bezug auf die Parametrierung des nachgeschalteten Fuzzy-Modellierungsverfahrens wird der Anwender entscheidend entlastet. Durch die automatisierte Generierung datenadaptierter Fuzzy-Zugehörigkeitsfunktionen für alle selektierten Einflussgrößen und der zu modellierenden Ausgangsgröße ergibt sich ein stark (*verminderter Parametrisierungsaufwand*) vor dem Start des eigentlichen Regelgenerierungsprozesses. Die gute Anpassung der optimierten Partitionierungen an die jeweiligen Datenstrukturen zeigt sich eindrucksvoll bei den untersuchten Beispielen in den erzielten hohen Modellierungsgüten.

Ausgehend von den überragenden Ergebnissen bei Modellierungsproblemen ist die Grundidee der evolutionären Merkmalsselektion, der iterativen graduellen Verfeinerung mit Optimierung, auf einen schrittweisen relevanzbasierten Entwurf von TSK–Fuzzy–Reglern für nichtlineare Systeme übertragen worden (Kapitel 6.2.5). Der wesentliche Vorteil dieser neuen evolutionären Methode zum Reglerentwurf besteht darin, dass eine strukturelle Anpassung und Parameteroptimierung kombiniert angewendet wird. Anstatt eine feste komplexe Reglerstruktur mit einer Vielzahl von Parametern direkt zu optimieren, wird die Strukturkomplexität des Reglers schrittweise erhöht und auf Grundlage bereits bekannter guter Lösungen mit geringerer Komplexität vorinitialisiert und anschließend evolutionär optimiert.

Anhang A

Eigenschaften einer Fuzzy-Menge

Die Zuordnung eines linguistischen Wertes zu einer Fuzzy-Menge \mathcal{F} gemäß Abschnitt 3.2 wird über eine Zugehörigkeitsfunktion $\mu_{\mathcal{F}}(x)$ definiert.

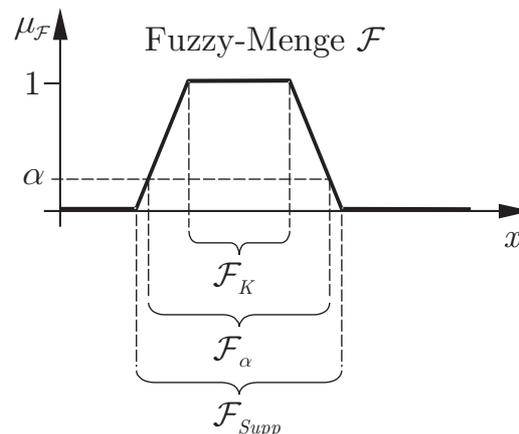


Abbildung A.1: Kenngrößen einer normalisierten Fuzzy-Menge.

Die wesentlichsten Kenngrößen einer Fuzzy-Menge sind in Abbildung A.1 illustriert. Dabei wird der *Support* \mathcal{F}_{Supp} (auch *Träger* genannt) einer Fuzzy-Menge definiert durch die Menge aller Werte x des Grundbereichs \mathcal{G} , für die die Zugehörigkeit $\mu_{\mathcal{F}}(x)$ positive Werte annimmt:

$$\mathcal{F}_{Supp} = \{x \in \mathcal{G} \mid \mu_{\mathcal{F}}(x) > 0\}. \quad (\text{A.1})$$

Der *Kern* \mathcal{F}_K der Fuzzy-Menge beinhaltet die Menge aller Werte x des Grundbereichs \mathcal{G} , deren Zugehörigkeit $\mu_{\mathcal{F}}(x)$ genau den Wert Eins annimmt:

$$\mathcal{F}_K = \{x \in \mathcal{G} \mid \mu_{\mathcal{F}}(x) = 1\}. \quad (\text{A.2})$$

Stimmen *Support* und *Kern* überein, so ist \mathcal{F} eine aus der Booleschen Logik bekannte scharfe Menge, die durch eine rechteckförmige Zugehörigkeitsfunktion charakterisiert wird.

Zusätzlich wird häufig der *Alpha-Schnitt* \mathcal{F}_α zur Charakterisierung einer Fuzzy-Menge herangezogen. Dieser beinhaltet die Menge aller Werte x des Grundbereichs \mathcal{G} , deren Zugehörigkeit $\mu_{\mathcal{F}}(x)$ Werte größer α annehmen:

$$\mathcal{F}_\alpha = \{x \in \mathcal{G} \mid \mu_{\mathcal{F}}(x) > \alpha\}. \quad (\text{A.3})$$

Aus den Kenngrößen für Zugehörigkeitsfunktionen ergeben sich zwei wichtige Eigenschaften für Fuzzy-Mengen:

Normalität: Eine Fuzzy-Menge, deren Zugehörigkeitsfunktion einen Maximalwert von Eins besitzt, wird *normal* genannt:

$$\max_{x \in \mathcal{G}} \{\mu_{\mathcal{F}}(x)\} = 1. \quad (\text{A.4})$$

Konverität: Eine Fuzzy-Menge, bei der alle Alpha-Schnitte \mathcal{F}_α mit $0 \leq \alpha \leq 1$ direkt zusammenhängen, wird *konvex* genannt. Somit muss für beliebige $x_1, x_2, x_3 \in \mathcal{X}$ gelten:

$$x_1 \leq x_2 \leq x_3 \quad \Rightarrow \quad \min\{\mu_{\mathcal{F}}(x_1), \mu_{\mathcal{F}}(x_3)\} \leq \mu_{\mathcal{F}}(x_2). \quad (\text{A.5})$$

Gilt anstelle von „ \leq “ die Relation „ $<$ “, so heißt die Fuzzy-Menge *streng konvex*.

Anhang B

Statistische Tests zur Regelbewertung im Fuzzy–ROSA–Verfahren

Die Grundlage für die Regeltest- und Bewertungsstrategien nach Definition B.1, B.2, B.3 und B.4 sind die gegebenenfalls durch Konfidenzintervalle abgeschätzten bedingten und unbedingten Wahrscheinlichkeiten für das Auftreten einer Konklusion $p(C|P)$ und $p(C)$. Da die Wahrscheinlichkeiten im Allgemeinen nicht exakt bestimmt werden können, werden sie im Fuzzy–ROSA–Verfahren, basierend auf den Lerndaten, wie folgt geschätzt [Kro99]:

$$\begin{aligned}\hat{p}(C) &= \frac{1}{N} \cdot \sum_{q=1}^N \mu_{Y_j}(y_q) \\ \hat{p}(C|P) &= \frac{1}{N_{sup}} \cdot \sum_{q=1}^N \mu_{Y_j}(y_q) \wedge \mu_P(\mathbf{x}_q) .\end{aligned}\tag{B.1}$$

Die Summation erfolgt über alle Datenpunkte $d_q = (\mathbf{x}_q, y_q)$ mit $q = 1, \dots, N$. Der Daten–Support N_{sup} ist wie folgt definiert:

$$N_{sup} := \sum_{q=1}^N \mu_P(\mathbf{x}_q) .\tag{B.2}$$

Zur statistischen Absicherung des Regeltests werden die Schätzer für die Wahrscheinlichkeiten nach Gleichung (B.1) durch die obere bzw. untere Grenze der einseitigen Konfidenzintervalle $I_L^\alpha = [0, \hat{p}_L^\alpha]$ und $I_U^\alpha = [\hat{p}_U^\alpha, 1]$ ersetzt [HEK95]. Die Konfidenzintervalle geben in Abhängigkeit der verwendeten Datenbasis den Bereich an, in denen sich die geschätzte Wahrscheinlichkeit \hat{p} mit einer vorgegebenen Irrtumswahrscheinlichkeit, auch α –Niveau genannt, befindet. Darauf aufbauend werden die Bewertungsindizes der Regeltest- und Bewertungsstrategien im Fuzzy–ROSA–Verfahren wie folgt definiert:

Definition B.1 (*Relevanzindex J_{RI}*) Der Regeltest und die Bewertung mit dem Relevanzindex J_{RI} ist wie folgt definiert:

$$\begin{aligned} \text{Test:} \quad & \hat{p}_U^\alpha(C|P) > \hat{p}_L^\alpha(C) \Rightarrow \text{positive Regel} \\ & \hat{p}_L^\alpha(C|P) < \hat{p}_U^\alpha(C) \Rightarrow \text{negative Regel} \end{aligned}$$

$$J_{RI} := \begin{cases} \frac{\hat{p}_U^\alpha(C|P) - \hat{p}_L^\alpha(C)}{1 - \hat{p}(C)} & \text{positive Regel} \\ \frac{\hat{p}_U^\alpha(C) - \hat{p}_L^\alpha(C|P)}{\hat{p}(C)} & \text{negative Regel} \end{cases} .$$

Getestet wird also, ob sich die bedingte Wahrscheinlichkeit für das Auftreten der Konklusion signifikant von der unbedingten Wahrscheinlichkeit unterscheidet.

Definition B.2 (*Relevante Trefferquote J_{RHR}*) Der Regeltest und die Bewertung mit der relevanten Trefferquote J_{RHR} ist wie folgt definiert:

$$\begin{aligned} \text{Test:} \quad & \hat{p}_U^\alpha(C|P) > \hat{p}_L^\alpha(C) \Rightarrow \text{positive Regel} \\ & \hat{p}_L^\alpha(C|P) < \hat{p}_U^\alpha(C) \Rightarrow \text{negative Regel} \end{aligned}$$

$$J_{RHR} := \begin{cases} \hat{p}(C|P) & \text{positive Regel} \\ 1 - \hat{p}(C|P) & \text{negative Regel} \end{cases} .$$

Im Gegensatz zum Relevanzindex wird nicht der normierte Abstand der Konfidenzintervalle für die Bewertung verwendet, sondern die Trefferquote auf den Lerndaten.

Definition B.3 (*Normierte Trefferquote J_{NHR}*) Der Regeltest und die Bewertung mit der normierten Trefferquote J_{NHR} ist wie folgt definiert:

$$\begin{aligned} \text{Test:} \quad & \hat{p}(C|P) > \theta \Rightarrow \text{positive Regel} \\ & \hat{p}(C|P) < \theta' \Rightarrow \text{negative Regel} \end{aligned}$$

$$J_{NHR} := \begin{cases} \frac{\hat{p}(C|P) - \theta}{1 - \theta} & \text{positive Regel} \\ \frac{\theta' - \hat{p}(C|P)}{\theta'} & \text{negative Regel} \end{cases} .$$

Getestet wird also, ob die Trefferquote auf den Lerndaten größer als eine bestimmte Schwelle θ ist. Dabei ist im Fuzzy-ROSA-Verfahren die Schwelle für positive Regeln mit $\theta = .5$ so gewählt, dass eine Regel mehr „Richtiges“ als „Falsches“ empfiehlt. Entsprechend ist der Standardwert für die Schwelle auch bei negativen Regeln durch $\theta' = 0,5$ gegeben.

Definition B.4 (*Konfidente Trefferquote J_{CHR}*) Der Regeltest und die Bewertung mit der konfidenten Trefferquote J_{CHR} ist wie folgt definiert:

$$\begin{aligned} \text{Test:} \quad & \hat{p}_U^\alpha(C|P) > \theta \Rightarrow \text{positive Regel} \\ & \hat{p}_L^\alpha(C|P) < \theta' \Rightarrow \text{negative Regel} \\ J_{CHR} & := \begin{cases} \frac{\hat{p}_U^\alpha(C|P) - \theta}{1 - \theta} & \text{positive Regeln} \\ \frac{\theta' - \hat{p}_L^\alpha(C|P)}{\theta'} & \text{negative Regeln} \end{cases} . \end{aligned}$$

Abweichend von der normierten Trefferquote wird hier die Trefferquote auf den Lerndaten durch das Konfidenzintervall nach unten abgeschätzt.

Im Gegensatz zu den bisher vorgestellten Bewertungsindizes wird für den mittelwertbasierten Index getestet, ob sich der bedingte Mittelwert der Ausgangsgröße \bar{y}_P in der durch die Prämisse P spezifizierten Eingangssituation signifikant von dem unbedingten Mittelwert \bar{y} der Ausgangsgröße unterscheidet. Im Fuzzy-Fall wird der Mittelwert \bar{y}_P und die Standardabweichung σ_P der bedingten Verteilung wie folgt geschätzt [JS98]:

$$\begin{aligned} \bar{y}_P &= \frac{1}{N_{sup}} \cdot \sum_{q=1}^N y_q \cdot \mu_P(\mathbf{x}_q) \quad \text{und} \\ \sigma_P &= \sqrt{\frac{\sum_{q=1}^N (y_q - \bar{y}_P)^2 \cdot \mu_P(\mathbf{x}_k)}{N_{sup} - 1}} . \end{aligned} \tag{B.3}$$

Definition B.5 (*Mittelwertbasierter Index J_{MVB}*) Der Regeltest und die Bewertung mit dem mittelwertbasierten Index J_{MVB} ist für positive Regeln wie folgt definiert:

$$\left| \frac{\bar{y}_P - \bar{y}}{\sigma_P / \sqrt{N_{sup}}} \right| > t(N_{sup}, \alpha) \Rightarrow J_{MVB} := e^{\sigma_P / \sigma} .$$

Der kritische Wert $t(N_{sup}, \alpha)$ für die Zurückweisung einer Hypothese kann durch Approximation berechnet oder Tabellen entnommen werden [HEK95]. Der linguistische Wert für die Konklusion der Regel wird so festgelegt, dass der Erfülltheitsgrad für den bedingten Mittelwert y_P maximal wird: $\mu_{Y,j}(\bar{y}_P) = \{\max \mu_{Y,j}(\bar{y}_P) : 1 \leq j \leq L_Y\}$.

Anhang C

Least–Squares–Algorithmus für allgemeine Funktionsansätze

Die $m + 1$ Parameter a_j einer Funktion

$$f(\mathbf{x}) = a_0 + \sum_{j=1}^m a_j \cdot g_j(\mathbf{x}) \quad (\text{C.1})$$

mit m konstanten Elementarfunktionen $g_j(\mathbf{x})$ sollen so identifiziert werden, dass die Datenmenge

$$\mathcal{D} = \{d_1, \dots, d_N\} \quad \text{mit} \quad d_i = (\mathbf{x}_i \quad y_i)^T \quad (\text{C.2})$$

durch die Funktion (C.1) im Sinne einer Minimierung des quadratischen Fehlers approximiert wird. Mit $\mathbf{g}(\mathbf{x}) = (1 \quad g_1(\mathbf{x}) \dots g_m(\mathbf{x}))^T$ und $\mathbf{a} = (a_0 \dots a_m)^T$ ergibt sich, dass eine Gütefunktion

$$J(\mathbf{a}) = \sum_{i=1}^N (\mathbf{g}^T(\mathbf{x}_i) \mathbf{a} - y_i)^2 \quad (\text{C.3})$$

minimiert werden muss. Zur Bestimmung des globalen Minimums muss das folgende Gleichungssystem gelöst werden:

$$\begin{aligned} \text{grad } J(\mathbf{a}) &= \mathbf{0} \\ \Leftrightarrow \text{grad} \left[\sum_{i=1}^N (\mathbf{g}^T(\mathbf{x}_i) \mathbf{a} - y_i)^2 \right] &= \mathbf{0} \\ \Leftrightarrow \sum_{i=1}^N 2(\mathbf{g}^T(\mathbf{x}_i) \mathbf{a} - y_i) \cdot \mathbf{g}(\mathbf{x}_i) &= \mathbf{0} \\ \Leftrightarrow \sum_{i=1}^N \mathbf{g}(\mathbf{x}_i) \mathbf{g}^T(\mathbf{x}_i) \mathbf{a} &= \sum_{i=1}^N y_i \cdot \mathbf{g}(\mathbf{x}_i) \quad . \end{aligned} \quad (\text{C.4})$$

Dies lässt sich mit

$$\mathbf{G} = \sum_{i=1}^N \mathbf{g}(\mathbf{x}_i) \mathbf{g}^T(\mathbf{x}_i) \quad \text{und} \quad \mathbf{b} = \sum_{i=1}^N y_i \cdot \mathbf{g}(\mathbf{x}_i) \quad (\text{C.5})$$

schreiben als

$$\mathbf{G} \cdot \mathbf{a} = \mathbf{b} \quad . \quad (\text{C.6})$$

Unter der Voraussetzung, dass $m + 1$ linear unabhängige Datenpunkte gegeben sind, lässt sich das Gleichungssystem (C.6) lösen. Die Funktion, die sich aus der Lösung des Gleichungssystems ergibt, approximiert die Punkte im Sinne der Minimierung der Fehlerquadrate. Eine mögliche Umsetzung des Verfahrens für die Anwendung auf einem Computer findet sich in [PTVF99].

Literaturverzeichnis

- [AAS96] ALBRECHT, J. ; ALBERS, K. ; STELZNER, P.: Sequentielle vorausschauende Vorsteuerung eines Kraftwerksparks durch ein Führungssystem. In: *at – Automatisierungstechnik* 44 (1996), Nr. 8, S. 381–390
- [ACPF96] ABREU, A. ; CUSTODIO, L. ; PINTO-FERREIRA, C.: Fuzzy Modelling: A Rule Based Approach. In: *Proceedings of the Fifth IEEE International Conference on Fuzzy Systems (FUZZ-IEEE '96), New Orleans, LA, USA* Bd. 1. Piscataway, NJ : IEEE Press, 1996, S. 162–168
- [AKAS96] ALBRECHT, J. ; KIENDL, H. ; ALBERS, K. ; STELZNER, P.: Vorausschauende wirtschaftliche Lastaufteilung mit Reserveoptimierung durch ein Kraftwerksführungssystem. In: *Tagungsband GMA-Kongreß 96: Meß- und Automatisierungstechnik*. Düsseldorf : VDI Verlag, 1996 (VDI-Berichte, Nr. 1282), S. 249–258
- [AKM⁺95] ALBRECHT, J. ; KIENDL, H. ; MICHALSKE, A. ; ALBERS, K. ; STELZNER, P.: *Verfahren zum Regeln der Leistung eines Kraftwerksparks*, 1995. – Patent, DE 195 10 342
- [Alb97] ALBRECHT, J.: *Vorausschauende optimale Steuer- und Regelstrategien zur Verbesserung der Kraftwerksführung*. Düsseldorf : VDI Verlag, 1997 (Fortschritt-Berichte VDI, Reihe 8, Nr. 616)
- [And35] ANDERSON, E.: The IRISes of the Gaspé Peninsula. In: *Bull. Amer. IRIS Soc.* 59 (1935), S. 2–5
- [Bas96] BASTIAN, A.: Sequential Fuzzy System Identification. In: *Control and Cybernetics* 25 (1996), Nr. 2, S. 199–223
- [Bäc94] BÄCK, T.: *Evolutionary Algorithms in Theory and Practice*, Fachbereich Informatik, Universität Dortmund, Diss., 1994

- [BDB97] BERSINI, H. ; DUCHATEAU, A. ; BRADSHAW, N.: Using Incremental Learning Algorithms in the Search for Minimal and Effective Fuzzy Models. In: *Proceedings of the Sixth IEEE International Conference on Fuzzy Systems (FUZZ-IEEE '97), Barcelona, Spanien* Bd. 3. Piscataway, NJ : IEEE Press, 1997, S. 1417–1422
- [Bel61] BELLMAN, R. E.: *Adaptive control processes: A guided tour*. Princeton : Princeton University Press, 1961
- [Ben97] BENMAKROUHA, F.: Parameter Identification in a Fuzzy System with Insufficient Data. In: *Proceedings of the Sixth IEEE International Conference on Fuzzy Systems (FUZZ-IEEE '97), Barcelona, Spanien, 1997* Bd. 1. Piscataway, NJ : IEEE Press, 1997, S. 537–542
- [Bez81] BEZDEK, J.C.: *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York, USA : Plenum Press, 1981
- [BHS97] BÄCK, T. ; HAMMEL, U. ; SCHWEFEL, H.-P.: Evolutionary Computation: Comments on the History and Current State. In: *IEEE Transactions on Evolutionary Computation* 1 (1997), Nr. 1, S. 3–17
- [BKC96] BONISSONE, P. ; KHEDKAR, P. ; CHEN, Y.: Genetic algorithms for automated tuning of fuzzy controllers: A transportation application. In: *Proceedings of the Fifth IEEE International Conference on Fuzzy Systems (FUZZ-IEEE '96), New Orleans, LA, USA* Bd. 2. Piscataway, NJ : IEEE Press, 1996, S. 674–680
- [BLSK00] BLUMENDELLER, W. ; LEUTNER, V. ; SCHWANE, U. ; KIENDL, H.: Regleroptimierung mit evolutionären Algorithmen. In: *2. Internationales Fluidtechnisches Kolloquium*. Dresden, 2000, S. 437–452
- [BP92] BEZDEK, J.C. ; PAUL, S.K.: Pattern Recognition: Models and Approaches. In: BEZDEK, J.C. (Hrsg.) ; PAUL, S.K. (Hrsg.): *Fuzzy Models for Pattern Recognition*. Piscataway, NJ : IEEE Press, 1992, Kapitel Background, Significance, and Key Points, S. 6–28
- [Bri98] BRINDA, T.: *Optimierungsverfahren zur Berücksichtigung des Ein-/Ausgangsverhaltens bei der Generierung relevanter Regeln für ein Fuzzy-System*, Universität Dortmund, Fachbereich Informatik, Diplomarbeit, 1998
- [BS96] BÄCK, T. ; SCHWEFEL, H.-P.: Evolutionary Computation: An Overview. In: *Proceedings of 1996 IEEE International Conference*

on Evolutionary Computation (ICEC '96), Nayoya, Japan. Piscataway, NJ, USA : IEEE-Press, 1996, S. 20–29

- [BSKNL96] BABUSKA, R. ; SETNES, M. ; KAYMAK, U. ; VAN NAUTA LEMKE, H.R.: Rule Base Simplification with Similarity Measures. In: *Proceedings of the Fifth IEEE International Conference on Fuzzy Systems (FUZZ-IEEE '96)*, New Orleans, USA, 1996 Bd. 3. Piscataway, NJ : IEEE Press, 1996, S. 1642–1647
- [CERK96] CECHIN, A.L. ; EPPERLEIN, U. ; ROSENSTIEL, W. ; KOPPENHOFER, B.: The Extraction of Sugeno Fuzzy Rules from Neural Networks. In: *Proceedings of the Fourth European Symposium on Artificial Neural Networks, ESANN '96*. Brüssel, Belgien : D facta, 1996, S. 49–54
- [CF97] CASTELLANO, G. ; FANELLI, A.M.: An Approach to Structure Identification of Fuzzy Models. In: *Proceedings of the Sixth IEEE International Conference on Fuzzy Systems (FUZZ-IEEE '97)*, Barcelona, Spanien Bd. 1. Piscataway, NJ : IEEE Press, 1997, S. 531–536
- [CFM96] CARSE, B. ; FOGARTY, T.C. ; MUNRO, A.: Evolving Fuzzy Rule Based Controllers Using Genetic Algorithms. In: *Fuzzy Sets and Systems* 80 (1996), S. 273–293
- [CH97] CORDON, O. ; HERRERA, F.: Evolutionary Design of TSK Fuzzy Rule-Based Systems Using (μ, λ) -Evolution Strategies. In: *Proceedings of the Sixth IEEE International Conference on Fuzzy Systems (FUZZ-IEEE '97)*, Barcelona, Spanien, 1997 Bd. 1. Piscataway, NJ : IEEE Press, 1997, S. 509–514
- [CHHL98] CHEN, C.-L. ; HSU, S.-H. ; HSIEH, C.-T. ; LIN, W.-K.: Generating Crisp-Type Fuzzy Models from Operating Data. In: *Proceedings of the Seventh IEEE International Conference on Fuzzy Systems (FUZZ-IEEE '98)*, Anchorage, USA Bd. 1. Piscataway, NJ : IEEE Press, 1998, S. 686–691
- [Chi96] CHIU, S.: Method and Software for Extracting Fuzzy Classification Rules by Subtractive Clustering. In: *Proceedings of the 1996 Biennial Conference of the North American Fuzzy Information Processing Society (NAFIPS)*, New York, USA. Piscataway, NJ : IEEE Press, 1996, S. 461–465

- [CHL96] CORDON, O. ; HERRERA, F. ; LOZANO, M.: A Three-Stage Method for Designing Genetic Fuzzy Systems by Learning From Examples. In: *Proceedings of the Fourth International Conference on Parallel Problem Solving from Nature, PPSN IV*. Berlin : Springer, 1996, S. 720–729
- [CHMV01] CORDON, O. ; HERRERA, F. ; MAGDALENA, L. ; VILLAR, P.: A genetic learning process for the scaling factors, granularity and contexts of the fuzzy rule-based system data base. In: *Information Sciences* 136 (2001), Nr. 1
- [CHV00] CORDON, O. ; HERRERA, F. ; VILLAR, M.: Analysis and guidelines to obtain a good uniform fuzzy partition granularity for fuzzy rule-based systems using simulated annealing. In: *International Journal of Approximate Reasoning* 25 (2000), Nr. 3, S. 187–216
- [CL98] CHEN, M.-Y. ; LINKENS, D.A.: A Fast Fuzzy Modelling Approach Using Clustering Neural Networks. In: *Proceedings of the Seventh IEEE International Conference on Fuzzy Systems (FUZZ-IEEE '98), Anchorage, USA* Bd. 2. Piscataway, NJ : IEEE Press, 1998, S. 1088–1093
- [CM98] CHERKASSKY, V. ; MULIER, F.: *Learning from Data*. New York, USA : John Wiley & Sons, 1998
- [CS94] CORCORAN, A. L. ; SANDIP, S.: Using Real-Valued Genetic Algorithms to Evolve Rule Sets for Classifications. In: *Proceedings of the First IEEE Conference on Evolutionary Computation (ICEC '94), Orlando, USA, 1994* Bd. 1. Piscataway, NJ : IEEE Press, 1994, S. 120–124
- [CTR99] COSTA, J.C. ; TAVARES, R. ; ROSA, A.: An experimental study on dynamic random variation of population size. In: *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics (IEEE SMC '99), Tokyo, Japan* Bd. 1. Piscataway, NJ : IEEE Press, 1999, S. 607–612
- [CW96] CHO, K.B. ; WANG, B.H.: Radial Basis Function Based Adaptive Fuzzy Systems and their Applications to System Identification and Prediction. In: *Fuzzy Sets and Systems* 83 (1996), S. 325–339
- [DKS95] DOUGHERTY, James ; KOHAVI, Ron ; SAHAMI, Mehran: Supervised and unsupervised discretization of continuous features. In:

- Proc. International Conference on Machine Learning*, 1995, S. 194–202
- [DM98] DENG, K. ; MOORE, A.: On the greediness of feature selection algorithms. In: *Proc. International Conference on Machine Learning (ICML)*, 1998
- [DS78] DIXON, L. C. W. ; SZEGÖ, G. P.: The Global Optimization Problem: An Introduction. In: DIXON, L. C. W. (Hrsg.) ; SZEGÖ, G. P. (Hrsg.): *Towards Global Optimization II*. Amsterdam, Niederlande : North-Holland Publishing Company, 1978
- [Eng03] ENGELBRECHT, Andries P.: *Computational Intelligence – An Introduction*. New York, USA : Wiley, 2003
- [Fag96] FAGARASAN, F.: A Genetic-Based Method Applied in Fuzzy Modeling. In: *Proceedings of 1996 IEEE International Conference on Evolutionary Computation (ICEC '96), New York, USA, 1996*. Piscataway, NJ : IEEE Press, 1996, S. 253–257
- [Gol89] GOLDBERG, D. E.: *Genetic Algorithms in Search, Optimization and Machine Learning*. Reading, USA : Addison Wesley, 1989
- [GSJ97] GOMEZ-SKARMETA, A.F. ; JIMENEZ, F.: Generating and Tuning Fuzzy Rules Using Hybrid Systems. In: *Proceedings of the Sixth IEEE International Conference on Fuzzy Systems (FUZZ-IEEE '97), Barcelona, Spanien* Bd. 1. Piscataway, NJ : IEEE Press, 1997, S. 247–252
- [Hae99] HAENDEL, L.: *Entwicklung und Implementierung eines adaptiven Fuzzy-Klassifikators zur akustischen Qualitätssicherung*, Universität Dortmund, Lehrstuhl für Elektrische Steuerung und Regelung, Diplomarbeit, 1999
- [HAK00] HINNEBURG, Alexander ; AGGARWAL, Charu C. ; KEIM, Daniel A.: What is the nearest neighbor in high dimensional spaces? In: *The VLDB Journal* (2000), S. 506–515
- [HC97] HUANG, Y.-P. ; CHEN, Y.-R.: Modeling a Fuzzy System by the Integrated Virtual and Genetic Algorithms. In: *Proceedings of the Sixth IEEE International Conference on Fuzzy Systems (FUZZ-IEEE '97), Barcelona, Spanien* Bd. 1. Piscataway, NJ : IEEE Press, 1997, S. 521–526

- [HC98] HONG, T.-P. ; CHEN, J.-B.: Building a Concise Decision Table for Fuzzy Rule Induction. In: *Proceedings of the Seventh IEEE International Conference on Fuzzy Systems (FUZZ-IEEE '98), Anchorage, USA, 1998* Bd. 2. Piscataway, NJ : IEEE Press, 1998, S. 997–1002
- [HC99] HOWE, N. ; CARDIE, C.: Weighting unusual feature types / Cornell University. 1999 (TR99-1735). – Forschungsbericht
- [HEK95] HARTUNG, J. ; ELPELT, B. ; KLOESENER, K.-H.: *Statistik, 10. Auflage*. München : Oldenbourg Verlag, 1995
- [HHN86] HOLLAND, J.H. ; HOLYOAK, K.J. ; NISBETT, R.E.: *Induction: Processes of Inference, Learning and Discovery*. Cambridge, Mass. [u.a.] : MIT Press, 1986
- [HHS05] HOFFMANN, F. ; HÖLEMANN, S. ; SCHAUTEN, D.: Augmentation of TSK Fuzzy Controllers by Evolutionary Optimization. In: *Proceedings of the International Workshop of Genetic Fuzzy Systems, (GFS 2005)*. Spain : Granada, 2005, S. 16–22
- [HKK97] HÖPPNER, F. ; KLAWONN, F. ; KRUSE, R.: *Fuzzy-Clusteranalyse*. Braunschweig : Vieweg, 1997
- [HL96] HALL, L.O. ; LANDE, P.: Generating Fuzzy Rules from Data. In: *Proceedings of the Fifth IEEE International Conference on Fuzzy Systems (FUZZ-IEEE '96), New Orleans, USA, 1996* Bd. 3. Piscataway, NJ : IEEE Press, 1996, S. 1757–1762
- [HLTM99] HUSSAIN, F. ; LIU, H. ; TAN, C. L. ; M., Dash: Discretization: An enabling technique / School of Computing, National University of Singapore. 1999 (TRC6/99). – Forschungsbericht
- [HLV93] HERRERA, F. ; LOZANO, M. ; VERDEGAY, J.L.: Generating Fuzzy Rules from Examples Using Genetic Algorithms / Department of Computer Science and Artificial Intelligence, Universität Granada, Spanien. 1993 (DECSAI-93115). – Forschungsbericht
- [HM92] HOMAIFAR, A. ; MCCORMICK, E.: Full Design of Fuzzy Controllers Using Genetic Algorithms. In: *Neural and Stochastic Methods in Image and Signal Processing, SPIE-1766*, The International Society for Optical Engineering, 1992, S. 393–404
- [Hol75] HOLLAND, J. H.: *Adaptation in Natural and Artificial Systems*. Michigan, USA : Universität Michigan Press, Ann Arbor, 1975

- [Hol98] HOLVE, R.: Investigation of Automatic Rule Generation for Hierarchical Fuzzy Systems. In: *Proceedings of the Seventh IEEE International Conference on Fuzzy Systems (FUZZ-IEEE '98), Anchorage, USA, 1998* Bd. 2. Piscataway, NJ : IEEE Press, 1998, S. 973–978
- [Hor97] HORN, J.: Multicriterion decision making. In: BÄCK, T. (Hrsg.) ; FOGEL, D. B. (Hrsg.) ; MICHALEWICZ, Z. (Hrsg.): *Handbook of Evolutionary Computation*. New York, NY, USA and Bristol, Großbritannien : IOP Publishing Ltd and Oxford University Press, 1997, S. F1.9: 1–15
- [HR78] HARRISON, D. ; RUBINFELD, D. L.: Hedonic Prices and the Demand for Clean Air. In: *Economics & Management* 5 (1978), S. 81–102
- [HS05] HOFFMANN, F. ; SCHAUTEN, D.: Strukturelle Evolution von Fuzzy-Reglern am Beispiel eines inversen Pendels. In: MIKUT, R. (Hrsg.) ; REISCHL, M. (Hrsg.): *Proceedings of the 15. Workshop Computational Intelligence*, 2005, S. 197–211
- [HSH07] HOFFMANN, F. ; SCHAUTEN, D. ; HÖLEMANN, S.: Incremental Evolutionary Design of TSK Fuzzy Controllers. In: *Special Issue des IEEE Transactions on Fuzzy Systems*. Piscataway, NJ : IEEE Press, 2007
- [IDC97] IZQUIERDO, J.M.C. ; DIMITRIADIS, Y.A. ; CORONADO, J.L.: Fas-Back: Matching–Error Based Learning for Automatic Generation of Fuzzy Logic Systems. In: *Proceedings of the Sixth IEEE International Conference on Fuzzy Systems (FUZZ-IEEE '97), Barcelona, Spanien* Bd. 3. Piscataway, NJ : IEEE Press, 1997, S. 1561–1566
- [INM96] ISHIBUCHI, H. ; NAKASHIMA, T. ; MURATA, T.: Genetic–Algorithm–Based Approaches to the Design of Fuzzy Systems for Multi–Dimensional Pattern Classification Problems. In: *Proceedings of 1996 IEEE International Conference on Evolutionary Computation (ICEC '96), New York, USA, 1996*. Piscataway, NJ : IEEE Press, 1996, S. 229–234
- [ISUT98] IMAMURA, K. ; SHINOHARA, K. ; UMANO, M. ; TAMURA, H.: Acquisition of Fuzzy Rules from Data Including Qualitative Attributes Using Fuzzy Neural Networks with Forgetting. In: *Proceedings of the Seventh IEEE International Conference on Fuzzy Systems*

- (FUZZ-IEEE '98), Anchorage, USA Bd. 2. Piscataway, NJ : IEEE Press, 1998, S. 1118–1123
- [IV94] IMAM, I. ; VAFAIE, H.: An empirical comparison between global and greedy-like search for variable selection. In: *Florida AI Research Symposium (FLAIRS)*, 1994
- [Jac91] JACKSON, J. E.: *A User's Guide to Principal Components*. New York : Wiley, 1991
- [JD88] JAIN, A.K. ; DUBES, R.C.: *Algorithms for Clustering Data*. New Jersey, USA : Prentice Hall, Englewood Cliffs, 1988
- [Jes00a] JESSEN, H.: *Test- und Bewertungsverfahren zur regelbasierten Modellierung und Anwendung in der Lastprognose*. Düsseldorf : VDI Verlag, 2000 (Fortschritt-Berichte VDI, Reihe 8, Nr. 836)
- [Jes00b] JESSEN, H.: Test and Rating Strategies for Automatic Fuzzy Rule Generation and Application to Load Prediction. In: MOHAMMADIAN, M. (Hrsg.): *New Frontiers in Computational Intelligence and its Applications*. Amsterdam, Niederlande : IOS Press, 2000, S. 11–21
- [JKP94] JOHN, G. H. ; KOHAVI, R. ; PFLEGER, K.: Irrelevant features and the subset selection problem. In: *Proc. International Conference on Machine Learning*, 1994. – Journal version in AIJ, S. 121–129
- [JMMB00] JÄKEL, J. ; MIKUT, R. ; MALBERG, H. ; BRETTHAUER, G.: Datenbasierte Regelsuche für Fuzzy-Systeme mittels baumorientierter Verfahren. In: *at – Automatisierungstechnik* 48 (2000), Nr. 7, S. 317–326
- [JS98] JESSEN, H. ; SLAWINSKI, T.: Test and Rating Strategies for Data-based Rule Generation. In: *Reihe Computational Intelligence*, CI-39/98, Sonderforschungsbereich 531, Universität Dortmund, 1998, S. 1–9
- [JSS98] JIN, Y. ; VON SEELEN, W. ; SENDHOFF, B.: An Approach to Rule-Based Knowledge Extraction. In: *Proceedings of the Seventh IEEE International Conference on Fuzzy Systems (FUZZ-IEEE '98)*, Anchorage, USA Bd. 2. Piscataway, NJ : IEEE Press, 1998, S. 1188–1193
- [KA97] KIENDL, H. ; ALBRECHT, J.: *Verfahren zur sequentiellen Vorsteuerung eines Prozesses*, 1997. – Patent, DE 195 10 343

- [Kar91] KARR, C.L.: Genetic Algorithms for Fuzzy Controllers. In: *AI Expert* (1991), S. 393–404
- [KFR95] KRONE, A. ; FRENCK, C. ; RUSSAK, O.: Design of a Fuzzy Controller for an Alkoxylation Process Using the ROSA Method for Automatic Rule Generation. In: *Proceedings of the Third European Congress on Intelligent Techniques and Soft Computing (EUFIT '95)* Bd. 2. Aachen : Verlag Mainz, 1995, S. 760–764
- [KG93] KARR, C.L ; GENTRY, E.: Fuzzy Control of pH Using Genetic Algorithms. In: *IEEE Transactions on Fuzzy Systems* 1 (1993), S. 46–53
- [Kie93a] KIENDL, H.: Fuzzy-Regler mit Hyperinferenz- und Hyperdefuzzifizierungsstrategie. In: REUSCH, B. (Hrsg.): *Fuzzy Logik – Theorie und Praxis, Reihe Informatik aktuell*. Berlin : Springer, 1993, S. 42–51
- [Kie93b] KIENDL, H.: Theorie für den Anwender: Fuzzy Control (Teil 9), Hyperinferenz und Hyperdefuzzifizierung. In: *at – Automatisierungstechnik* 41 (1993), Nr. 11, S. A33–A36
- [Kie94a] KIENDL, H.: The inference filter. In: *Second European Congress on Intelligent Techniques and Soft Computing (EUFIT '94)* Bd. 1. Aachen : Verlag Mainz, 1994, S. 443–452
- [Kie94b] KIENDL, H.: Verfahren zur Erzeugung von Stellgrößen am Ausgang eines Fuzzy-Reglers und Fuzzy-Regler hierfür. 1994. – Patent. DE 43 08 083
- [Kie96] KIENDL, H.: Verfahren zur Defuzzifizierung für signalverarbeitende Fuzzy-Baueinheiten. 1996. – Patent. DE 19 640 635
- [Kie97a] KIENDL, H.: *Fuzzy Control methodenorientiert*. München : Oldenbourg, 1997
- [Kie97b] KIENDL, H.: Non-translation-invariant Defuzzification. In: *Proceedings of the Sixth IEEE International Conference on Fuzzy Systems (FUZZ-IEEE '97), Barcelona, Spanien, 1997*. Piscataway, NJ : IEEE Press, 1997, S. 737–742
- [Kie99] KIENDL, H.: Decision Analysis by Advanced Fuzzy Systems. In: L. ZADEH, J. K. (Hrsg.): *Computing with Words in Information/Intelligent Systems*. Heidelberg : Physica-Verlag, 1999, S. 223–242

- [Kin94] KINNEBROCK, W.: *Neuronale Netze*. 2nd. München Wien : Oldenbourg, 1994
- [KJ97] KOHAVI, K. ; JOHN, G. H.: Wrappers for feature subset selection. In: *Artificial Intelligence 97* (1997), Nr. 1–2, S. 273–324
- [KK89] KIENDL, H. ; KRABS, M.: Ein Verfahren zur Generierung regelbasierter Modelle für dynamische Systeme. In: *at – Automatisierungstechnik* 37 (1989), Nr. 11, S. 423–430
- [KK94] KRONE, A. ; KIENDL, H.: Automatic Generation of Positive and Negative Rules for Two–Way Fuzzy Controllers. In: *Proceedings of the Second European Congress on Intelligent Techniques and Soft Computing (EUFIT '94)* Bd. 1. Aachen : Verlag Mainz, 1994, S. 438–447
- [KK95] KRABS, M. ; KIENDL, H.: Anwendungsfelder der automatischen Regelgenerierung mit dem ROSA–Verfahren. In: *at – Automatisierungstechnik* 43 (1995), Nr. 6, S. 269–276
- [KK97] KRONE, A. ; KIENDL, H.: An Evolutionary Concept for Generating Relevant Fuzzy Rules from Data. In: *International Journal of Knowledge–based Intelligent Engineering Systems* 1 (1997), Nr. 4, S. 207–213
- [KKK94] KINZEL, J. ; KLAWONN, F. ; KRUSE, R.: Anpassung genetischer Algorithmen zum Erlernen und Optimieren von Fuzzy–Reglern. In: *Tagungsband der 4. Dortmunder Fuzzy–Tage*, 1994, S. 103–110
- [KKN96] KIENDL, H. ; KNICKER, R. ; NIEWELS, F.: Two Way Fuzzy Controllers Based on Hyperinference and Inference Filter. In: *Proceedings of World Automation Congress (WAC '96)*, Montpellier, Frankreich. Albuquerque, NM, USA : TSI Enterprise Inc., 1996
- [KKN97] KRUSE, R. ; KLAWONN, F. ; NAUCK, D.: Erlernen von Fuzzy–Regeln. In: *Informatik Forschung und Entwicklung*. Berlin : Springer, 1997, S. 2–6
- [KKS00a] KRAUSE, P. ; KRONE, A. ; SLAWINSKI, T.: Fuzzy System Identification by Generating and Evolutionary Optimizing Fuzzy Rule Bases Consisting of Relevant Fuzzy Rules, 2000
- [KKS00b] KRONE, A. ; KRAUSE, P. ; SLAWINSKI, T.: A New Rule Reduction Method for Finding Interpretable and Small Rule Bases in High

- Dimensional Search Spaces. In: *Proceedings of the Ninth IEEE International Conference on Fuzzy Systems, (FUZZ-IEEE '00), San Antonio, USA* Bd. 2. Piscataway, NJ : IEEE Press, 2000, S. 696–699
- [Kra94] KRABS, M.: *Das ROSA-Verfahren zur Modellierung dynamischer Systeme durch Regeln mit statistischer Relevanzbewertung*. Düsseldorf : VDI Verlag, 1994 (Fortschritt-Berichte VDI, Reihe 8, Nr. 404)
- [Kra99] KRAUSE, P.: Demonstration of the Use of Hyperinference: Position Control Using an Electric Disc-Motor. In: REUSCH, B. (Hrsg.): *Computational Intelligence: Theory and Applications (Proceedings of 6th Fuzzy-Days), Dortmund*. Berlin : Springer, 1999, S. 671–675
- [Kra00] KRAUSE, P.: Generierung von Takagi-Sugeno-Fuzzy-Systemen aus relevanten Fuzzy-Regeln. In: *Proceedings 10. Workshop Fuzzy Control des GMA-FA 5.22, Dortmund*. Karlsruhe : Wissenschaftliche Berichte Forschungszentrum Karlsruhe (FZKA 6509), 2000, S. 84–97
- [Kra01] KRAUSE, P.: *Datenbasierte Generierung von transparenten und genauen Fuzzy-Modellen für mehrdeutige Daten und komplexe Systeme*. Düsseldorf : VDI Verlag, 2001 (Fortschritt-Berichte VDI, Reihe 10, Nr. 691)
- [Kro96] KRONE, A.: Advanced Rule Reduction Concepts for Optimizing Efficiency of Knowledge Extraction. In: *Proceedings of the Fourth European Congress on Intelligent Techniques and Soft Computing (EUFIT '96)* Bd. 2. Aachen : Verlag Mainz, 1996, S. 919–923
- [Kro99] KRONE, A.: *Datenbasierte Generierung von relevanten Fuzzy-Regeln zur Modellierung von Prozesszusammenhängen und Bedienstrategien*. Düsseldorf : VDI Verlag, 1999 (Fortschritt-Berichte VDI, Reihe 10, Nr. 615)
- [KS96] KRONE, A. ; SCHWANE, U.: Generating Fuzzy Rules from Contradictory Data of Different Control Strategies and Control Performances. In: *Proceedings of the Fifth IEEE International Conference on Fuzzy Systems (FUZZ-IEEE '96), New Orleans, USA, 1996* Bd. 1. Piscataway, NJ : IEEE Press, 1996, S. 492–497
- [KS98] KRONE, A. ; SLAWINSKI, T.: Data-Based Extraction of Unidimensional Fuzzy Sets for Fuzzy Rule Generation. In: *Proceedings of the*

Seventh IEEE International Conference on Fuzzy Systems (FUZZ-IEEE '98), Anchorage, USA, 1998 Bd. 2. Piscataway, NJ : IEEE Press, 1998, S. 1032–1037

- [KS01] KRAUSE, P. ; SLAWINSKI, T.: Das Fuzzy–ROSA–Verfahren: Von der regelorientierten statistischen Analyse zur datenbasierten Generierung von interpretierbaren Takagi–Sugeno–Systemen. In: *at – Automatisierungstechnik* 9 (2001), Nr. 49, S. 391–399
- [KSK00] KRONE, A. ; SLAWINSKI, T. ; KRAUSE, P.: Search Space Structuring as a Key to Cope with Different Problem Sizes in the Field of Fuzzy Modeling. In: JAMSHIDI, M (Hrsg.) ; FATHI, M. (Hrsg.) ; FURUHASHI, T. (Hrsg.): *Soft Computing, Multimedia, and Image Processing (Proceedings of the Fourth World Automation Congress (WAC '00), Maui, USA)* Bd. 11. Albuquerque, NM, USA : TSI Press, 2000, S. 346–354
- [KT98] KRONE, A. ; TAEGER, H.: Relevance Test for Fuzzy Rules. In: *Reihe Computational Intelligence, CI–40/98*, Sonderforschungsbereich 531, Universität Dortmund, 1998
- [KW00] KOJADINOVIC, I. ; WOTTKA, T.: Comparison between a filter and a wrapper approach to variable subset selection in regression problems. In: *Proc. European Symposium on Intelligent Techniques (ESIT)*, 2000
- [KWS00] KRAUSE, P ; WIESMANN, D. ; SLAWINSKI, T.: Parallel Evolutionary Algorithms for Optimizing Data–Based Generated Fuzzy Systems. In: *Reihe Computational Intelligence, CI–101/00*, Sonderforschungsbereich 531, Universität Dortmund, 2000
- [Lie99] LIESKE, D.: *Objektive Geräuschprüfung zur Qualitätssicherung von Automatikgetrieben*, Technische Universität Dresden, Diss., 1999
- [LT93] LEE, M.A. ; TAKAGI, H.: Dynamic Control of Genetic Algorithms Using Fuzzy Logic Techniques. In: *Proceedings of the Fifth International Conference on Genetic Algorithms* (1993), S. 76–83
- [LW98] LUTZ, H. ; WENDT, W.: *Taschenbuch der Regelungstechnik*. 2nd. Frankfurt am Main, Thun : Harry Deutsch, 1998
- [MA75] MAMDANI, E.H. ; ASSILIAN, S.: An experiment in linguistic synthesis with a fuzzy logic controller. In: *International Journal of Man Machine Studies* 7 (1975), S. 1–13

- [Mam77] MAMDANI, E.H.: Application of Fuzzy Logic to Approximate Reasoning Using Linguistic Systems. In: *IEEE Transaction on Computing* 26 (1977), S. 1182–1191
- [Mat96] MATHAR, R.: *Informationstheorie*. Stuttgart : Teubner Verlag, 1996
- [MBC⁺99] MIKUT, R. ; BÖHLMANN, S. ; CUNO, B. ; JÄKEL, J. ; KROLL, A. ; RAUSCHENBACH, T. ; PFEIFFER, B.-M. ; SLAWINSKI, T.: Fuzzy-Logic und Fuzzy-Control — Begriffe und Definitionen. In: *Tagungsband des 9. Workshop Fuzzy Control des GMA 1.4.2, Dortmund*. Dortmund : Forschungsbericht der Fakultät für Elektrotechnik, Nr. 0499, Universität Dortmund, 1999, S. 76–91
- [MG77] MACKEY, M. ; GLASS, L.: Oscillation and Chaos in Physiological Control Systems. In: *Science* 197 (1977), S. 287–289
- [MG81] MAMDANI, E. H. ; GAINES, B. R.: *Fuzzy Reasoning and its Applications*. London : Academic Press, 1981
- [MST94] MICHIE, D. ; SPIEGELHALTER, D.J. ; TAYLOR, C.C.: *Machine learning, Neural and Statistical Classification*. Hemel Hempstead, Großbritannien : Ellis Horwood, 1994
- [MT97] MASTOROCOSTAS, P. ; THEOCHARIS, J.: Orthogonal Least Squares Fuzzy Modeling of Nonlinear Dynamical Systems. In: *Proceedings of the Sixth IEEE International Conference on Fuzzy Systems (FUZZ-IEEE '97), Barcelona, Spanien, 1997* Bd. 2. Piscataway, NJ : IEEE Press, 1997, S. 1147–1152
- [NFM96] NELLES, O. ; FISCHER, M. ; MÜLLER, B.: Fuzzy Rule Extraction by a Genetic Algorithm and Constrained Nonlinear Optimization of Membership Functions. In: *Proceedings of the Fifth IEEE International Conference on Fuzzy Systems (FUZZ-IEEE '96), New Orleans, USA, 1996* Bd. 1. Piscataway, NJ : IEEE Press, 1996, S. 213–219
- [NGF96] NAKOULA, Y. ; GALICHET, S. ; FOULLOY, L.: Simultaneous Learning of Rules and Linguistic Terms. In: *Proceedings of the Fifth IEEE International Conference on Fuzzy Systems (FUZZ-IEEE '96), New Orleans, USA, 1996* Bd. 3. Piscataway, NJ : IEEE Press, 1996, S. 1743–1749

- [Nis97] NISSEN, V.: *Einführung in Evolutionäre Algorithmen Optimierung nach dem Vorbild der Evolution*. Wiesbaden : Vieweg Verlag, 1997
- [NK97] NAUCK, D. ; KRUSE, R.: What are neuro-fuzzy classifiers? In: *Proc. 7th International Fuzzy Systems Association World Congress IFSA '97* Bd. 4, Academica Prague, 1997, S. 228–233
- [NK98] NAUCK, D. ; KRUSE, R.: A Neuro-fuzzy Approach to Obtain Interpretable Fuzzy Systems for Function Approximation. In: *Proceedings of the Seventh IEEE International Conference on Fuzzy Systems (FUZZ-IEEE '98)*, Anchorage Bd. 2. Piscataway, NJ : IEEE Press, 1998, S. 1106–1111
- [NKBS01] NICOLAUS, B. ; KIENDL, H. ; BLUMENDELLER, W. ; SCHWANE, U.: Evolutionary Optimization of an Industrial Hydraulic Valve with the help of a Fuzzy Performance-Index. In: *Proceedings of the Tenth IEEE International Conference on Fuzzy Systems (FUZZ-IEEE '01)*, Melbourne Bd. CD-ROM. Piscataway, NJ : IEEE Press, 2001
- [ON96] OULIDDREN, K. ; NOWE, A.: An Approach to Incremental Fuzzy Modelling of Dependencies in Complex Systems. In: *1996 Biennial Conference of the North American Fuzzy Information Processing Society (NAFIPS)*, New York, USA. Piscataway, NJ : IEEE Press, 1996, S. 479–483
- [PKS99] PRACZYK, J. ; KIENDL, H. ; SLAWINSKI, T.: Finding Relevant Process Characteristics with a Method for Data-Based Complexity Reduction. In: REUSCH, B. (Hrsg.): *Computational Intelligence: Theory and Applications (Proceedings of 6th Fuzzy-Days)*. Berlin : Springer, 1999, S. 548–555
- [Poh00] POHLHEIM, H.: *GEATbx: Genetic and Evolutionary Algorithm Toolbox for use with MATLAB*, 2000. – (Online-Handbuch, <http://www.geatbx.com/docu/index.html>)
- [PPB97] PAL, N. R. ; PAL, K. ; BEZDEK, J. C.: A Mixed c-Means Clustering Model. In: *Proceedings of the Sixth IEEE International Conference on Fuzzy Systems (FUZZ-IEEE '97)*, Barcelona, Spanien, 1997 Bd. 1. Piscataway, NJ : IEEE Press, 1997, S. 11–21
- [PR96] PINGAN, Z. ; RENHOU, L.: A New Approach to Fuzzy Identification for Complex Systems. In: *Proceedings of the Fifth IEEE International Conference on Fuzzy Systems (FUZZ-IEEE '96)*, New Orleans, USA, 1996. Piscataway, NJ : IEEE Press, 1996, S. 1308–1313

- [Pra99] PRACZYK, J.: *Entwurf von Fuzzy-Gütemaßen zur Prozeßbewertung*. Düsseldorf : VDI Verlag, 1999 (Fortschritt-Berichte VDI, Reihe 8, Nr. 796)
- [Pre94] PRECHELT, L.: PROBEN 1 – A Set of Neural Network Benchmark Problems and Benchmarking Rules / Fakultät für Informatik, Universität Karlsruhe. 1994 (21). – Forschungsbericht
- [PSK98] PRACZYK, J. ; SLAWINSKI, T. ; KIENDL, H.: Auswahl relevanter Prozeßmerkmale für einen Fuzzy-Klassifikator durch ein Verfahren zur datenbasierten Komplexitätsreduktion. In: *Tagungsband des 8. Workshop Fuzzy Control VDI/VDE GMA-FA 5.22*, Forschungsbericht der Fakultät für Elektrotechnik, Nr. 0298, Universität Dortmund, 1998
- [PTVF99] PRESS, W. H. ; TEUKOLSKY, S. A. ; VETTERLING, W. T. ; FLANERNY, B. P.: *Numerical Recipes in C*. 2nd. Cambridge, England : Cambridge University Press, 1999
- [Qui93] QUINLAN, J. R.: *C4.5: Programs for Machine Learning*. San Mateo : Morgan Kaufman, 1993
- [Rob99] Robert Bosch GmbH, Geschäftsbereich Automationstechnik: *Elektrohydraulische Proportional- und Regelungssysteme*. 1999
- [Roj96] ROJAS, R.: *Theorie der neuronalen Netze*. Berlin : Springer, 1996
- [Ros60] ROSENBROCK, H. H.: An Automatic Method for Finding the Greatest or Least Value of a Function. In: *The Computer Journal* 3 (1960), Nr. 3, S. 175–184
- [RW96] RONG, L. ; WANG, Z.: An Algorithm of Extracting Fuzzy Rules Directly from Numerical Examples by Using FNN. In: *Proceedings of 1996 IEEE International Conference on Systems, Man and Cybernetics. Information Intelligence and Systems, New York, USA*. Piscataway, NJ : IEEE Press, 1996, S. 1067–1072
- [RY96] RENHOU, L. ; YI, Z.: Fuzzy Logic Controller Based on Genetic Algorithms. In: *Fuzzy Sets and Systems* 83 (1996), S. 1–10
- [Sar02] SARLE, Warren S.: *Neural Network FAQ*. 2002. – Official FAQ of the Usenet newsgroup `comp.ai.neural-nets` (<ftp://ftp.sas.com/pub/neural/FAQ.html>)

- [Sch77] SCHWEFEL, H.-P.: *Numerische Optimierung von Computer-Modellen mittels der Evolutionsstrategie*. Basel und Stuttgart : Birkhäuser, 1977
- [Sch87] SCHWEFEL, H.-P.: Collective Phenomena in Evolutionary Systems. In: *Preprints of the 31st Annual Meeting of the International Society for General System Research, Budapest, Ungarn* Bd. 2, 1987, S. 1025–1033
- [Sch95] SCHWEFEL, H.-P.: *Evolution and Optimum Seeking*. New York, USA : Sixth-Generation Computer Technology Series, 1995
- [Sch02] SCHÄFER, S.: *Entwicklung und Erprobung von Methoden zur adaptiven Diskretisierung von Evolutionsstrategien zur Effizienzsteigerung und Bestimmung robuster Lösungen*, Lehrstuhl für Elektrische Steuerung und Regelung, Universität Dortmund, Diplomarbeit, 2002
- [SHF94] SCHÖNEBURG, E. ; HEINZMANN, F. ; FEDDERSEN, S.: *Genetische Algorithmen und Evolutionsstrategien*. Bonn : Addison-Wesley, 1994
- [SK97] SPRICKMANN-KERKERINCK, A.: *Entwicklung eines Fuzzy-Logik gestützten Klassifikationssystems zur akustischen Serienprüfung von Automatikgetrieben*, Universität Dortmund, Lehrstuhl für Elektrische Steuerung und Regelung, Diplomarbeit, 1997
- [SKBB98] SETNES, M. ; KOENE, A. ; BABUSKA, R. ; BRUIJN, P.M.: Data-Driven Initialization and Structure Learning in Fuzzy Neural Networks. In: *Proceedings of the Seventh IEEE International Conference on Fuzzy Systems (FUZZ-IEEE '98), Anchorage* Bd. 2. Piscataway, NJ : IEEE Press, 1998, S. 1147–1152
- [SKH⁺99] SLAWINSKI, T. ; KRONE, A. ; HAMMEL, U. ; WIESMANN, D. ; KRAUSE, P.: A Hybrid Evolutionary Search Concept for Data-based Generation of Relevant Fuzzy Rules in High Dimensional Spaces. In: *Proceedings of IEEE International Conference on Fuzzy Systems (FUZZ-IEEE '99) Seoul, Korea, 1999* Bd. 3. Piscataway, NJ : IEEE Press, 1999, S. 1432–1437
- [SKK⁺99a] SLAWINSKI, T. ; KRAUSE, P. ; KRONE, A. ; HAMMEL, U. ; WIESMANN, D.: Fuzzy Adaption der evolutionären Regelsuche im Fuzzy-ROSA-Verfahren. In: *Tagungsband des 9. Workshops Fuzzy Control des GMA-FA 5.22, Dortmund*. Dortmund : Forschungsbericht

der Fakultät für Elektrotechnik, Nr. 0499, Universität Dortmund, 1999, S. 14–39

- [SKK99b] SLAWINSKI, T. ; KRONE, A. ; KRAUSE, P.: Efficient Design of a Complete Rule Search in Sparsely Populated Search Spaces. In: *Reihe Computational Intelligence*, CI-77/99, Sonderforschungsbereich 531, Universität Dortmund, 1999
- [SKKK01] SLAWINSKI, T. ; KRONE, A. ; KRAUSE, P. ; KIENDL, H.: The Fuzzy-ROSA Method: A Statistically Motivated Fuzzy Approach for Data-Based Generation of Small Interpretable Rule Bases in High-Dimensional Search Spaces. In: LAST, M. (Hrsg.) ; KANDEL, A. (Hrsg.) ; BUNKE, H. (Hrsg.): *Data Mining and Computational Intelligence*. Heidelberg : Physica-Verlag, 2001, S. 141–166
- [SKMM03] SCHAUTEN, D. ; KIENDL, H. ; MEYER, J. ; MACHE, D.H.: Rekonstruktionsbasierte Selektion relevanter Einflussgrößen. In: *Reihe Computational Intelligence*, CI-154/03, Sonderforschungsbereich 531, Universität Dortmund, 2003
- [Sla01] SLAWINSKI, T.: *Analyse und effiziente Generierung von relevanten Fuzzy-Regeln in hochdimensionalen Suchräumen*. Düsseldorf : VDI Verlag, 2001 (Fortschritt-Berichte VDI, Reihe 10, Nr. 686)
- [SM98] SHI, Y. ; MIZUMOTO, M.: An Improvement of Neuro-Fuzzy Learning Algorithm for Tuning Fuzzy Rules Based on Fuzzy Clustering Method. In: *Proceedings of the Seventh IEEE International Conference on Fuzzy Systems (FUZZ-IEEE '98)*, Anchorage Bd. 2. Piscataway, NJ : IEEE Press, 1998, S. 991–996
- [Smi80] SMITH, S.F. *A Learning System Based on Genetic Adaptive Systems*. University of Pittsburgh, PhD Dissertation. 1980
- [Spr95] SPRAVE, J.: Evolutionäre Algorithmen zur Parameteroptimierung. In: *Automatisierungstechnik* 43 (1995), S. 109–115
- [Spr99] SPRAVE, Joachim: A Unified Model of Non-panmictic Population Structures in Evolutionary Algorithms. In: ANGELINE, P. J. (Hrsg.) ; PORTO, V. W. (Hrsg.): *Proceedings of the 1999 Congress on Evolutionary Computation (CEC'99)* Bd. 2. Piscataway NJ : IEEE Press, July 6–9, 1999. – ISBN 0-7803-5536-9, S. 1384–1391

- [SPS⁺99] SLAWINSKI, T. ; PRACZYK, J. ; SCHWANE, U. ; KRONE, A. ; KIENDL, H.: Data-based Generation of Fuzzy Rules for Classification, Prediction and Control with the Fuzzy-ROSA Method. In: *European Control Congress (ECC '99), Karlsruhe* Bd. CD-ROM, 1999
- [SS95] SCHLITTEGEN, R. ; STREITBERG, B.: *Zeitreihenanalyse*. 6. überarbeitete und erweiterte Auflage. München : Oldenbourg Verlag, 1995
- [SSK01] SCHAUTEN, D. ; SLAWINSKI, T. ; KIENDL, H.: Datenbasierte Generierung von situationsbezogenen Entscheidungsregeln für die Kraftwerksführung. In: *Reihe Computational Intelligence, CI-110/01*, Sonderforschungsbereich 531, Universität Dortmund, 2001
- [SSP⁺98] SLAWINSKI, T. ; SCHWANE, U. ; PRACZYK, J. ; KRONE, A. ; JESSEN, H. ; KIENDL, H. ; LIESKE, D.: Application of WINROSA for Controller Adaptation in Robotics and Classification in Quality Control. In: *Reihe Computational Intelligence, CI-46/98*, Sonderforschungsbereich 531, Universität Dortmund, 1998
- [Sur94] SURMANN, H. *Systematischer rechnergestützter Entwurf unscharfer regelbasierter Systeme*. Universität Dortmund, Fakultät für Elektrotechnik, Dissertation. 1994
- [SV01] SCHMITT, A. ; VERSTEGE, J.: Multikriterielle Optimierung auf Basis von Evolutionsstrategien zur Bereitstellung von Systemdienstleistungen. In: *Tagungsband zur VDI-Tagung Optimierung in der Energieversorgung, Veitshöchheim*. Düsseldorf : VDI Verlag, 2001. – (VDI Berichte 1627), S. 225–234
- [TH97] TAN, G.V. ; HU, X.: More on Designing Fuzzy Controllers Using Genetic Algorithms: Guided Constrained Optimisation. In: *Proceedings of the Sixth IEEE International Conference on Fuzzy Systems (FUZZ-IEEE '97), Barcelona, Spanien, 1997* Bd. 1. Piscataway, NJ : IEEE Press, 1997, S. 497–502
- [TS85] TAKAGI, T. ; SUGENO, M.: Fuzzy Identification of Systems and its Application to Modeling and Control. In: *IEEE Transactions on Systems, Man, and Cybernetics* 15 (1985), Nr. 1, S. 116–132
- [Wei02] WEICKER, K.: *Evolutionäre Algorithmen*. Stuttgart, Deutschland : Teubener Verlag, 2002

- [WM00] WILSON, D. R. ; MARTINEZ, Tony R.: An integrated instance-based learning algorithm. In: *Computational Intelligence* 16 (2000), Nr. 1, S. 1–28
- [Wol94] WOLF, T.: Optimization of Fuzzy Systems Using Neural Networks and Genetic Algorithms. In: *Proceedings of the Fourth European Congress on Intelligent Techniques and Computing (EUFIT '94)* Bd. 1. Aachen : Verlag Mainz, 1994, S. 544–551
- [WR96] WILLIAMS, C. K. I. ; RASMUSSEN, C. E.: Gaussian Processes for Regression. In: D. S. TOURETZKY, M. E. H. (Hrsg.): *Advances in Neural Information Processing Systems*. Cambridge, Mass. [u.a.] : MIT Press, 1996
- [Wys97] WYSOTZKI, F.: Maschinelles Lernen. In: *at – Automatisierungstechnik* 45 (1997), Nr. 11, S. 526–536
- [Xio01] XION, N.: *Designing Compact and Comprehensible Fuzzy Controllers Using Genetic Algorithms (Entwurf kompakter und interpretierbarer Fuzzy Controller mittels Genetischer Algorithmen)*. Aachen : Shaker Verlag, 2001 (Berichte aus der Automatisierungstechnik)
- [YWG98] YEN, J. ; WANG, L. ; GILLEPSIE, W.: A Global–Local Learning Algorithm for Identifying Takagi–Sugeno–Kang Fuzzy Models. In: *Proceedings of the Seventh IEEE International Conference on Fuzzy Systems (FUZZ–IEEE '98), Anchorage, 1998* Bd. 2. Piscataway, NJ : IEEE Press, 1998, S. 967–972
- [Zad65] ZADEH, L.A.: Fuzzy Sets. In: *Information Control* 8 (1965), S. 338–353
- [Zad68] ZADEH, L.H.: Probability measures of fuzzy events. In: *Journal of Mathematical Analysis and Applications* 23 (1968), S. 421–427
- [ZGW96] ZHAO, J. ; GOREZ, R. ; WERTZ, V.: Genetic Algorithms for the Elimination of Redundancy and/or Rule Contribution Assessment in Fuzzy Models. In: *Mathematics and Computers in Simulation* Bd. 41. Niederlande : Elsevier, 1996, S. 139–148
- [Zhe92] ZHENG, L.: A practical guide to tune of proportional and integral (PI) like fuzzy controllers. In: *Proceedings of the Firth IEEE International Conference on Fuzzy Systems (FUZZ–IEEE '92), San Diego, 1992* Bd. 2. Piscataway, NJ : IEEE Press, 1992, S. 633–640

[Zim95] ZIMMERMANN, H.-J.: *Datenanalyse*. Düsseldorf : VDI Verlag, 1995