

Robuste datenbasierte Fuzzy-Modellierung komplexer Systeme: neue Strategieelemente und systematische Entwurfsstrategien

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

von

Martin Sternke

Dortmund 2007

Tag der mündlichen Prüfung:	14. Mai 2007
Hauptreferent	Prof. em. Dr. rer. nat. H. Kiendl
Korreferent	Prof. Dr.-Ing. V. Krebs

Absence of Evidence
is not
Evidence of Absence

C. Sagan
Astrophysiker und Schriftsteller

Vorwort

Die vorliegende Arbeit entstand während meiner Tätigkeit am Lehrstuhl für Systemanalyse und am Lehrstuhl für Regelungssystemtechnik der Universität Dortmund. Die Arbeit wurde angeregt und betreut von Herrn Prof. em. Dr. rer. nat. H. Kiendl. Auf ihn geht die Grundidee des Fuzzy-ROSA-Verfahrens zurück. Ihm danke ich herzlich für die stete Förderung meiner wissenschaftlichen Arbeit. Herrn Prof. Dr.-Ing. V. Krebs danke ich für die Übernahme des Korreferats.

Der Deutschen Forschungsgemeinschaft danke ich für die finanzielle Förderung meiner Arbeit im Rahmen des Transferbereichs 37 *Datenbasierte Generierung und Optimierung von Modellen mit Methoden der Computational Intelligence zur Lösung industrieller Anwendungsprobleme* der Universität Dortmund. Herrn Prof. a.D. Dr.-Ing. H.-P. Schwefel, der diesen Transferbereich initiiert und als Sprecher geleitet hat, danke ich ebenfalls herzlich.

Für die gute Zusammenarbeit bedanke ich mich bei der NuTech Solutions GmbH, Dortmund, die als Kooperationspartner auf Seiten der Industrie im Transferbereich mitgewirkt hat. Mein besonderer Dank gilt dabei Herrn Dr.-Ing. P. Krause.

Ferner bedanke ich mich bei allen Mitarbeiterinnen und Mitarbeitern des Lehrstuhls für Regelungssystemtechnik für die vielfältige Unterstützung und die gute Zusammenarbeit. Besonderer Dank gilt dabei den Herren Dr. rer. nat. F. Hoffmann, T. Nierobisch und D. Schauten. Wertvolle Hinweise formaler Art für die endgültige Fassung des Manuskripts verdanke ich außerdem Frau G. Kasimir. Insbesondere bedanke ich mich auch bei Herrn Univ.-Prof. Dr.-Ing. Prof. h. c. T. Bertram als Inhaber des Lehrstuhls für Regelungssystemtechnik für die Möglichkeit, meine Arbeit wie gewünscht abzuschließen.

Dortmund, Mai 2007

Martin Sternke

Inhaltsverzeichnis

1	Einleitung	1
2	Einordnung und Zielsetzung	5
2.1	Modelle	5
2.2	Modellierung	6
2.3	Übersicht etablierter Modellansätze	9
2.4	Ziele dieser Arbeit	11
3	Grundlagen der Fuzzy-Modellierung	13
3.1	Fuzzy-Logik	13
3.2	Mamdani-Fuzzy-Systeme	14
3.3	Takagi-Sugeno-Kang-Fuzzy-Systeme	18
4	Das Fuzzy-ROSA-Verfahren	19
4.1	Grundidee des Verfahrens	19
4.2	Projektdefinition	20
4.3	Generierung, Analyse und Optimierung einer Regelbasis	24
5	Analyse des Verfahrens im Hinblick auf Verbesserungspotenzial	29
5.1	Anforderungen industrieller Anwender	29
5.2	Abgleich des Verfahrens mit den Anforderungen	31

5.3	Anwenderorientierte Erweiterungen des Verfahrens	33
6	Neue Strategieelemente für die Generierung von Mamdani-Regeln	37
6.1	Test- und Bewertungsverfahren basierend auf Datenhäufigkeiten	37
6.2	Test- und Bewertungsverfahren basierend auf Datenverteilungen	40
6.3	Medianbasiertes Test- und Bewertungsverfahren	44
6.4	Regelgenerierung nach dem Konzept des Inkrementellen Relevanzindex . .	48
7	Neue Strategieelemente für Takagi-Sugeno-Kang-Systeme	51
7.1	Transformation einer Mamdani-Regelbasis in eine TSK-Regelbasis	51
7.2	Bisher angewandte Strategien zur Parameterbestimmung	57
7.2.1	Unterbestimmte lineare Gleichungssysteme	57
7.2.2	Überbestimmte lineare Gleichungssysteme	58
7.3	Parameterbestimmung über QR-Zerlegung mit Spaltenpivotierung	60
7.4	Parameterbestimmung über Iteratively Reweighted Least-Squares	62
7.4.1	Motivation	62
7.4.2	Verfahrensweise	63
7.4.3	Anwendungsbeispiel	69
7.5	Berechnung des Ausgabewertes über den gewichteten Median	70
8	Hybrides Fuzzy-System	73
8.1	Mamdani-Systeme mit negativen Regeln	74
8.2	Zur Übertragbarkeit des Prinzips negativer Regeln auf TSK-Systeme . . .	78
8.3	Ansatz für ein hybrides Fuzzy-System	79
8.4	Anwendungsbeispiel	81

9 Einsatz für komplexe Anwendungen und Ableitung von Einstellregeln	85
9.1 Motivation und Zielsetzungen	86
9.2 Experiment-Design	87
9.3 Einsatz der neuen Strategieelemente für komplexe Anwendungen	94
9.4 Zusammenfassende Darstellung und Bewertung der Modellierungsergebnisse	132
9.5 Qualitative Empfehlungen für die Einstellung freier Modellparameter . . .	136
9.6 Modul zur Prognose des Modellierungsfehlers	138
10 Zusammenfassung und Ausblick	142
Anhang	
A Test- und Bewertungsverfahren für die Regelgenerierung	146
B Vorstellung der untersuchten Anwendungen und Benchmarkprobleme	149
C Maße zur Beurteilung der Modellierungsgüte	161
Literaturverzeichnis	164

Kapitel 1

Einleitung

Modellen kommt in allen Bereichen der Wissenschaft eine große Bedeutung zu. Modelle helfen dem Menschen, komplexe Wirkungszusammenhänge und Vernetzungen besser verstehen zu können und ermöglichen es ihm letztlich, entsprechende Systeme seinen Wünschen und Vorstellungen gemäß beeinflussen zu können. Die Anforderungen, die an Modelle gestellt werden, sind dabei je nach primär verfolgter Zielsetzung durchaus unterschiedlich. In den Bereichen der Wirtschafts- und Sozialwissenschaften steht häufig das Ziel im Vordergrund, die zu Grunde liegenden Wirkungszusammenhänge innerhalb eines Systems aufzudecken und so ein besseres Verständnis über das System zu gewinnen. Die Systeme, die in diesen Wissenschaftsbereichen betrachtet werden, basieren in der Regel gerade nicht auf physikalischen Gesetzmäßigkeiten, sondern sind vielmehr von menschlichen Handlungsweisen, deren Interaktion und daraus resultierenden Effekten bestimmt, die sich nur schwer in Formeln fassen lassen, wie beispielsweise die Entwicklung an Aktienmärkten. In den Bereichen der Natur- und Ingenieurwissenschaften steht hingegen meist die Forderung nach einem möglichst exakten Modell im Vordergrund. Die dort betrachteten Systeme unterliegen physikalischen Gesetzmäßigkeiten, die jedoch in vielen Fällen derartig komplex sind, dass deren präzise Formulierung zu aufwendig ist. Die Wirklichkeit muss durch das Modell einerseits so weit vereinfacht werden, dass der Aufwand für die Modellerstellung und der Umfang des Modells vertretbar sind. Andererseits muss die Wirklichkeit durch das Modell aber auch hinreichend genau angenähert werden, wie beispielsweise bei der Simulation von Auto-Crashtests. Ein Modell bildet den Ausgangspunkt für eine numerische Simulation, eine Optimierung oder eine Regelung des modellierten Systems. Durch die stetig fortschreitende Entwicklung der Computertechnologie und der daraus resultierenden Rechenleistung haben sich in den letzten Jahren neue Anwendungsfelder für die Modellierung erschlossen bzw. werden sich auch zukünftig noch erschließen, die bisher aufgrund ihrer Komplexität kaum oder gar nicht zugänglich waren, wie beispielsweise Modelle zur langfristigen Klimaentwicklung.

Der Prozess der Modellbildung umfasst zwei wesentliche Teilaufgaben: Strukturwahl und

Parameterbestimmung. Die Strukturwahl legt fest, welcher Modellansatz (zum Beispiel Differentialgleichungen) zum Einsatz kommt. Nach dieser Festlegung werden die in der gewählten Struktur enthaltenen freien Parameter (zum Beispiel die Koeffizienten in den Differentialgleichungen) so bestimmt, dass das Modellverhalten möglichst gut mit dem realen Systemverhalten übereinstimmt. Im Hinblick auf den zu wählenden Modellansatz kann grundsätzlich zwischen quantitativen und qualitativen Modellen unterschieden werden. Quantitative Modelle zielen darauf ab, das Eingangs-/Ausgangsverhalten des zu modellierenden Systems möglichst genau nachzubilden, ohne dabei Wert auf die Interpretierbarkeit des Modells zu legen. Der Vorteil liegt in diesem Fall darin, dass im Prinzip keinerlei Wissen über die Ursache-Wirkung-Zusammenhänge des zu modellierenden Systems vorhanden sein muss. Der gravierende Nachteil liegt aber darin, dass derartige Modelle aufgrund der geringen bzw. fehlenden Interpretierbarkeit von potenziellen Anwendern meist nicht akzeptiert werden. Dies ist auch verständlich, da industrielle Anwender die Modelle so nicht mit unter Umständen vorhandenem Systemwissen abgleichen können, gleichzeitig aber die Verantwortung für die korrekte Arbeitsweise der Modelle übernehmen müssen. Qualitative Modelle werden hingegen so ausgelegt, dass sie Aussagen über die Ursache-Wirkung-Zusammenhänge liefern und daher auch interpretierbar sind. Hierfür werden beispielsweise nachvollziehbare physikalische Gleichungen oder verständliche WENN-DANN-Regeln verwendet. Durch die Bestimmung freier Parameter wird ein solches Modell in ein quantitatives Modell überführt, das weiterhin gut zu interpretieren ist. Modelle dieser Art werden von potenziellen Anwendern in der Regel akzeptiert.

Zusammenfassend lässt sich festhalten, dass die parallele Verfolgung der Aspekte Modellinterpretierbarkeit und Modellgenauigkeit erstrebenswert ist. Ferner ist es von Vorteil, wenn auch bereits vorhandenes Wissen über das zu modellierende System in geeigneter Weise in den Prozess der Modellierung einfließen kann. Im Hinblick auf die genannten Zielsetzungen stoßen rigorose mathematische Modellierungsansätze bei immer komplexer werdenden Systemen an ihre Grenzen. Im Bereich der Computational Intelligence (CI) sind viel versprechende Ansätze zur Lösung auch solcher Modellierungsprobleme entwickelt worden. Unter dem Dachbegriff der CI werden eine Reihe verschiedener Methoden zusammengefasst, insbesondere sind dies evolutionäre Algorithmen (EA), künstliche neuronale Netze (NN) und Fuzzy-Logik (FL). Evolutionäre Algorithmen ahmen die aus der biologischen Evolution bekannten Schritte Reproduktion, Mutation und Selektion nach und werden oftmals zur Lösung hochdimensionaler Optimierungsprobleme eingesetzt. Sie ermöglichen eine effiziente Durchmusterung entsprechender Suchräume, stellen aber in diesem Sinne keinen eigenständigen Modellierungsansatz dar. Das Prinzip künstlicher neuronaler Netze ist ebenfalls der Natur entlehnt, da der Aufbau solcher Netze dem menschlichen Gehirn nachempfunden ist. Fuzzy-Logik bzw. Fuzzy-Modelle verwenden sehr anschauliche WENN-DANN-Regeln und orientieren sich an der Art und Weise, wie Menschen ihnen bewusstes Wissen formulieren. Künstliche neuronale Netze zeichnen sich durch eine vergleichsweise hohe erreichbare Modellgenauigkeit aus. Allerdings ist die Interpretierbarkeit solcher Modelle eher gering und es ist auch nur sehr bedingt möglich, bereits vorhandenes Wissen zu nutzen. Im Hinblick auf diese Aspekte haben Fuzzy-Modelle deutliche Vorteile.

Als Vorstufe von Fuzzy-Modellen können regelbasierte Modelle angesehen werden, die Regeln der Art WENN <Bedingung> DANN <Ausgabewert> verwenden. Der gravierende Nachteil liegt darin, dass eine Regel entweder aktiviert (die Bedingung der Regel ist vollständig erfüllt) oder aber deaktiviert (die Bedingung der Regel ist überhaupt nicht erfüllt) wird. Hieraus resultiert eine Diskretisierung des Modellausgabewertes, was wiederum zu einer entsprechend geringen Modellgenauigkeit führen kann. Das beschriebene Verhalten ist das aus der booleschen Logik bekannte Prinzip, dass eine Aussage entweder wahr (1) oder falsch (0) ist. Diese Sichtweise entspricht allerdings kaum der Art und Weise, in der Menschen Aussagen bewerten. Menschen äußern sich meist eher so, dass für sie eine Aussage *mehr oder weniger* zutrifft. Dieses Prinzip wird durch die von ZADEH eingeführte Fuzzy-Logik aufgegriffen. Eine Aussage trifft dabei in einem Grade $\mu \in [0, 1]$ zu. Dieses Prinzip wird in Fuzzy-Modellen so genutzt, dass jede Regel in einem bestimmten Maße aktiviert wird und so fließende Übergänge zwischen den Ausgabewerten der einzelnen Regeln ermöglicht werden. Die bei regelbasierten Modellen auftretende Diskretisierung des Modellausgabewertes wird dadurch aufgebrochen, was die Modellgenauigkeit meist erheblich verbessert wird. Fuzzy-Modelle zeichnen sich somit durch eine gute Interpretierbarkeit und eine hohe Genauigkeit aus. Ein weiterer Vorteil besteht darin, dass Modelle dieser Art wartbar und einfach erweiterbar sind, da weitere Regeln einfach zu der vorhandenen Regelbasis hinzugefügt werden können. Auf diese Weise kann auch bereits vorhandenes Wissen über das System relativ einfach in das Modell integriert werden.

In der Anfangszeit der Fuzzy-Modellierung wurden die Modelle wissensbasiert aufgestellt. Experten mit entsprechender Kenntnis des zu modellierenden Systems formulieren die Regeln dabei *von Hand*. Diese Vorgehensweise ist jedoch recht aufwendig. Da das Wissen von Experten unter Umständen auch nur unvollständig ist bzw. verschiedene Experten verschiedene Ansichten vertreten, sind dieser Vorgehensweise Grenzen gesetzt. Dieser Ansatz scheidet auch dann aus, wenn kein oder nur sehr wenig Wissen über das System vorliegt. Aus diesen Gründen kommt der datenbasierten Fuzzy-Modellierung eine große Bedeutung zu. Grundlage für die Erstellung eines Modells sind dabei in erster Linie Messdaten, die an dem System erfasst werden. Die Vorteile dieses Ansatzes bestehen zum einen darin, dass im Prinzip kein Wissen über das System notwendig ist, zum anderen kann die Modellierung häufig in vergleichsweise kurzer Zeit durchgeführt werden, da dieser Prozess meist gut automatisiert werden kann.

Die Arbeit beschäftigt sich mit der datenbasierten Generierung von Fuzzy-Modellen. Hierfür wird das von KIENDL, KRABS und KRONE in Vorarbeiten entwickelte Fuzzy-ROSA-Verfahren¹ eingesetzt. Ziel der Arbeit ist es, das Verfahren in seiner Leistungsfähigkeit im Hinblick auf die Anforderungen in industriellen Problemstellungen zu erweitern und Anwenden systematische Strategien zur Lösung komplexer Modellierungsaufgaben bereitzustellen, damit diese das Verfahren erfolgreich einsetzen können, ohne sich mit Details hierzu befassen zu müssen.

¹Regel-Orientierte Statistische Analyse

Die Arbeit ist wie folgt gegliedert: In Kapitel 2 wird die der Arbeit zu Grunde liegende Problemstellung erläutert. Die daraus resultierenden Ziele werden herausgearbeitet und es wird eine Einordnung in bestehende Arbeiten vorgenommen. Kapitel 3 gibt eine Einführung in die Fuzzy-Logik und in die beiden gängigen Arten von Fuzzy-Modellen (Mamdani-Modelle und Takagi-Sugeno-Kang-Modelle). Das in dieser Arbeit eingesetzte Fuzzy-ROSA-Verfahren wird in Kapitel 4 detaillierter beschrieben. Kapitel 5 befasst sich mit der Analyse des Verfahrens, wie es zu Beginn dieser Arbeit vorlag. Aufbauend auf der Analyse werden neue Strategieelemente, die Verbesserungspotenzial versprechen, vorgeschlagen. In den Kapiteln 6 und 7 werden neue Strategien sowohl für die Regelgenerierung als auch die Regeltransformation vorgestellt. Für die Generierung relevanter Mamdani-Regeln im Fall einer nur kleinen Datenbasis wird ein neues Test- und Bewertungsverfahren, basierend auf dem Median, entwickelt, das in Kapitel 6 beschrieben wird. Zur Erhöhung der Modellgenauigkeit kann ein mit dem Fuzzy-ROSA-Verfahren generiertes Mamdani-Modell nachgelagert in ein korrespondierendes Takagi-Sugeno-Kang-Modell transformiert werden, wobei die Interpretierbarkeit der Regeln erhalten bleibt. In Kapitel 7 wird für diese Transformation eine Strategie vorgestellt, die eine bessere Robustheit gegenüber fragwürdigen Daten aufweist als die bisher eingesetzte Strategie. In Kapitel 8 wird ein Ansatz für ein hybrides Fuzzy-Modell vorgestellt, das die Nutzung der von KIENDL vorgeschlagenen negativen Mamdani-Regeln in Verbindung mit Takagi-Sugeno-Kang-Modellen ermöglicht. Kapitel 9 befasst sich mit der systematischen Exploration der Einstellmöglichkeiten des Fuzzy-ROSA-Verfahrens und der daraus abgeleiteten Formulierung systematischer Lösungsstrategien. Die Leistungsfähigkeit der neu entwickelten Strategieelemente wird in diesem Zusammenhang aufgezeigt. Eine Zusammenfassung der Arbeit sowie ein Ausblick werden in Kapitel 10 gegeben.

Kapitel 2

Einordnung und Zielsetzung

In diesem Kapitel wird zunächst der Begriff des Modells erläutert und die Bedeutung von Modellen herausgearbeitet. Die Unterschiede zwischen einer wissensbasierten und einer datenbasierten Modellierung werden in Abschnitt 2.2 behandelt. Abschnitt 2.3 gibt eine Übersicht über etablierte Modellansätze im Bereich der datenbasierten Modellierung. Die Ziele dieser Arbeit werden in Abschnitt 2.4 beschrieben.

2.1 Modelle

Wie einführend erläutert, sind Modelle sowohl in den Wirtschafts- und Sozialwissenschaften als auch in den Natur- und Ingenieurwissenschaften von großer Bedeutung. Je nach Anwendungsfeld werden bestimmte Eigenschaften eines Modells gewünscht bzw. präferiert. Daher sind unterschiedlichste Modellarten entwickelt worden, um den verschiedensten Ansprüchen möglichst gerecht zu werden. Allen Modellarten ist dabei aber die Eigenschaft gemeinsam, dass sie ein System abbilden. Ein System wird häufig definiert als *eine Menge von Elementen, die miteinander durch Beziehungen verbunden sind und gemeinsam einen bestimmten Zweck zu erfüllen haben*. Ein System besteht aus Systemelementen, Beziehungen (Relationen) zwischen den Elementen und einer Systemgrenze, die das System von der Umwelt klar abgrenzt. Der Begriff *System* ist damit im Hinblick auf die sehr unterschiedlichen Anwendungsfelder sehr weit gefasst. Unter einem System können beispielsweise vernetzte Wirtschaftsabläufe verstanden werden oder aber auch industrielle Produktionsanlagen. Ein System muss dabei nicht zwangsläufig real existieren. Beispielsweise werden in der Automobilindustrie aus Zeit- und Kostengründen zunehmend potenzielle Prototypen zunächst nur virtuell entworfen und durch entsprechende Modelle erprobt und optimiert, bevor sie dann tatsächlich physikalisch realisiert und getestet werden. Die Abbildung eines Systems durch ein Modell geschieht typischerweise vereinfachend. Die Gründe hierfür liegen darin, dass zum einen in der Regel nie der ganze Umfang des zu modellierenden Systems

erfasst werden kann, und zum anderen, dass ein Modell in seiner gewählten Struktur auf einen bestimmten Umfang bzw. eine gewisse Komplexität begrenzt ist. Die Vereinfachung darf dabei nicht zu stark ausfallen, da sonst die Modellierung wertlos ist. Die Kompromissbildung bzw. Abwägung zwischen der vorgenommenen Vereinfachung und dem dadurch verursachten Verlust an Genauigkeit der Nachbildung ist eine überaus schwierige Aufgabe.

In den Wirtschafts- und Sozialwissenschaften steht häufig die transparente Beschreibung von Wirkungszusammenhängen des Systems durch das Modell im Vordergrund. In den Natur- und Ingenieurwissenschaften hingegen sind die Genauigkeit und die Anwendbarkeit des Modells von Interesse. Ein Modell wird in diesen Bereichen typischerweise zur Simulation und Optimierung eines Systems eingesetzt. Grundsätzlich sind beide Modelleigenschaften – Transparenz und Genauigkeit – erstrebenswert, oftmals muss aber ein Aspekt zugunsten des anderen vernachlässigt werden. Quantitative Modelle lassen sich im Hinblick auf den Grad ihrer Transparenz in verschiedene Kategorien einordnen, wobei die Grenzen zwischen diesen Kategorien fließend sind. So genannte White-Box-Modelle werden auch als explikative Modelle bezeichnet und zeichnen sich dadurch aus, dass ihre Wirkungsmechanismen völlig transparent sind. Auf der anderen Seite existieren so genannte Black-Box-Modelle, die auch als deskriptive Modelle bezeichnet werden. Bei solchen Modellen sind die Wirkungsmechanismen völlig intransparent. Da sich ein Modell in der Regel zwischen diesen Extremen einordnet, hat sich hierfür der Begriff der Grey-Box-Modelle etabliert.

2.2 Modellierung

Um eine mathematische Modellierung eines Systems vornehmen zu können, wird dieses auf sein Eingangs-/Ausgangsverhalten abstrahiert. Das Eingangs-/Ausgangsverhalten des Systems wird durch den Satz $\mathcal{X} = \{X_1, \dots, X_n\}$ von Eingangsgrößen und den Satz $\mathcal{Y} = \{Y_1, \dots, Y_m\}$ von Ausgangsgrößen beschrieben. Bezüglich der Art der gemessenen Größen sind drei verschiedene Typen zu unterscheiden. Nominale Größen nehmen ausschließlich symbolische Werte an, die nicht in einer Rangfolge bezüglich einer Kleiner-Größer-Relation angeordnet werden können (Beispiel: Farbe eines Objektes). Ordinale Größen nehmen ebenfalls ausschließlich symbolische Werte an, die sich jedoch anhand einer Kleiner-Größer-Relation ordnen lassen (Beispiel: Schulnoten). Kontinuierliche Größen können beliebige reelle Werte annehmen (Beispiel: Temperaturwerte in Grad Celsius). Je nach Anzahl der Eingangs- und Ausgangsgrößen eines Systems unterscheidet man zwischen vier Klassen: SISO-, MISO-, SIMO- und MIMO-Systeme¹. Eine an dem System protokollierte Eingangs-/Ausgangssituation wird durch das Datentupel $\mathbf{d}_i = (\mathbf{x}_i, \mathbf{y}_i)$ vollständig beschrieben. Dabei sind $\mathbf{x}_i = (x_{i,1}, \dots, x_{i,n})$ und $\mathbf{y}_i = (y_{i,1}, \dots, y_{i,m})$ die reellen Werte der Eingangs- und Ausgangsgrößen. Ein Modell des Systems verwendet den gleichen Satz \mathcal{X} von Eingangsgrößen (oder aber auch nur eine Teilmenge hiervon) und liefert entsprechend einen Satz

¹Single Input Single Output, Multi Input Single Output, Single Input Multi Output und Multi Input Multi Output.

$\hat{\mathcal{Y}} = \{\hat{Y}_1, \dots, \hat{Y}_m\}$ von modellierten Ausgangsgrößen. Die Aufgabe einer quantitativen mathematischen Modellierung ist es, eine geeignete Abbildung zu realisieren, sodass für einen gegebenen Vektor $\mathbf{x}_i = (x_{i,1}, \dots, x_{i,n})$ von Eingangsgrößenwerten der von dem Modell gelieferte Vektor mit modellierten Ausgangsgrößenwerten $\hat{\mathbf{y}}_i = (\hat{y}_{i,1}, \dots, \hat{y}_{i,m})$ möglichst gut mit dem Vektor der protokollierten Ausgangsgrößenwerte $\mathbf{y}_i = (y_{i,1}, \dots, y_{i,m})$ übereinstimmt.

Ein mathematisches Modell eines Systems kann entweder durch eine theoretische Analyse (nachfolgend als *wissensbasierte* Modellierung bezeichnet) oder aber durch eine experimentelle Analyse (nachfolgend als *datenbasierte* Modellierung bezeichnet) erstellt werden. Eine wissensbasierte Modellierung beruht auf der Formulierung der Wirkungszusammenhänge des zu modellierenden Systems. Dies geschieht oftmals in der Art, dass auf der Basis physikalischer Gesetzmäßigkeiten Modellgleichungen formuliert werden, beispielsweise in Form von Differentialgleichungen. Darin enthaltene freie Parameter des Modells (zum Beispiel die Koeffizienten der einzelnen Differentialgleichungen) werden dann bestimmt und gegebenenfalls geeignet angepasst, sodass das Modellverhalten möglichst gut dem realen Systemverhalten entspricht. Eine solche wissensbasierte Modellierung hat den Vorteil, dass das System selbst für die Modellierung nicht benötigt wird, es müssen also beispielsweise keine Messdaten erhoben werden. Ein weiterer Vorteil dieser Vorgehensweise liegt darin, dass tiefgehende Erkenntnisse über die innere Struktur des Systems gewonnen werden. Ein solches Modell ist daher überaus transparent, da die Wirkungsweise des Modells absolut einsichtig ist. Nachteilig ist jedoch, dass diese Vorgehensweise ein tiefes Verständnis des zu modellierenden Systems, also entsprechendes Expertenwissen, voraussetzt. Solches Expertenwissen ist, insbesondere im Hinblick auf immer komplexer werdende Systeme und deren zunehmende Vernetzung, oftmals nicht in dem Maße vorhanden, wie es für eine erfolgreiche wissensbasierte Modellierung benötigt wird. Selbst in dem Fall, dass entsprechendes Wissen vorhanden ist, ist die Modellierung dann aber trotzdem noch zeit- und damit kostenintensiv. Dieser Aspekt ist vor dem Hintergrund eines zunehmenden Wettbewerbs in allen Bereichen der Industrie von entscheidender Bedeutung. Es ist für industrielle Anwender zwingend notwendig, anspruchsvolle Projekte kosteneffizient durchzuführen. Die entstehenden Kosten sind dabei oftmals in erster Linie von der Zeitspanne abhängig, die für die Bearbeitung eines Projektes benötigt wird.

Vor diesem Hintergrund und der sich abzeichnenden zukünftigen Entwicklung gewinnt die datenbasierte Modellierung immer mehr an Bedeutung. Vorteilhaft dabei ist, dass kein bzw. nur sehr wenig Expertenwissen über das zu modellierende System vorausgesetzt wird. Dieser Ansatz verspricht daher insbesondere auch in solchen Fällen Vorteile, bei denen dem zu modellierenden System keine physikalischen Gesetzmäßigkeiten zu Grunde liegen, sondern vielmehr soziologische und psychologische Mechanismen. Solche vergleichsweise neuen Anwendungsfelder, wie beispielsweise die Bewertung der Kreditwürdigkeit anhand kundenspezifischer Merkmale oder die Erforschung des Konsumverhaltens, lassen sich mit klassischen Modellierungsansätzen oft nur schwer erschließen. In solchen Fällen werden verstärkt Methoden und Modelle aus dem Bereich der Computational Intelligence (CI) eingesetzt, auf die in Abschnitt 2.3 näher eingegangen wird. Die Suche nach Zusammen-

hängen in Daten, deren Aufdeckung und Formulierung wird häufig auch allgemein unter dem Oberbegriff *Data Mining* zusammengefasst.

Es haben sich verschiedene Modellansätze für eine datenbasierte Modellierung etabliert, die in Abschnitt 2.3 kurz vorgestellt werden. Bestimmte Arbeitsschritte sind allen Ansätzen gemeinsam und werden daher nachfolgend erläutert. Eine datenbasierte Modellierung gliedert sich in folgende Teilschritte: Datenerfassung, Datenauswahl, Datenaufbereitung, Merkmalsbildung, Merkmalsauswahl und die eigentliche Modellbildung. Die ersten drei Teilschritte können unter dem übergeordneten Begriff *Datenaquisition* zusammengefasst werden. Für die Modellierung müssen Daten an dem zu modellierenden System erfasst werden, die oft auch als Rohdaten bezeichnet. Da hierbei Fehler oder Störungen auftreten können, ist es oftmals notwendig, eine Auswahl der Daten vorzunehmen (zum Beispiel bestimmte Messdaten zu löschen) und diese in gewisser Weise aufzubereiten (zum Beispiel bestimmte Messdaten zu korrigieren). Die so vorverarbeiteten Daten werden oft auch als Nutzdaten bezeichnet. Es wird nachfolgend angenommen, dass diese Daten das Systemverhalten hinreichend gut beschreiben und alle relevanten Eingangsgrößen erfasst sind. Wenn dies nicht der Fall ist, also die Daten wenig aussagekräftig sind, sind die Chancen für eine erfolgreiche datenbasierte Modellierung unabhängig von dem gewählten Ansatz von vornherein entsprechend gering. Für die nachfolgende Modellierung kann es von Vorteil sein, diese Daten noch weiter zu bearbeiten. Beispielsweise können aus den protokollierten Größen neue Größen abgeleitet oder aggregierte Größen gebildet werden. Insbesondere kann es bei der Modellierung von Prozessen, die einen zeitlichen Verlauf aufweisen, beispielsweise bei der Lastprognose in Stromnetzen, vorteilhaft sein, nicht nur den aktuellen Wert einer Größe als Eingangsgröße des Modells zu verwenden, sondern zusätzlich auch zeitlich zurückliegende Werte. Bei einem üblichen Fuzzy-System ist der aktuelle Ausgangsgrößenwert eine eindeutige Funktion der aktuellen Eingangsgrößenwerte. Das Fuzzy-System arbeitet also ohne Erinnerung. Durch die oben beschriebene Maßnahme kann so ein Erinnerungsvermögen für das Fuzzy-System geschaffen werden [Kie97a]. Im Prinzip ergeben sich an dieser Stelle beliebige Freiheiten, wobei die physikalische Anschaulichkeit abgeleiteter Größen unter Umständen eher gering ist. Um die Modellbildung in annehmbarer Zeit durchführen zu können, muss gegebenenfalls die Anzahl der Merkmale, also die Anzahl der Eingangsgrößen des späteren Modells, durch eine wie auch immer geartete Selektion auf eine handhabbare Menge reduziert werden. Die beiden Teilschritte Merkmalsbildung und Merkmalsauswahl sind für den Erfolg der nachfolgenden Modellbildung entscheidend, sie sind aber nicht wesentlicher Bestandteil dieser Arbeit. Wenn die oben genannten Schritte abgeschlossen sind, erfolgt die eigentliche Modellbildung. Es muss zunächst eine Modellstruktur gewählt werden, beispielsweise ein Fuzzy-Modell oder ein neuronales Netz. Ist dies geschehen, müssen die freien Modellparameter auf Grundlage der Daten identifiziert werden. Damit ist die Modellbildung im Prinzip abgeschlossen. In der Praxis setzt darauf aufbauend jedoch ein iterativer Prozess ein, bei dem zunächst die Qualität des so erstellten Modells bewertet wird (zum Beispiel anhand eines Fehlermaßes) und dann gegebenenfalls Modifikationen der Struktur (zum Beispiel Einfügen weiterer Neuronen in einem neuronalen Netz) vorgenommen werden. Dieser Vorgang wiederholt sich so lange, bis die Modellierung ein zufrie-

denstellendes Ergebnis liefert. Um abschließend zu prüfen, ob das Modell tatsächlich die Wirkungszusammenhänge des Systems nachbildet und nicht nur die Messdaten reproduziert, erfolgt meist ein Test des Modells auf bis dahin unbekanntem Daten, die nicht für die Erstellung des Modells verwendet worden sind. Wenn auch diese Daten zufriedenstellend modelliert werden, kann das Modell eingesetzt werden.

2.3 Übersicht etablierter Modellansätze

In dem Bereich der datenbasierten Modellierung haben sich verschiedene Modellansätze etabliert, die im Folgenden skizziert werden. Zwischen den verschiedenen Ansätzen existieren teilweise starke Verbindungen, die die Überführung eines Modelltyps in einen anderen erlauben. Zunehmend gewinnen auch hybride Modelltypen, also Mischformen aus verschiedenen Ansätzen, an Bedeutung.

Entscheidungsbäume:

Entscheidungsbäume werden typischerweise für Klassifikationsaufgaben eingesetzt. Eine Übertragung des Baumprinzips auf den Fall der Approximation ist jedoch leicht möglich. Der Unterschied liegt lediglich in der Art der Ausgabe der Blätter des Baumes. Die Induktion von Bäumen erfolgt oftmals mit Hilfe des von QUINLAN entwickelten ID3-Algorithmus [Qui86, Qui93]. Weitergehende Ansätze nutzen statt harter Entscheidungsgrenzen Fuzzy-Mengen für die Verzweigungen innerhalb eines Baumes (*Fuzzy Decision Trees*) [YS95, Jan98], wodurch eine Verbesserung der Robustheit angestrebt wird. Eine Baumstruktur lässt sich in einen Satz von WENN-DANN-Regeln überführen, sodass eine enge Verbindung zu Fuzzy-Modellen besteht und diese Verfahren daher auch im Bereich der datenbasierten Fuzzy-Modellierung angewandt werden [JMMB00].

Clusterverfahren:

Clusterverfahren [And73, Bez81, XW05] werden oftmals in Form unüberwachter Lernverfahren eingesetzt, um größere Mengen von Datenpunkten anhand ihrer Lage im Raum der Eingangsgrößen sinnvoll zu gruppieren. Diese Gruppierung dient häufig dazu, im Weiteren statt der ursprünglichen Daten die Prototypen der Cluster zu verwenden, da dann mit einer wesentlich kleineren Anzahl an Daten gearbeitet werden kann. Clusterverfahren können aber auch in Form überwachter Lernverfahren realisiert werden und dann als Modell dienen. Ein solches Modell kann ebenfalls als ein Satz von Regeln interpretiert werden, wobei die Regeln dann im Prinzip folgende Form haben:

WENN <Eingangsgrößenvektor innerhalb des Clusters> DANN <Konklusion> .

Beispiel hierzu finden sich in [GSJ97, Hae03]. Nachteilig ist hierbei, dass für jede Regel jeweils neue linguistische Werte festgelegt werden müssen, die abhängig sind

von der Ausdehnung des zugehörigen Clusters. Je nach Anzahl der Cluster kann die Anzahl der erzeugten linguistischen Werte daher schnell ansteigen und da sich die linguistischen Werte zudem beliebig überlappen können, ist die Transparenz und Einsichtigkeit der so erzeugten Regelbasis eingeschränkt.

Neuronale Netze:

Neuronale Netze [Kin94, Roj96] sind in ihrer Struktur dem Aufbau des menschlichen Gehirns angelehnt. Der große Vorteil neuronaler Netze liegt in ihrer Lernfähigkeit. Es existieren verschiedene Lernverfahren für die Anpassung eines neuronalen Netzes an die gegebenen Lerndaten. Häufig wird das so genannte Backpropagation-Verfahren eingesetzt, das im Wesentlichen darauf beruht, den Fehler bei der Modellierung der Lerndaten auf die einzelnen gewichteten Verbindung zurückzurechnen und die Gewichte so zu ändern, dass der Fehler kleiner wird. Dieser Vorgang wird wiederholt und der Fehler so iterativ minimiert. Bei Vorgabe einer speziellen Netzstruktur lassen sich neuronale Netze in Fuzzy-Modelle überführen. Neuronale Netze können auch für eine Optimierung von Fuzzy-Modellen eingesetzt werden. Neuro-Fuzzy-Systeme stellen einen hybriden Ansatz dar, der versucht, die Vorzüge beider Ansätze miteinander zu verbinden.

Support Vector Machine:

Die Support Vector Machine (SVM) geht auf VAPNIK zurück [Vap95]. Im Fall einer Klassifikationsaufgabe versucht die SVM, die gegebenen Lerndaten durch eine trennende Hyperebene linear in zwei Klassen zu separieren. Falls dies möglich ist, wird aus der Menge der möglichen Hyperebenen diejenige gewählt, die die Trennung mit der größtmöglichen Robustheit darstellt. Dies ist die Hyperebene, die zu den Datenpunkten, die der Hyperebene am nächsten liegen (den Support-Vektoren), den größtmöglichen Abstand hat. Lassen sich die Datenpunkte nicht linear separieren, so werden die Daten aus dem ursprünglichen Raum (Input Space) durch eine nichtlineare Funktion in einen Raum höherer Dimension (Feature Space) transformiert, in dem sie sich linear separieren lassen. Es existieren verschiedene Strategien, um mit der SVM auch Mehrklassenprobleme zu bearbeiten. Für die Bearbeitung von Approximationsproblemen wurde der Ansatz der Support Vector Regression entwickelt.

Fuzzy-Modelle:

Fuzzy-Modelle arbeiten mit WENN-DANN-Regeln und sind daher am ehesten mit der Art und Weise vergleichbar, wie Menschen ihnen bewusstes Wissen ausdrücken. In den Anfängen der Fuzzy-Modellierung sind Regelbasen überwiegend wissensbasiert erstellt worden. Diese Vorgehensweise setzt ein entsprechendes Expertenwissen des zu modellierenden Systems voraus und zielt in erster Linie darauf ab, das bereits vorhandene Wissen in einer (gegebenenfalls vereinfachten) Form für die maschinelle Verarbeitung geeignet zu repräsentieren. Diese Vorgehensweise hat zwei zentrale Nachteile: Zum einen stößt sie bei komplexen Systemen im Hinblick auf den Aufwand zwangsläufig schnell an Grenzen. Zum anderen kann so vorhandenes Wissen – was nicht zwangsläufig das System auch tatsächlich zutreffend beschreiben muss – zwar

reproduziert werden, aber es können unbekannte Zusammenhänge – die das System tatsächlich zutreffend beschreiben – so nicht aufgedeckt werden. Daher sind frühzeitig Bemühungen unternommen worden, Verfahren zu entwickeln, die eine zutreffende Regelbasis anhand vorhandener Messdaten generieren. In diese Kategorie von Verfahren fällt auch das dieser Arbeit zu Grunde liegende Fuzzy-ROSA-Verfahren.

2.4 Ziele dieser Arbeit

Diese Arbeit verfolgt die nachfolgend beschriebenen Ziele.

Erweiterung des Fuzzy-ROSA-Verfahrens für industrielle Anwendungen:

Um die Anwendbarkeit für industrielle Problemstellungen und die damit letztlich verbundene Akzeptanz des Verfahrens bei potenziellen Anwendern zu erhöhen, werden in dieser Arbeit verschiedene Erweiterungen des Fuzzy-ROSA-Verfahrens vorgenommen. Dies betrifft verschiedenste Aspekte der datenbasierten Fuzzy-Modellierung. Hierzu zählen: Selektion geeigneter Eingangsgrößen für die nachfolgende Modellierung, Entwicklung eines Test- und Bewertungsverfahrens zur Regelgenerierung für den Fall einer kleinen und/oder gestörten Datenbasis sowie Erhöhung der Genauigkeit bzw. Verbesserung der Robustheit der Transformation von Mamdani-Regeln in TSK-Regeln.

Ausdehnung der Einsatzmöglichkeiten negativer Regeln:

Ein herausstellendes Merkmal des Fuzzy-ROSA-Verfahrens gegenüber anderen Modellierungsansätzen ist die Unterstützung negativer Mamdani-Regeln. Durch negative Regeln wird die prinzipielle Leistungsfähigkeit von Mamdani-Systemen zwar nicht erweitert, da gezeigt wurde, dass sich mit diesen prinzipiell universelle Approximatoren realisieren lassen. Negative Regeln ermöglichen es aber, Zusammenhänge bzw. Wissen, das in Form von Warnungen oder Verboten vorliegt, sehr anschaulich und kompakt zu formulieren. Neben Mamdani-Systemen sind TSK-Systeme wegen ihrer meist guten Modellgenauigkeit weit verbreitet. Eine Möglichkeit zur direkten Übertragung des Prinzips negativer Regeln auf Systeme solchen Typs ist bisher nicht bekannt. Daher wird in dieser Arbeit ein Ansatz für ein hybrides Fuzzy-System entwickelt, der eine nutzbringende Anwendung negativer Mamdani-Regeln in Verbindung mit TSK-Systemen erlaubt.

Exploration der Einstellmöglichkeiten und Entwicklung systematischer Lösungsstrategien:

Zusammenfassend zielt die datenbasierte Modellierung darauf ab, auch solche Systeme und Prozesse modellieren zu können, für die keine oder nur eine sehr geringe Kenntnis der Wirkungszusammenhänge vorhanden ist. Eine Modellierung soll also ohne nennenswertes Expertenwissen möglich sein. Erfordert nun aber das eingesetzte

Modellierungsverfahren seinerseits wieder solches Expertenwissen, so wird der Aufwand für einen Anwender insgesamt nicht reduziert, sondern nur an eine andere Stelle verlagert. Es ist daher zwingend erforderlich, dass Anwendern die komplexen Methoden des Fuzzy-ROSA-Verfahrens auch ohne eine tief gehende Einarbeitung in Details zugänglich sind. Um dies zu gewährleisten, werden zwei Ziele formuliert. Das erste Ziel besteht darin, dem Anwender eine Unterstützung bei der Modellerstellung anzubieten. Dieses Support-System gibt abhängig von dem zu modellierenden Problem bestimmte Empfehlungen für zum Beispiel die Wahl von Modellparametern, die endgültige Entscheidung trifft aber der Anwender. Das weiterführende Ziel besteht darin, diesen Vorgang vollständig zu automatisieren, also anhand charakteristischer Problemkenngrößen eine möglichst geeignete Einstellung der freien Modellparameter vorzunehmen und ein entsprechendes Modell zu generieren. Eine solche Vorgehensweise ist zum einen für Anwender interessant, die sich mit dem Modellierungsverfahren möglichst wenig befassen möchten. Es ist aber insbesondere dann von Vorteil, wenn die Modellgenerierung in einen übergeordneten Prozess eingebunden werden muss. Da dieser übergeordnete Prozess nicht für längere Zeit unterbrochen werden darf, ist es dann nicht möglich, ein Modell von Hand zu erstellen. Um die genannten Ziele zu erreichen, werden die Einstellmöglichkeiten des Fuzzy-ROSA-Verfahrens an unterschiedlichsten Problemen systematisch untersucht. Mit Hilfe der so gewonnenen Daten werden das Support-System und die automatisierte Modellgenerierung aufgebaut.

Kapitel 3

Grundlagen der Fuzzy-Modellierung

In diesem Kapitel wird auf Grundlagen eingegangen, die für diese Arbeit sowohl wegen der verwendeten Begriffe als auch inhaltlich von zentraler Bedeutung sind. Es wird in Abschnitt 3.1 zunächst eine Einführung in die von ZADEH eingeführte Fuzzy-Logik gegeben [Zad65, Zad68]. Die Fuzzy-Logik bildet die Grundlage für den Entwurf und Einsatz von Fuzzy-Reglern und Fuzzy-Modellen. Unabhängig von ihrem späteren Verwendungszweck können Fuzzy-Regler und Fuzzy-Modelle allgemein auch als Fuzzy-Systeme bezeichnet werden. In den Abschnitten 3.2 und 3.3 werden zwei gängige Typen von Fuzzy-Systemen vorgestellt.

3.1 Fuzzy-Logik

In der booleschen Logik können nur die Wahrheitswerte 1 (wahr) oder 0 (falsch) angenommen werden. Im Gegensatz dazu erlaubt die Fuzzy-Logik auch Zwischenwerte aus dem Intervall $[0, 1]$. Die Fuzzy-Logik basiert auf der Verwendung von Fuzzy-Mengen. Eine Fuzzy-Menge \mathcal{S} besteht aus geordneten Zahlenpaaren:

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

Die Werte x stammen aus dem Universum \mathcal{U} , das im Rahmen dieser Arbeit mit der Menge \mathbb{R} der reellen Zahlen gleichgesetzt werden kann. Die Zugehörigkeitsfunktion $\mu_{\mathcal{S}}(x)$ bestimmt dabei den Grad der Zugehörigkeit eines Wertes x zu der Fuzzy-Menge \mathcal{S} . Eine Zugehörigkeitsfunktion $\mu_{\mathcal{S}}(x)$ kann durch eine Reihe von Kenngrößen charakterisiert werden. Der Support \mathcal{S}_0 ist der Teil des Definitionsbereiches einer Zugehörigkeitsfunktion, in der diese Werte größer als Null annimmt:

$$\mathcal{S}_0 = \{x \mid \mu_{\mathcal{S}}(x) > 0\} \quad . \quad (3.2)$$

Der Alpha-Schnitt \mathcal{S}_α ist der Teil des Supports einer Zugehörigkeitsfunktion, in der diese Werte größer α annimmt:

$$\mathcal{S}_\alpha = \{x \mid \mu_{\mathcal{S}}(x) > \alpha\} \quad . \quad (3.3)$$

Der Kern \mathcal{S}_1 ist der Teil des Supports einer Zugehörigkeitsfunktion, in der diese den Wert Eins annimmt:

$$\mathcal{S}_1 = \{x \mid \mu_S(x) = 1\} \quad . \quad (3.4)$$

Aus den oben genannten Kenngrößen resultieren zwei zentrale Eigenschaften für Fuzzy-Mengen. Eine Fuzzy-Menge, deren Zugehörigkeitsfunktion einen Maximalwert von Eins besitzt, wird normalisiert genannt. Eine Fuzzy-Menge, bei der alle Alpha-Schnitte zusammenhängen, wird konvex genannt.

Um mit Fuzzy-Mengen arbeiten zu können, sind die aus der booleschen Logik bekannten Operatoren zur logischen Verknüpfung auf Fuzzy-Mengen übertragen worden. Für die Realisierung eines Fuzzy-Operators ergeben sich unterschiedliche Möglichkeiten. Eine grundlegende Forderung ist jedoch oftmals, dass für den Grenzfall, dass beide Operanden einen Wert aus der Menge $\{0, 1\}$ annehmen, das Ergebnis ebenfalls einen Wert aus dieser Menge annimmt. Mit anderen Worten: Für den Grenzfall soll das Ergebnis dem der booleschen Logik entsprechen. Die wichtigsten Operatoren sind UND, ODER und NICHT. Für die Realisierung eines UND- bzw. ODER-Operators wird typischerweise eine t-Norm bzw. eine t-Conorm (oder auch s-Norm) verwendet. Eine t-Norm ist auf dem Einheitsintervall $[0, 1]$ definiert, $T : [0, 1] \times [0, 1] \longrightarrow [0, 1]$. Gleiches gilt für eine t-Conorm. Eine t-Norm muss folgende Eigenschaften aufweisen: Assoziativität, Kommutativität, Monotonie und 1 ist das neutrale Element. Geläufige t-Normen sind das Minimum, das algebraische Produkt sowie die beschränkte Differenz. Eine t-Conorm muss folgende Eigenschaften aufweisen: Assoziativität, Kommutativität, Monotonie und 0 ist das neutrale Element. Geläufige t-Conormen sind das Maximum, die algebraische Summe sowie die beschränkte Summe.

3.2 Mamdani-Fuzzy-Systeme

Ein Mamdani-System [MG81] beinhaltet die folgenden Teilmodule: Fuzzifizierung, Inferenz und Defuzzifizierung. Das Teilmodul Inferenz ist wiederum aus den Teilmodulen Aggregation, Aktivierung und Akkumulation zusammengesetzt. Auf die einzelnen Teilmodule wird im Folgenden genauer eingegangen.

Fuzzifizierung

Um unscharfe qualitative Aussagen zum Beispiel der Form *Temperatur ist hoch* verarbeiten zu können, muss dem numerischen Wert einer Größe (zum Beispiel Temperatur = 35° C) ein linguistischer Wert (zum Beispiel *hoch*) zugeordnet werden. Ein linguistischer Wert ist eine natursprachliche Bezeichnung, um die Eigenschaften einer Größe zu charakterisieren. Die gegenseitige Zuordnung zwischen numerischen und linguistischen Werten einer Größe erfolgt über Fuzzy-Mengen. Die Definition linguistischer Werte durch Festlegung von Zugehörigkeitsfunktionen wird auch als Partitionierung oder Granularisierung bezeichnet.

Inferenz

Unter Inferenz wird die Auswertung einer Regelbasis \mathcal{R} für einen gegebenen Eingangsgrößenvektor \mathbf{x} verstanden, also die Erzeugung einer ausgangsseitigen Zugehörigkeitsfunktion $\mu_Y(y)$ aus den fuzzifizierten Eingangsgrößen. Eine Mamdani-Regel hat die Form:

$$\text{WENN } \langle P \rangle \text{ DANN } \langle C \rangle \quad . \quad (3.5)$$

Der WENN-Teil einer Regel wird als Prämisse P bezeichnet, der DANN-Teil als Konklusion C . Die Prämisse ist eine Elementaraussage bzw. eine Verknüpfung mehrerer Elementaraussagen, bezogen auf die Eingangsgrößen. Dabei kann es sich grundsätzlich um eine rein konjunktive oder rein disjunktive Verknüpfung von Elementaraussagen handeln oder um eine Mischform aus beiden. Die Konklusion ist eine Elementaraussage, bezogen auf die Ausgangsgröße.

Eine Elementaraussage, bezogen auf die Ausgangsgröße Y , ist gegeben durch

$$C = (Y = \mathcal{S}_{Y,j}) \quad . \quad (3.6)$$

Entsprechend ist eine Elementaraussage, bezogen auf eine Eingangsgröße X_i , gegeben durch

$$E = (X_i = \mathcal{S}_{X_i,j}) \quad . \quad (3.7)$$

Bei einer rein konjunktiven Verknüpfungen von Elementaraussagen gilt somit für die Prämisse:

$$P = \bigwedge_{l=1}^c E_l \quad . \quad (3.8)$$

Die Anzahl der in der Prämisse vorkommenden Elementaraussagen wird als *Kombinationstiefe* c bezeichnet. Die Erfülltheit der Prämisse P ergibt sich in diesem Fall durch die Verknüpfung der Erfülltheitsgrade der einzelnen Elementaraussagen zu:

$$\mu_P(\mathbf{x}) = \bigwedge_{l=1}^c \mu_{E_l} \quad . \quad (3.9)$$

Die Bestimmung der Erfülltheit der Prämisse P wird als *Aggregation* bezeichnet. Aus der Erfülltheit der Prämisse wird in dem Schritt der *Aktivierung* (auch als *Komposition* bezeichnet) die Erfülltheit der Zugehörigkeitsfunktion der Konklusion bestimmt:

$$\mu_C(\mathbf{x}, y) = \mu_P(\mathbf{x}) \wedge \mu_{Y,j}(y) \quad . \quad (3.10)$$

Im Schritt der *Akkumulation* wird durch Überlagerung der einzelnen Empfehlungen der R Regeln die ausgangsseitige Zugehörigkeitsfunktion $\mu_Y(\mathbf{x}, y)$ gebildet:

$$\mu_Y(\mathbf{x}, y) = \bigvee_{i=1}^R \mu_{C_i}(\mathbf{x}, y) \quad . \quad (3.11)$$

Defuzzifizierung

Unter Defuzzifizierung wird die Bestimmung eines scharfen numerischen Ausgangsgrößenwertes $y_D = f(\mu_Y(y))$ aus der ausgangsseitigen Zugehörigkeitsfunktion durch ein geeignetes Funktional verstanden. Eine Reihe unterschiedlicher Defuzzifizierungsmethoden sind geläufig: Center Of Gravity (COG), Center Of Area (COA) und Mean Of Maximum (MOM). Je nach Anwendungsfall kommen weitere Methoden in Betracht: Smallest Of Maximum (SOM) und Largest Of Maximum (LOM). Für den Fall, dass ausgangsseitig Singletons als Zugehörigkeitsfunktionen verwendet werden, kommen folgende Defuzzifizierungsmethoden in Betracht: Center Of Singletons (COS) und Maximum Of Singletons (MOS).

Neben diesen weit verbreiteten Methoden existieren noch andere Ansätze für die Defuzzifizierung, beispielsweise die Drehmoment-Methode und das Inferenzfilter, die von KIENDL eingeführt wurden [Kie97a]. Der Drehmoment-Methode [Kie97b] liegt die Idee zu Grunde, dass bei gleichsinnigen Empfehlungen eher ein Verstärkungseffekt und keine Kompromissbildung eintreten sollte. Das folgende Beispiel aus Abbildung 3.1 demonstriert einen solchen Fall anhand ausgangsseitiger Singletons. Das Ergebnis der Defuzzifizierung bei Anwendung der COS-Methode ist jeweils als Kreis dargestellt, das der Drehmoment-Methode als Dreieck.

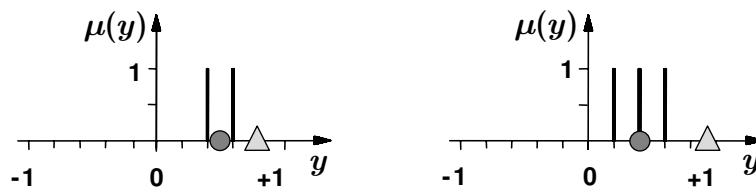


Abbildung 3.1: Beispiele für die Defuzzifizierung nach der Drehmoment-Methode in einem Mamdani-System.

Dem Inferenzfilter [Kie94b, Kie94a, Nie96] liegt die Überlegung zu Grunde, dass die ausgangsseitige Zugehörigkeitsfunktion $\mu_Y(y)$ nicht immer das *wahre* Empfehlungsgebirge widerspiegelt. So ergibt sich bei Anwendung der COG-Defuzzifizierung unter Umständen ein scharfer Ausgangsgrößenwert y_D , für den $\mu_Y(y_D) \neq \max(\mu_Y(y))$ gilt. Der so bestimmte Ausgangsgrößenwert entspricht also nicht zwangsläufig demjenigen Wert, dessen Empfehlung in $\mu_Y(y)$ am stärksten ist. Unter diesem Gesichtspunkt ist die MOM-Defuzzifizierung vorteilhaft. Aber auch diese kann in bestimmten Fällen zu Ergebnissen führen, die nicht mit den Entscheidungen eines Menschen übereinstimmen, beispielsweise in dem Fall, dass benachbarte Werte des am stärksten empfohlenen Ausgangsgrößenwertes kaum oder gar nicht empfohlen werden. Ein Mensch würde in diesem Fall aufgrund der Morphologie des Empfehlungsgebirges eher einen solchen Ausgangsgrößenwert bevorzugen, dessen benach-

barte Werte ebenfalls in einem signifikanten Maße empfohlen werden. Mit anderen Worten: $\mu_Y(y)$ stellt nicht die wahre Attraktivität eines Ausgangsgrößenwertes dar. Dies führt zu der Idee, aus $\mu_Y(y)$ durch ein geeignetes Filter eine ausgangsseitige Zugehörigkeitsfunktion $\tilde{\mu}_Y(y)$ zu erzeugen, die die wahre Attraktivität der Ausgangsgrößenwerte repräsentiert. In Abbildung 3.2 ist dies exemplarisch dargestellt. Die durchgezogene Linie stellt dabei $\mu_Y(y)$ und die gestrichelte Linie $\tilde{\mu}_Y(y)$ dar.

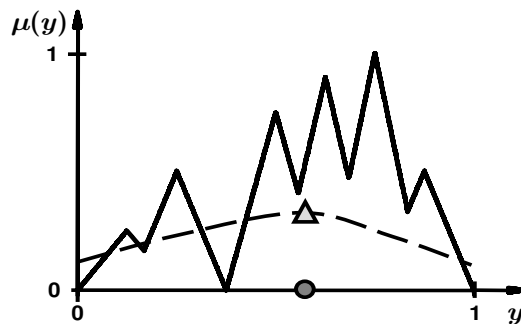


Abbildung 3.2: Beispiel für die Anwendung eines Inferenzfilters für die Defuzzifizierung in einem Mamdani-System.

Um den tatsächlich am stärksten empfohlenen Ausgangsgrößenwert zu erhalten, wird konsequenterweise die Maximum-Defuzzifizierung auf $\tilde{\mu}_Y(y)$ angewandt. Das Ergebnis bei Anwendung der COG-Defuzzifizierung ist in Abbildung 3.2 als Kreis dargestellt, das des Inferenzfilters als Dreieck. In diesem Fall wird das Inferenzfilter gerade so parametrisiert, dass das Ergebnis mit dem der COG-Defuzzifizierung übereinstimmt. Die wahre Attraktivität eines Ausgangsgrößenwertes berücksichtigt dabei, wie stark oder schwach benachbarte Werte eines Ausgangsgrößenwertes in $\mu_Y(y)$ empfohlen werden. Die Stärke der Berücksichtigung und der betrachtete Abstandshorizont können über Parameter des Inferenzfilters bzw. der darin eingesetzten Abstandsfunktion angepasst werden. Bei entsprechender Wahl der Parameter lassen sich so einerseits die bekannten Defuzzifizierungsmethoden realisieren (siehe das Beispiel aus Abbildung 3.2), andererseits aber auch zusätzlich fließende Übergänge zwischen diesen. Das Inferenzfilter kann in identischer Form auch auf ausgangsseitige Zugehörigkeitsfunktionen in Form von Singletons angewandt werden. Die hier skizzierte Realisierung eines Inferenzfilters ist eine von vielen möglichen. Sie basiert auf der Verwendung einer Abstandsfunktion, die über freie Parameter angepasst werden kann und daher bereits eine vergleichsweise hohe Flexibilität aufweist. Die Nachbildung der Art und Weise, wie ein menschlicher Experte eine Defuzzifizierung bzw. eine Entscheidungsfindung vornimmt, bleibt bei diesem Ansatz aber dennoch problematisch. Eine andere Realisierung eines Inferenzfilters, die eher den Ansatz einer musterbasierten Defuzzifizierung verfolgt, wird in [KK01, Kra02] vorgestellt.

3.3 Takagi-Sugeno-Kang-Fuzzy-Systeme

Der Aufbau der Prämisse einer Regel ist bei Mamdani-Systemen und bei TSK-Systemen [TS85] identisch. Der fundamentale Unterschied zwischen den beiden Systemtypen liegt in dem Aufbau der Konklusion einer Regel. Bei einer Mamdani-Regel ist die Konklusion eine Aussage bezüglich eines linguistischen Wertes. Im Gegensatz dazu ist bei einer TSK-Regel die Konklusion eine wie auch immer geartete Funktion der Eingangsgrößen:

$$\text{WENN } \langle P \rangle \text{ DANN } \langle f(\mathbf{x}) \rangle \quad . \quad (3.12)$$

Es ergeben sich zwangsläufig Unterschiede zwischen einem Mamdani-System und einem TSK-System: Ein TSK-System beinhaltet wohl die Teilmodule Fuzzifizierung und Aggregation, nicht aber die Teilmodule Aktivierung, Akkumulation und Defuzzifizierung. An deren Stelle tritt in einem TSK-System das Teilmodul Funktionsauswertung. Die Empfehlungen der einzelnen Regeln werden dabei typischerweise mit dem jeweiligen Erfülltheitsgrad der Prämisse gewichtet und der Ausgabewert des TSK-Systems wird beispielsweise durch eine gewichtete Mittelwertbildung zu

$$y_D(\mathbf{x}) = \frac{\sum_{i=1}^R \mu_{P_i}(\mathbf{x}) \cdot f_i(\mathbf{x})}{\sum_{i=1}^R \mu_{P_i}(\mathbf{x})} \quad (3.13)$$

berechnet. Es sei an dieser Stelle angemerkt, dass die Bestimmung des Ausgabewertes für ein TSK-System gemäß Gl. (3.13) im Prinzip der Defuzzifizierung nach der COS-Methode für ein Mamdani-System mit ausgangsseitigen Singletons entspricht.

Kapitel 4

Das Fuzzy-ROSA-Verfahren

Dieses Kapitel gibt einen zusammenfassenden Überblick über das in dieser Arbeit zur Modellierung eingesetzte Fuzzy-ROSA-Verfahren. Zentraler Bestandteil des Verfahrens ist die von KIENDL und KRABS vorgeschlagene **Regel-Orientierte Statistische Analyse** (ROSA) zur automatischen Generierung regelbasierter Systeme [KK89, Kra94, KK95]. Dieser Ansatz ist von KIENDL und KRONE auf den Fall der Fuzzy-Regeln übertragen worden (Fuzzy-ROSA) [KK94, Kro99, SKKK01]. Eine Beschreibung des Verfahrens sowie hiermit erzielte Ergebnisse für unterschiedliche Anwendungen finden sich in [KKSS02a]. Weitere erfolgreiche Anwendungen des Verfahrens werden in [TDT99, BE99, SJP⁺00, PJK⁺02, KKR06] vorgestellt.

Im Folgenden wird zunächst auf die grundlegende Idee des Verfahrens eingegangen. Daran anschließend werden einzelne Aspekte des Verfahrens genauer vorgestellt.

4.1 Grundidee des Verfahrens

Ziel des Verfahrens ist es, ein Modell zu erstellen, das, basierend auf Fuzzy-Regeln, das Verhalten des zu modellierenden Systems inhaltlich korrekt, möglichst exakt und möglichst kompakt nachbildet. Die Gesamtheit aller verwendeten Fuzzy-Regeln wird dabei als Regelbasis \mathcal{R} des Modells bezeichnet. Die zentrale Aufgabe besteht darin, eine im Sinne der verfolgten Ziele möglichst geeignete Regelbasis anhand der protokollierten Messdaten zu erstellen. Für diese Aufgabe kommen zwei fundamental unterschiedliche Ansätze in Frage, die jeweils gewisse Vor- und Nachteile aufweisen. Unter dem so genannten Pittsburgh-Ansatz [Smi80] wird verstanden, dass eine wie auch immer erzeugte Regelbasis gesucht wird, die insgesamt das Systemverhalten möglichst gut beschreibt bzw. nachbildet. Unter dem so genannten Michigan-Ansatz [Hol75, HHN86] wird hingegen verstanden, dass nacheinander einzelne Regeln geprüft werden und die Regelbasis so sukzessive aus einzelnen

plausiblen Regeln aufgebaut wird. Der Michigan-Ansatz reduziert so die Aufgabe der Suche nach einer geeigneten Regelbasis auf die Suche nach einzelnen geeigneten Regeln. Der Vorteil des Pittsburgh-Ansatzes liegt darin, dass die Interaktion der Regeln direkt bewertet wird. Nachteilig ist zum einen, dass hierbei Mengen von Regeln untersucht bzw. optimiert werden müssen und die resultierende kombinatorische Vielfalt einen entsprechend großen Suchraum induziert. Zum anderen kann die Plausibilität einzelner Regeln nicht gewährleistet werden. Insbesondere können mehrere unplausible Regeln im Zusammenwirken dennoch zu dem gewünschten Modellverhalten führen. Dieser mögliche Effekt erschwert das Erkennen solcher Regeln. Die Vorteile des Michigan-Ansatzes liegen darin, dass zum einen der Suchraum wesentlich kleiner ist als bei der Suche nach einer Regelbasis und zum anderen in der gesicherten Plausibilität der generierten Regeln. Der Nachteil dieses Ansatzes liegt darin, dass die spätere Interaktion der Regeln während der Generierungsphase nicht beachtet wird. Es ist daher nicht auszuschließen, dass die Regeln im Zusammenwirken nicht zu dem gewünschten Modellverhalten führen, obwohl sie individuell betrachtet alle vernünftig erscheinen.

Das Fuzzy-ROSA-Verfahren basiert auf dem oben beschriebenen Michigan-Ansatz. In den nachfolgenden Abschnitten werden die wesentlichen Schritte der Modellgenerierung, Modellanalyse und Modelloptimierung zusammenfassend vorgestellt.

4.2 Projektdefinition

In diesem Abschnitt wird beschrieben, in welchen Schritten die datenbasierte Generierung eines Mamdani-Modells mit dem Fuzzy-ROSA-Verfahren durchgeführt wird.

Mit dem Fuzzy-ROSA-Verfahren lassen sich Systeme mit einer oder mehreren Eingangsgröße(n) und einer Ausgangsgröße modellieren. Falls das zu modellierende System mehr als eine Ausgangsgröße aufweist, müssen entsprechend mehrere Fuzzy-Modelle erstellt werden, wobei sich jedes Modell dann auf eine der Ausgangsgrößen bezieht. Das zu modellierende System wird auf sein Eingangs-/Ausgangsverhalten abstrahiert, das gemäß der obigen Einschränkung durch den Satz $\mathcal{X} = \{X_1, \dots, X_n\}$ von Eingangsgrößen und $\mathcal{Y} = \{Y_1\}$ von Ausgangsgrößen beschrieben wird. Durch Beobachtung des zu modellierenden Systems wird ein Datensatz \mathcal{D} gewonnen, der aus N unabhängigen Datenpunkten $\mathbf{d}_i = (\mathbf{x}_i, \mathbf{y}_i)$ besteht, mit $\mathbf{x}_i = (x_{i,1}, \dots, x_{i,n})$, $\mathbf{y}_i = (y_{i,1})$ und $1 \leq i \leq N$.

Auswahl der zu beachtenden Eingangsgrößen

Das Verhalten des zu modellierenden Systems wird durch die n Eingangsgrößen und die Ausgangsgröße beschrieben. Für die Modellgenerierung können alle Eingangsgrößen verwendet werden oder aber nur eine Teilmenge $\tilde{\mathcal{X}} \subseteq \mathcal{X}$ hiervon. Die Auswahl einer geeigneten

Teilmenge kann aus unterschiedlichen Gründen von Vorteil sein. Zum einen können in der Menge \mathcal{X} redundante bzw. irrelevante Größen enthalten sein, die die Modellgenerierung eher erschweren. Zum anderen ist zu bedenken, dass mit zunehmender Anzahl an Eingangsgrößen die Dimension des Suchraumes für relevante Regeln anwächst, wodurch die benötigte Rechenzeit für die Modellgenerierung steigt.

Strukturierung des Suchraumes

Nach geeigneter Auswahl der zu beachtenden Eingangsgrößen muss der Suchraum, also die Menge aller möglichen Regeln, für die nachfolgende Suche strukturiert werden. Entscheidend hierfür ist die Partitionierung. Unter Partitionierung wird dabei sowohl die Festlegung der Anzahl der linguistischen Werte je Eingangs- bzw. Ausgangsgröße verstanden als auch die Festlegung der Zugehörigkeitsfunktionen für die linguistischen Werte. Dies kann sowohl wissensbasiert als auch datenbasiert [KS98] oder aber heuristisch erfolgen.

Im Fuzzy-ROSA-Verfahren kann die Zugehörigkeitsfunktion für einen linguistischen Wert rechteckförmig, trapezförmig oder dreieckförmig sein. Eine Zugehörigkeitsfunktion lässt sich in diesen Fällen sehr einfach durch vier Stützstellen beschreiben. Es wird ferner gefordert, dass ein Fuzzy-Informationssystem vorliegt, also die Summe der Zugehörigkeitsgrade über alle linguistischen Werte einer Größe für einen gegebenen reellen Wert stets Eins beträgt. Je nach gewählter Form können sich die Zugehörigkeitsfunktionen also mehr oder weniger stark überlappen.

Die Prämisse einer Regel definiert einen k -dimensionalen achsparallelen Hyperquader im n -dimensionalen Raum der betrachteten Eingangsgrößen, wobei $1 \leq k \leq n$ gilt. Ein solcher Hyperquader beinhaltet eine Teilmenge des zu Grunde liegenden Lerndatensatzes \mathcal{D} (Grundmenge). Die prinzipielle Idee der Regelgenerierung im Fuzzy-ROSA-Verfahren basiert darauf, zu prüfen, ob es zwischen einer so definierten Teilmenge und der Grundmenge einen wie auch immer gearteten signifikanten Unterschied gibt. Ist dies der Fall, kann eine entsprechende Regel formuliert werden, andernfalls nicht. Wird die Trennung der Daten *ungünstig* vorgenommen, also letztlich die Strukturierung des Suchraumes ungünstig gewählt, können *eigentlich vorhandene Unterschiede* (Regeln) eventuell nicht aufgedeckt werden.

Einschränkung des Suchraumes

Nach geeigneter Festlegung linguistischer Werte und Zugehörigkeitsfunktionen hierzu für die zu beachtenden Eingangs- und Ausgangsgröße(n) ist der Suchraum für die nachfolgende Regelsuche strukturiert. Die Anzahl potenziell relevanter vollständiger Regeln, in deren Prämisse *alle* ausgewählten Eingangsgrößen konjunktiv miteinander verknüpft sind, wächst mit der Anzahl an Eingangsgrößen aufgrund der resultierenden kombinatorischen Vielfalt sehr stark an [Kro99]. Werden nicht alle, sondern nur eine Teilmenge der Eingangsgrößen

in der Regelprämisse beachtet, so werden Regeln dieser Art als *generalisierende Regeln* [Kro99] bezeichnet. Die maximal zulässige Anzahl an Eingangsgrößen, die in einer Prämisse auftreten dürfen, wird im Folgenden als maximal zulässige Kombinationstiefe c_{max} bezeichnet.

Regelsuche

Prinzipielle Aufgabe der Regelsuche ist es, potenziell relevante Regeln als Hypothesen zu formulieren und diese einem geeigneten statistischen Test zu unterziehen. Wird eine Hypothese durch den Test bestätigt, so wird sie als relevante Regel angesehen, mit einem Glaubensmaß $\rho \in]0, 1]$ bewertet und der Regelbasis hinzugefügt. Wird eine Hypothese durch den Test widerlegt, so wird sie verworfen. Es existieren eine Reihe von Test- und Bewertungsverfahren, die in Kapitel 6 und Anhang A detaillierter beschrieben werden. Bevor eine Hypothese jedoch einem Test unterzogen wird, ist gegebenenfalls zu prüfen, ob sie bestimmte Restriktionen [Kro99, Sla01] erfüllt. Wird eine der gewünschten Restriktionen nicht erfüllt, wird die Hypothese ohne weiteren Test verworfen. Die Restriktionen werden nachfolgend kurz skizziert.

Kombinationsrestriktion:

Eine Hypothese, deren Prämisse Kombinationen von linguistischen Ausdrücken enthält, die sich auf die gleiche linguistische Variable beziehen, wird bei aktivierter Restriktion direkt verworfen. Ein Beispiel hierfür ist:

$$\text{WENN } \langle \text{Druck} = \text{niedrig} \wedge \text{Druck} = \text{hoch} \rangle \text{ DANN } \langle \text{Konklusion} \rangle \quad .$$

Komplementrestriktion:

Eine Hypothese, deren Prämisse negierte linguistische Ausdrücke enthält, wird bei aktivierter Restriktion direkt verworfen. Ein Beispiel hierfür ist:

$$\text{WENN } \langle \text{Druck ist nicht niedrig} \rangle \text{ DANN } \langle \text{Konklusion} \rangle \quad .$$

Minimaler Daten-Support:

Eine Hypothese, deren Prämisse nicht von einer ausreichenden Menge an Datenpunkten bzw. über alle Datenpunkte nicht in einem gewissen Mindestmaß gestützt wird, wird bei aktivierter Restriktion direkt verworfen. Der Daten-Support D_{sup} einer Prämisse ist im Fuzzy-ROSA-Verfahren folgendermaßen definiert:

$$D_{sup} = \sum_{i=1}^D \mu_P(\mathbf{x}_i) \quad . \quad (4.1)$$

Nur positive Regeln:

Bei aktivierter Restriktion wird die Generierung negativer Regeln (siehe Abschnitt 8.1) verhindert.

Die Durchmusterung eines strukturierten und gegebenenfalls eingeschränkten Suchraumes ist insbesondere bei hochdimensionalen Problemen, also Modellierungsproblemen mit vielen Eingangsgrößen, ein kritischer Faktor, da die Suche sehr zeitaufwendig werden kann. Vor diesem Hintergrund sind für das Fuzzy-ROSA-Verfahren verschiedene Suchalgorithmen entwickelt worden, von denen die wichtigsten im Folgenden vorgestellt werden.

Komplette Regelsuche:

Dieser Ansatz [Kro99] arbeitet *regelerorientiert*. Es werden dabei alle möglichen und im Sinne der gewählten Restriktionen zulässigen Prämissen aufgestellt. Jede Prämisse wird mit jedem linguistischen Wert der Ausgangsgröße zu einer Hypothese kombiniert und getestet. Auf diese Weise ist gewährleistet, dass alle im Sinne des angewandten Tests relevanten Regeln gefunden werden. Der Aufwand ist entsprechend groß und stößt bei hochdimensionalen Problemen aufgrund der großen Anzahl von zu prüfenden Hypothesen schnell an Grenzen bezüglich des vertretbaren Zeitaufwands.

Baumorientierte komplette Regelsuche:

Dieser Ansatz [Sla01] ist vom Prinzip her mit der kompletten Regelsuche identisch, er arbeitet jedoch *datenorientiert*. Es werden dabei alle möglichen und im Sinne der gewählten Restriktionen zulässigen Prämissen aufgestellt, die auch von Datenpunkten gestützt werden. Eine Prämisse (bzw. eine Hypothese), die von keinen Datenpunkten gestützt wird, kann auch keine Regel sein und muss daher gar nicht erst getestet werden. Gegenüber der kompletten Suche kann so eine deutliche Effizienzsteigerung erzielt werden, insbesondere bei hochdimensionalen Problemen, da in diesen Fällen der von den Eingangsgrößen aufgespannte Raum häufig nur spärlich durch Datenpunkte abgedeckt wird.

Hybride evolutionäre Regelsuche:

Bei einem sehr hochdimensionalen Suchraum kann es unter Umständen auch mit einer effizienten Suche nicht mehr möglich sein, den Raum komplett zu durchmustern. In solchen Fällen kann alternativ eine evolutionäre Regelsuche [KBT96, KK97, SKH⁺99, Sla01] zum Einsatz kommen. Das Ziel besteht dabei darin, möglichst *aussichtsreiche* Hypothesen aufzustellen. Durch die an die natürliche Evolution angelehnten Schritte Mutation, Rekombination und Selektion sollen aus guten Individuen (Regeln) so nach und nach noch bessere Individuen hervorgehen. Dabei ist es von Vorteil, den evolutionären Algorithmus entsprechend dem Verlauf der Suche zu adaptieren, also beispielsweise den Selektionsdruck zu verringern oder zu erhöhen. Dies erfolgt zum Beispiel mit Hilfe von Fuzzy-Regeln [Sla01], die anhand von Indikatoren, die den Suchverlauf sowohl qualitativ wie auch quantitativ beschreiben, eine Anpassung der freien Parameter des evolutionären Algorithmus vornehmen.

4.3 Generierung, Analyse und Optimierung einer Regelbasis

Die in Abschnitt 4.2 erwähnten Schritte (Auswahl der zu beachtenden Eingangsgrößen, Strukturierung des Suchraumes und Festlegung von Restriktionen) sind Vorstufen der eigentlichen Regelgenerierung. Im Folgenden wird zunächst genauer auf den Regelgenerierungsprozess eingegangen. Aufgrund des im Fuzzy-ROSA-Verfahren gewählten Ansatzes, einzelne plausible Regeln zu erzeugen und keine ganzen Regelbasen, ist es sinnvoll, das Zusammenwirken der so generierten Regeln nachgelagert genauer zu analysieren. Hierzu zählen verschiedene Aspekte, wie beispielsweise die Überprüfung der Vollständigkeit der Regelbasis oder aber die Betrachtung der Regeln im Hinblick auf widersprüchliche Aussagen. Nach einer evtl. vorgenommenen Reduktion der Regelbasis kann gegebenenfalls eine Optimierung der Zusammensetzung der Regelbasis angestrebt werden. Ziel dabei ist es, die Modellierungsgenauigkeit möglichst noch zu verbessern und gleichzeitig weniger Regeln hierfür zu benötigen.

Generierung relevanter Mamdani-Regeln

Das Fuzzy-ROSA-Verfahren generiert Regeln vom Typ Mamdani. Das Ergebnis des Regelgenerierungsprozesses hängt – bei gegebener Suchraumstrukturierung und bei gegebenen Restriktionen – von dem gewählten Test- und Bewertungsverfahren ab. Der Entwicklung der verschiedenen Test- und Bewertungsverfahren [Kro99, Jes00a] für die Regelgenerierung liegen unterschiedliche Anforderungen und Zielsetzungen zu Grunde, die sich wie folgt gruppieren lassen.

Relevanz:

Ziel dieser Strategie ist es, Regeln zu finden, die einen kausalen Zusammenhang zwischen der in der Regelprämisse spezifizierten linguistischen Eingangssituation und dem in der Regelkonklusion spezifizierten linguistischen Ausgangsgrößenwert beschreiben.

Treffericherheit:

Ziel dieser Strategie ist es, Regeln zu finden, die zu der in der Regelprämisse spezifizierten linguistischen Eingangssituation den korrekten linguistischen Ausgangsgrößenwert empfehlen.

Approximationsgüte:

Ziel dieser Strategie ist es, Regeln zu finden, die zu der in der Regelprämisse spezifizierten linguistischen Eingangssituation denjenigen linguistischen Ausgangsgrößenwert empfehlen, der im Sinne einer Minimierung des Approximationsfehlers am besten geeignet ist.

Die oben genannten Anforderungen und Zielsetzungen sind teilweise widersprüchlich oder gegenläufig. In der Praxis können sie daher im Allgemeinen nicht alle gleichzeitig voll erfüllt werden.

Die im Fuzzy-ROSA-Verfahren zur Verfügung stehenden Test- und Bewertungsverfahren lassen sich in zwei Klassen zusammenfassen. Eine Klasse bilden diejenigen Verfahren, die auf Grundlage der relativen Datenhäufigkeiten arbeiten. Eine weitere Klasse bilden diejenigen Verfahren, die sich auf Datenverteilungen stützen. Die Strategien der erstgenannten Klasse verfolgen eher die oben genannten Zielsetzungen Relevanz und Treffsicherheit, die der anderen Klasse primär das Ziel einer hohen Approximationsgüte. Ein weiterer zentraler Unterschied besteht darin, dass mit Hilfe der Strategien der erstgenannten Klasse auch *negative* Mamdani-Regeln generiert werden können, die eine Besonderheit des Fuzzy-ROSA-Verfahrens darstellen und im Zusammenhang mit dem Ansatz für ein hybrides Fuzzy-System in Kapitel 8 genauer erläutert werden. Die in der anderen Klasse zusammengefassten Test- und Bewertungsverfahren unterstützen die Generierung von Regeln dieser Art hingegen nicht.

Ein wesentlicher Gedanke des Fuzzy-ROSA-Verfahrens besteht darin, jeder generierten Regel ein Glaubensmaß $\rho \in]0, 1]$ zuzuweisen, das eine Aussage darüber liefert, wie gut die Regel durch die zu Grunde liegenden Lerndaten statistisch abgesichert ist [Kra94, Kro99]. Je höher das Glaubensmaß ist, desto gesicherter ist die Regel.

Auf weitere Details zu den verschiedenen Test- und Bewertungsverfahren wird an dieser Stelle nicht eingegangen, eine ausführliche Vorstellung der Verfahren folgt in Abschnitt 6.2 und in Anhang A.

Transformation von Mamdani-Regeln in Takagi-Sugeno-Kang-Regeln

Mamdani-Systeme zeichnen sich durch ihre gute Interpretierbarkeit und die damit verbundene Transparenz aus. Im Hinblick auf die Modellgenauigkeit schneiden sie jedoch meist deutlich schlechter ab als Takagi-Sugeno-Kang-Systeme. Aufgrund des Aufbaus der Konklusion einer Regel sind TSK-Systeme jedoch weniger gut nachvollziehbar. Beide Eigenschaften, Interpretierbarkeit und Genauigkeit, sind grundsätzlich wünschenswert bzw. sie werden von industriellen Anwendern gefordert. Oftmals muss jedoch eine der beiden Eigenschaften zugunsten der anderen vernachlässigt werden. Zur Überwindung dieses Problems wurde im Fuzzy-ROSA-Verfahren die Möglichkeit geschaffen, eine nachgelagerte Transformation eines datenbasiert generierten Mamdani-Systems in ein korrespondierendes TSK-System vorzunehmen [Kra00, Kra02]. Die Transformation ist dabei so konzipiert worden, dass die Interpretierbarkeit der Regeln möglichst erhalten bleibt und gleichzeitig die Approximationsgüte verbessert wird. Hierauf wird in Abschnitt 7.1 genauer eingegangen.

Analyse, Reduktion und Optimierung einer Fuzzy-Regelbasis

Bei dem gewählten Ansatz für das Fuzzy-ROSA-Verfahren wird im Rahmen der Regelsuche nur die Relevanz einzelner Regeln bewertet, nicht aber die der gesamten Regelbasis. Jede in die Regelbasis übernommene Regel wird als lokal vernünftig im Sinne des angewandten Testverfahrens angesehen. Dennoch kann nicht vollständig ausgeschlossen werden, dass die resultierende Regelbasis sowohl redundante als auch widersprüchliche Regeln enthält. Durch eine geeignete Reduktion der Regeln kann zum einen der Umfang der Regelbasis reduziert und zum anderen die Modellierungsgenauigkeit noch verbessert werden. Im Fuzzy-ROSA-Verfahren stehen verschiedene Reduktionsmechanismen [Kro99, Sla01, KKS00a, KKS00b] zur Verfügung, die im Folgenden kurz skizziert werden.

Online-Reduktion:

Dieser Ansatz verfolgt das Ziel, redundante Regeln schon während der Regelsuche zu vermeiden. Eine im Sinne des angewandten Test- und Bewertungsverfahrens lokal vernünftige Regel wird dabei als redundant angesehen, wenn sie von einer generelleren Regel, die einen höheren Bewertungsindex als die betrachtete Regel aufweist, überdeckt wird.

Situationsbasierte Konfliktreduktion:

Hierbei werden nur vollständige Regeln betrachtet. Es werden solche Regeln entfernt, die in keiner vollständigen linguistischen Eingangssituation zu den Regeln mit den höchsten Bewertungsindizes gehören. Nach Durchführung dieser Reduktion werden genauso viele linguistische Eingangssituationen von Regeln abgedeckt wie vor der Reduktion.

Strukturbasierte Konfliktreduktion:

Im Gegensatz zu der situationsbasierten Konfliktreduktion werden hier generalisierende linguistische Eingangssituationen betrachtet. Auch diese Reduktion arbeitet abdeckungserhaltend.

Datenbasierte Konfliktreduktion:

Dieser Ansatz zielt auf eine möglichst gleichmäßige Ausnutzung der zu Grunde liegenden Daten ab. Es werden solche Regeln reduziert, deren Support überwiegend aus solchen Daten besteht, die auch andere Regeln mit höherer Relevanz stützen. Die Abdeckungserhaltung kann in diesem Fall nicht garantiert werden.

Optimierende Konfliktreduktion:

Dieser Ansatz [KKS00a] zielt primär auf die Erhöhung der Modellierungsgenauigkeit ab. Aus der Menge aller vorhandenen Regeln soll diejenige Teilmenge gefunden werden, die den kleinsten Modellierungsfehler liefert. Andere Kriterien, wie etwa die Abdeckung der linguistischen Eingangssituationen durch die Regeln oder die Ausnutzung der Daten, werden dabei nicht beachtet. Eine entsprechende Teilmenge der

Regeln wird dabei als Bit-String codiert. Mit Hilfe eines genetischen Algorithmus wird anhand des jeweils erzielten Modellierungsfehlers die Zusammensetzung der Teilmengen optimiert.

Bewertung einer Fuzzy-Regelbasis

Das Fuzzy-ROSA-Verfahren basiert auf der Idee, die Suche nach einer geeigneten Regelbasis auf die Suche nach einzelnen relevanten Regeln zu reduzieren. Daher geht aber in den Generierungsprozess das Zusammenwirken der späteren Regelbasis nicht direkt mit ein. Eine Bewertung des Gesamtverhaltens erfolgt somit erst nach Abschluss der Regelgenerierung und ist insofern schwierig, als eine Reihe unterschiedlicher Bewertungskriterien sinnvoll erscheinen, die jedoch auch teilweise gegenläufig sein können. Es werden im Folgenden verschiedene Gütemaße vorgestellt, die im Zusammenwirken eine möglichst umfassende Bewertung einer Regelbasis ermöglichen. Weitere Gütemaße werden in [Sla01] vorgestellt und erläutert.

Modellierungsgüte:

Die erzielte Modellierungsgüte bzw. der Modellierungsfehler kann als wesentliches Gütemaß angesehen werden. Ist die erzielte Modellierungsgüte im Sinne des Anwenders nicht ausreichend, wird das Modell in der Praxis nicht zum Einsatz kommen. Die Wahl eines geeigneten Fehlermaßes hängt in erster Linie von der Aufgabenstellung ab. Bei Klassifikationsaufgaben wird typischerweise der relative Klassifikationsfehler (RCE) angegeben. Für Approximationsaufgaben werden meist der mittlere absolute Fehler (MAE) und der mittlere quadratische Fehler (MSE) angegeben. Die Fehlermaße sind in Anhang C angegeben. Ein Nachteil der oben genannten Fehlermaße MAE und MSE besteht darin, dass die erzielten Fehlerwerte problemspezifisch zu bewerten sind, das heißt im Kontext des Wertebereichs der Ausgangsgrößenwerte interpretiert werden müssen. Ein weiterer Nachteil aller oben genannten Fehlermaße liegt darin, dass sie keine Aussage darüber erlauben, wie gut das Modell im Verhältnis zu dem betriebenen Aufwand ist. Das denkbar einfachste Modell liefert zu jedem beliebigen Eingangsgrößenvektor \mathbf{x} stets den gleichen Ausgangsgrößenwert \hat{y}_{def} als Prognose. Mit Hilfe eines solchen trivialen Modells lassen sich normierte und anschauliche Fehlermaße definieren, die in Anhang C aufgeführt sind. Ein normiertes Fehlermaß von $\epsilon_{norm} = 1$ bedeutet, dass das betrachtete Fuzzy-Modell die gleiche Modellierungsqualität aufweist wie dieses denkbar einfachste Modell. Neben der Einstufung des Aufwand-Nutzen-Verhältnisses liegt ein weiterer Vorteil der normierten Fehlermaße darin, dass sie von der Größenordnung der Ausgangsgrößenwerte unabhängig sind. Daher ermöglichen sie auch einen besseren Vergleich der Performance auf verschiedenen Benchmarkproblemen. Ein weiteres gängiges Maß für die Modellierungsqualität, das ebenfalls eine gute Vergleichbarkeit ermöglicht, ist die Korrelation der prognostizierten zu den gemessenen Ausgangsgrößenwerten. In der Praxis wird ein Anwender

neben diesen quantitativen Maßen aber auch immer visuell überprüfen, wie gut ein in den Ausgangsgrößenwerten vorhandener Trend tatsächlich modelliert wird. Eine solche morphologische Bewertung erfolgt anhand individueller Präferenzen und lässt sich formelmäßig schlecht erfassen. Daher wird diese Art der Bewertung in dieser Arbeit nicht weiter berücksichtigt.

Vollständigkeit:

Neben einer ausreichenden Modellierungsgüte ist die Vollständigkeit einer Regelbasis von Interesse. Ein vollständiger Regelsatz garantiert, dass für jeden Eingangsgrößenvektor \mathbf{x} eine Regel der Regelbasis aktiviert wird und das Fuzzy-System somit einen Ausgangsgrößenwert \hat{y} liefert. Bei einer unvollständigen Regelbasis ist dies nicht der Fall. Falls es gewünscht ist, dass das Fuzzy-System auch in einem solchen Fall einen Ausgangsgrößenwert liefert, muss eine Strategie hierfür festgelegt werden. Ein naheliegender Ansatz besteht darin, einen Standard-Ausgabewert (Default-Wert) festzulegen. Dieser dürfte in den meisten Fällen allerdings nur eine Behelfslösung sein und eine vergleichsweise ungenaue Modellierung bzw. Prognose liefern. Daher ist es von Interesse, in wie vielen Fällen auf diese Behelfslösung zurückgegriffen werden muss.

Transparenz und Interpretierbarkeit:

Die Fuzzy-Modellierung führt vom Grundansatz her zu interpretierbaren Modellen und hat daher prinzipiell Vorteile gegenüber rein deskriptiven Modellierungsverfahren. Es besteht jedoch gerade bei einer datenbasierten Modellgenerierung die Gefahr, dass dieser Aspekt vernachlässigt wird und so dieser grundsätzliche Vorteil teilweise verloren gehen kann. Transparenz und Interpretierbarkeit sind keine fest definierten Begriffe und werden von Menschen unterschiedlich ausgelegt. Die Überlegungen konzentrieren sich daher an dieser Stelle auf zwei Aspekte: zum einen auf die grundsätzliche Vorgehensweise zur Erstellung einer Regelbasis und zum anderen auf den Umfang der generierten Regelbasis und die Komplexität der darin enthaltenen Regeln. Wie in Abschnitt 4.1 ausgeführt, basiert das Fuzzy-ROSA-Verfahren auf dem Michigan-Ansatz und erzeugt daher solche Regeln, die lokal vernünftig sind und zuvor einen entsprechenden Test bestanden haben. Diese Vorgehensweise hat im Hinblick auf die Transparenz des erzeugten Modells Vorteile gegenüber der als Pittsburgh-Ansatz bekannten Vorgehensweise. Auch wenn die einzelnen Regeln lokal vernünftig und nachvollziehbar sind, so wird es doch mit zunehmendem Umfang der Regelbasis immer problematischer, das Zusammenwirken der Regeln nachzuvollziehen. Es sollte immer angestrebt werden, eine möglichst genaue, aber auch gleichzeitig möglichst kompakte Regelbasis zu erstellen. Die Anzahl der Regeln ist jedoch auch immer in Verbindung mit der Komplexität der erzeugten Regeln zu sehen. So kann beispielsweise eine Regelbasis, deren Regeln nur konjunktive Verknüpfungen in der Prämisse enthalten, dadurch stark komprimiert werden, dass Regeln mit gleicher Konklusion zusammengefasst werden, indem disjunktive Verknüpfungen in der Prämisse zugelassen werden. So ist die Regelbasis zwar kleiner, aber die Regeln sind auch deutlich komplexer.

Kapitel 5

Analyse des Verfahrens im Hinblick auf Verbesserungspotenzial

Im Sinne des angestrebten Transfers der Methoden zur datenbasierten Fuzzy-Modellierung in die Praxis ist es notwendig, zunächst die Anforderungen potenzieller Anwender zu erkunden und diese zu formulieren. Darauf aufbauend wird das Fuzzy-ROSA-Verfahren im Hinblick auf die formulierten Anforderungen analysiert, um mögliche Schwachstellen des Verfahrens bzw. Potenzial für Verbesserungen zu orten. Um das Verfahren noch besser für industrielle Problemstellungen einsetzen zu können, werden eine Reihe von Anpassungen bzw. Erweiterungen an dem Verfahren vorgenommen, die aber nicht den Schwerpunkt dieser Arbeit bilden und daher in Abschnitt 4.3 vergleichsweise knapp beschrieben werden. Die zur Überwindung bisheriger Schwachstellen und zur Erweiterung der Leistungsfähigkeit des Verfahrens weiter- bzw. neu entwickelten Methoden werden in den Kapiteln 6 bis 8 vorgestellt.

5.1 Anforderungen industrieller Anwender

Die Anforderungen, die potenzielle Anwender an ein Modellierungsverfahren stellen, sind immer sehr stark von der jeweiligen Modellierungsaufgabe, aber auch von internen Rahmenbedingungen, wie beispielsweise der zur Verfügung stehenden Zeit für die Bearbeitung eines Projektes, abhängig. Eine umfassende Beschreibung aller Anforderungen ist daher kaum möglich und auch nicht Ziel dieser Ausführungen. Es geht vielmehr darum, solche Anforderungen zu formulieren und zusammenzufassen, die problemunabhängig auftreten und deren angemessene Beachtung für die Akzeptanz des Verfahrens in der Praxis von großer Bedeutung ist. Die von Anwendern häufig formulierten Forderungen und Wünsche lassen sich unter verschiedenen Aspekten folgendermaßen zusammenfassen.

Aufwand für den Anwender des Modellierungsverfahrens:

Für Anwender ist es von großer Bedeutung, dass ihnen die Methoden für die Modellierung in einer einheitlichen Programmumgebung zur Verfügung stehen und nicht unterschiedliche Programme eingesetzt werden müssen. So wird den Anwendern die zeit- und kostenaufwendige Einarbeitung in die Bedienung verschiedener Programme sowie der damit verbundene Datentransfer erspart. Neben den Anforderungen an die methodische Leistungsfähigkeit ist für Anwender auch die Bedienerfreundlichkeit der Software entscheidend. Neben einer grafischen Benutzeroberfläche, einer transparenten Bedienstruktur und einer leicht verständlichen Hilfe ist es zudem oftmals wünschenswert, unterschiedliche Benutzertypen – zum Beispiel Anfänger, Fortgeschrittener und Experte – zu definieren, um so den jeweiligen Erfahrungsstand des Anwenders im Umgang mit dem eingesetzten Modellierungsverfahren angemessen zu berücksichtigen. Anfangs ist es für Anwender oftmals hilfreich, wenn sie sich zunächst nur mit vergleichsweise wenig Einstelloptionen befassen müssen und so bereits akzeptable Modellierungsergebnisse erhalten. Durch den Umgang mit dem Modellierungsverfahren sammeln Anwender eigene Erfahrungen und Erfolge und sind daher zunehmend bereit, sich auch mit weiteren Einstelloptionen des Modellierungsverfahrens, die die Modellgenauigkeit weiter verbessern können, zu befassen.

Zeitaufwand für die Modellerstellung:

Durch den zunehmenden Wettbewerb in allen Bereichen der Industrie ist es zwingend notwendig, anspruchsvolle Projekte kosteneffizient durchzuführen. Die entstehenden Kosten sind dabei oftmals in erster Linie von der Zeitspanne abhängig, die für die Bearbeitung eines Projektes benötigt wird. Die Modellierung großer und komplexer Systeme ist in der Regel nicht auf Anhieb erfolgreich, sondern ist vielmehr ein iterativer Prozess, bei dem schrittweise neue Erkenntnisse gesammelt werden. Um dem Aspekt der Kosteneffizienz gerecht zu werden, muss es Anwendern daher möglich sein, in relativ kurzer Zeit ein Modell zu erstellen, um so möglichst viele Iterationen in der zur Verfügung stehenden Zeit durchführen zu können.

Messaufwand für die Modellerstellung:

Grundlage einer datenbasierten Modellierung sind die zu einem System oder Prozess erhobenen Messdaten. In vielen praktischen Anwendungen können aus Zeit- und Kostengründen oft nur sehr wenige Messdaten für eine Modellierung bereitgestellt werden. Bei der Untersuchung von chemischen Mischungen beispielsweise sind die eingesetzten Rohstoffe oftmals teuer und daher knapp bemessen. Auch in solchen Fällen muss ein akzeptables Modellierungsergebnis erzielt werden. Die Erfassung von Messwerten ist oftmals auch mit einem erheblichen technischen Aufwand – zum Beispiel in Form von teuren Sensoren – verbunden. Es ist daher erstrebenswert, Einflussgrößen zu orten, die für die Modellierung nicht erforderlich sind oder sich aus anderen Einflussgrößen rekonstruieren lassen. Auf diese Weise können unter Umständen teure Sensoren und damit verbundene Kosten eingespart werden.

Robustheit der Modellerstellung:

In industriellen Anwendungen kann eine fehlerbehaftete Datenerfassung nie ausgeschlossen werden. Die Quellen möglicher Fehler bei der Datenerhebung sind dabei vielfältig: Störung von Sensorwerten durch Rauschen, defekte Sensoren oder Fehler bei der manuellen Dateneingabe. Eine große Schwierigkeit besteht bereits darin, zu erkennen, dass Daten fehlerhaft sind. Insbesondere bei hochdimensionalen Problemen, die sich weniger gut visualisieren lassen, ist dies ein nicht triviales Problem. Die Analyse und Aufbereitung eines Datensatzes in dieser Hinsicht erfordert Erfahrung und einen nicht unerheblichen Aufwand. Industrielle Anwender verfügen oftmals nicht über eine entsprechende Erfahrung oder sind nicht bereit, vor der Durchführung einer Modellierung einen entsprechenden Zeit- und Kostenaufwand für die Datenaufbereitung zu betreiben. Hieraus leitet sich die Forderung ab, dass das eingesetzte Modellierungsverfahren eine hohe Robustheit gegenüber solchen Daten aufweist und auch in einem solchen Fall ein akzeptables Modellierungsergebnis liefert.

Transparenz der generierten Modelle:

Neben der Genauigkeit der Modelle ist für Anwender auch deren Transparenz von großer Bedeutung. Ein Modell wird erst dann vollständig akzeptiert, wenn es einen Plausibilitätstest besteht, das heißt mit dem vorhandenen qualitativen Prozesswissen erfolgreich abgeglichen werden kann. Dies ist auch verständlich, weil die Anwender letztlich die Verantwortung für die ordnungsgemäße Funktion der Modelle übernehmen müssen. Hieraus leitet sich der Anspruch ab, dass das Modell möglichst kompakt sein sollte, das heißt so wenig Eingangsgrößen wie nötig verwendet und einen möglichst kleinen Satz von Fuzzy-Regeln benötigt. Die Kompaktheit des Modells erleichtert so die Durchführung des Plausibilitätstests zum Beispiel durch die Überprüfung einzelner verständlicher Fuzzy-Regeln. Ein weiterer Anspruch ist, dass überall dort, wo eine Erhöhung der Genauigkeit des Modells erzielt werden kann, dies möglichst unter Erhaltung der Transparenz geschehen soll.

5.2 Abgleich des Verfahrens mit den Anforderungen

Die in Abschnitt 5.1 formulierten Anforderungen potenzieller Anwender werden mit dem Fuzzy-ROSA-Verfahren in der Form, wie es zu Beginn dieser Arbeit vorlag, abgeglichen.

Aufwand für den Anwender des Modellierungsverfahrens:

Bei diesem Aspekt steht die geeignete softwaretechnische Umsetzung des Verfahrens im Vordergrund, nicht die Erweiterung der methodischen Leistungsfähigkeit. Diese Arbeit wurde überwiegend von dem Kooperationspartner NuTech Solutions geleistet.

Zeitaufwand für die Modellerstellung:

Der Zeitaufwand für die Erstellung eines Modells hängt in erster Linie davon ab, wie

viele Eingangsgrößen bei der Regelgenerierung beachtet werden. Es ist daher erstrebenswert, Eingangsgrößen zu orten, die für die Modellierung nicht erforderlich oder sogar eher störend sind. Bisher war es innerhalb des Verfahrens nur bedingt möglich, eine datenbasierte Selektion möglichst geeigneter Eingangsgrößen vorzunehmen, beispielsweise mit Hilfe eines Verfahrens zur datenbasierten Komplexitätsreduktion [Pra99]. Eine entsprechende Auswahl musste daher typischerweise wissensbasiert erfolgen oder mit externen Programmen vorgenommen werden. Hier liegt eine Schwachstelle des Verfahrens.

Messaufwand für die Modellerstellung:

In praktischen Anwendungen können aus Kostengründen oftmals nur wenige Messdaten für eine Modellierung zur Verfügung gestellt werden. Die Regelgenerierung ist in solchen Fällen unter Umständen problematisch, weil die Test- und Bewertungsverfahren keine befriedigenden Resultate liefern. Wie in Abschnitt 4.3 erwähnt, lassen sich die Verfahren in zwei übergeordnete Klassen aufteilen. Die Verfahren, die auf relativen Datenhäufigkeiten basieren, arbeiten größtenteils mit Konfidenzintervallen zu den geschätzten relativen Häufigkeiten. Bei einer geringen Anzahl an Daten fallen die Konfidenzintervalle entsprechend groß aus und sind daher oftmals nicht disjunkt. In solchen Fällen wird dann keine Regel generiert, wodurch die Modellierung dann unter Umständen entsprechend schlecht gelingt. Die Verfahren, die auf Datenverteilungen basieren, arbeiten zwar nicht mit Konfidenzintervallen, machen aber gewisse Annahmen über die zu Grunde liegenden Daten. Beispielsweise liegt den Verfahren die Annahme zu Grunde, dass die Daten normalverteilt sind. Diese Annahme ist zwar in vielen praktischen Fällen bei einer entsprechenden Datenmenge vertretbar, gerade aber bei nur wenigen Daten kann diese Annahme jedoch nicht zutreffend sein. Für die Modellierung im Fall einer nur kleinen Datenbasis sind daher neue Strategien zu entwickeln.

Robustheit der Modellerstellung:

In industriellen Anwendungen kann eine fehlerbehaftete Datenerfassung nie ausgeschlossen werden. Eine Analyse bzw. darauf aufbauende Aufbereitung der Daten ist eine nicht triviale Aufgabe, und gerade in industriellen Anwendungen kann nicht davon ausgegangen werden, dass die für die Modellierung verwendeten Daten zuvor entsprechend aufbereitet werden. Fehlerhafte Daten können sich sowohl bei der Generierung von Mamdani-Regeln als auch bei deren Transformation in TSK-Regeln negativ auswirken. Bei der Regelgenerierung sind zwei Fälle voneinander zu unterscheiden: Zum einen können eigentlich vorhandene Zusammenhänge durch fehlerhafte Daten verdeckt werden, zum anderen können aber auch solche Zusammenhänge vorhanden erscheinen, die eigentlich gar nicht vorhanden sind. Bei der Transformation einer Mamdani-Regel in eine TSK-Regel können bereits wenige fehlerhafte Daten einen signifikanten Einfluss auf die Bestimmung der freien Parameter in der Konklusion der TSK-Regel haben. Dies gilt insbesondere auch für die Parameterbestimmung über den herkömmlichen Least-Squares-Ansatz (siehe Abschnitt 7.2.2). In den

Teilmodulen Regelgenerierung und Regeltransformation soll die Robustheit gegenüber fehlerhaften Daten verbessert werden.

Transparenz der generierten Modelle:

Die Transparenz eines Modells lässt sich nur bedingt beschreiben, da es keine festen Kriterien hierfür gibt. Fuzzy-Modelle bzw. Fuzzy-Regeln sind aufgrund der Art, wie sie Zusammenhänge beschreiben, in sich bereits vergleichsweise transparent. Damit ergeben sich Vorteile gegenüber anderen Ansätzen, wie beispielsweise neuronalen Netzen. Aufgrund des im Fuzzy-ROSA-Verfahrens verfolgten Michigan-Ansatzes sind auch die einzelnen Regeln im Sinne des zu Grunde liegenden Testverfahrens sinnvoll. Dies erleichtert die Überprüfung bzw. den Abgleich der so erzeugten Regeln mit dem vorhandenen qualitativen Prozesswissen. Auch die optionale Transformation der Mamdani-Regeln in TSK-Regeln ist so gestaltet, dass die Interpretierbarkeit der Regeln möglichst gut erhalten bleibt. Ein entscheidender Faktor in diesem Zusammenhang ist jedoch die Größe der Regelbasis. Auch wenn jede einzelne Regel plausibel und verständlich ist, fällt es doch mit steigender Anzahl an Regeln immer schwerer, die Plausibilität einer Regelbasis als Ganzes zu beurteilen. Ein weiteres Ziel des Fuzzy-ROSA-Verfahrens besteht deshalb darin, eine möglichst kompakte Regelbasis zu verwenden. Hierfür sind zum einen verschiedene Reduktionsverfahren entwickelt worden. Zum anderen zielen aber auch die negativen Regeln in diese Richtung, da sie eine überaus kompakte Darstellung bzw. Formulierung von Zusammenhängen dieser Art erlauben. Entsprechende Zusammenhänge lassen sich im Prinzip auch durch positive Regeln ausdrücken, allerdings ist die Repräsentation dann wesentlich umfangreicher und daher auch weniger anschaulich. Der Aspekt der Kompaktheit ist bei TSK-Systemen verbesserungsfähig, da negative Regeln hierfür bisher nicht genutzt werden konnten. Die Nutzbarmachung von negativen Regeln für TSK-Systeme verspricht ein erhebliches Verbesserungspotenzial.

5.3 Anwenderorientierte Erweiterungen des Verfahrens

Nachfolgend werden einige Erweiterungen, die an dem Fuzzy-ROSA-Verfahren vorgenommen wurden, beschrieben. Diese stellen jedoch nicht den Schwerpunkt der Arbeit dar und werden daher an dieser Stelle vergleichsweise knapp behandelt.

Um industriellen Anwendern eine komfortable Verwaltung ihrer Daten und Modelle zu ermöglichen, sind Anbindungen an gängige Datenbanktypen (Microsoft Access, Microsoft MySQL-Server und MySQL) geschaffen worden. Ferner sind eine grafische Benutzeroberfläche und ein Hilfesystem realisiert worden, um Anwendern die komplexen Modellierungs- und Optimierungsmethoden des Verfahrens leicht zugänglich zu machen. Diese Arbeiten wurden in erster Linie vom Kooperationspartner NuTech durchgeführt.

In Abschnitt 4.3 wird darauf hingewiesen, dass der Zeitaufwand für die datenbasierte Erstel-

lung eines Modells mit dem Fuzzy-ROSA-Verfahren primär von der Anzahl der zu beachtenden Eingangsgrößen abhängt, da mit zunehmender Anzahl der Eingangsgrößen die Anzahl der zu prüfenden Hypothesen exponentiell ansteigt und eine vollständige Durchmusterung des Suchraumes oftmals nicht mehr in vertretbarer Zeit möglich ist. Eine mögliche Vorgehensweise kann in diesem Fall darin bestehen, eine evolutionäre Regelsuche durchzuführen (siehe Abschnitt 4.2). Eine andere Herangehensweise besteht darin, die Anzahl der Eingangsgrößen so weit zu reduzieren, dass sich ein Suchraum handhabbarer Größe ergibt. Die Auswahl relevanter Eingangsgrößen ist aber auch deshalb von Interesse, da sich so unter Umständen auch Kosten einsparen lassen, beispielsweise teure Sensoren.

Verfahren für die Selektion von Eingangsgrößen lassen sich anhand von drei Kriterien kennzeichnen: Erstens ist danach zu fragen, ob in dem Verfahren die Möglichkeit zur Bewertung kompletter Sätze von Eingangsgrößen und nicht nur von individuellen Größen vorgesehen ist. Zweitens ist nach der verwendeten Suchstrategie für Sätze von Eingangsgrößen und drittens nach der Bewertungsfunktion zu fragen, die zur Relevanzbewertung eingesetzt wird. Die Möglichkeit zur Bewertung kompletter Sätze von Eingangsgrößen muss hier gewährleistet sein, um die Interaktion der Eingangsgrößen bereits bei deren Auswahl zu berücksichtigen, da diese Interaktion für den Erfolg der nachfolgenden datenbasierten Modellierung von entscheidender Bedeutung ist. Es müssen ferner geeignete Strategien für die Suche nach relevanten Sätzen von Eingangsgrößen zur Verfügung stehen, um die Selektion in annehmbarer Zeit durchführen zu können. Bei der eingesetzten Bewertungsfunktion kann generell zwischen zwei Ansätzen unterschieden werden. Bei dem *Filter-Ansatz* [KW00] wird ein meist einfach und schnell zu berechnendes Maß (zum Beispiel die aus der Statistik und der Informationstheorie bekannten Maße Korrelation und Transinformation [Mat96]) verwendet. Im Gegensatz dazu wird bei dem *Wrapper-Ansatz* [KJ97] für die Bewertung eines Satzes von Eingangsgrößen die Güte des tatsächlichen Modells herangezogen, die sich bei der Verwendung des gewählten Satzes von Eingangsgrößen ergibt.

Eine vollständige Durchmusterung aller möglichen Sätze von Eingangsgrößen ist aufgrund der kombinatorischen Vielfalt und dem damit verbundenen Zeitaufwand in der Praxis oftmals nicht möglich. Häufig wird daher eine *Greedy-Suche* eingesetzt, bei der ein Satz von Eingangsgrößen, ausgehend von einer Eingangsgröße, sukzessive erweitert oder, ausgehend von allen Eingangsgrößen, sukzessive reduziert wird. Bei diesem Suchansatz ist der Aufwand zwar geringer als bei einer vollständigen Suche, es ist aber nicht gewährleistet, dass der so gefundene Satz von n Eingangsgrößen auch derjenige aller möglichen Sätze mit n Eingangsgrößen ist, der die höchste Relevanz aufweist. Daher wird eine dritte Suchstrategie entwickelt, die als *eingeschränkte vollständige Suche* bezeichnet wird. Diese Strategie generiert und testet all diejenigen Sätze, die eine bestimmte Anzahl an Eingangsgrößen beinhalten. Der Nachteil der Greedy-Suche wird so behoben, und da der Suchraum bei dieser Strategie meist deutlich kleiner ist als bei einer vollständigen Suche, kann die Selektion in einer für den Anwender vertretbaren Zeit durchgeführt werden. Der Wrapper-Ansatz zur Bewertung der Relevanz eines betrachteten Satzes von Eingangsgrößen scheidet hier aus, da der Zeitaufwand zu groß ist und sich so nicht die gewünschte Beschleunigung

der Modellerstellung erzielen lässt. Es werden daher Selektionsverfahren eingesetzt, die auf dem Filter-Ansatz basieren. Die Verfahren arbeiten hier rekonstruktionsbasiert, das heißt es wird auf Grundlage eines Satzes ausgewählter Eingangsgrößen ein vergleichsweise einfaches und schnell zu berechnendes Modell erzeugt und bewertet. Der erzielte Modellierungsfehler wird als Bewertungsmaß für die Relevanz des betrachteten Satzes von Eingangsgrößen aufgefasst. Je kleiner der Fehlerwert ist, desto höher ist die Relevanz. Der Satz von Eingangsgrößen mit der höchsten Relevanz wird für die nachfolgende datenbasierte Modellierung mit dem Fuzzy-ROSA-Verfahren eingesetzt. Dem ersten Selektionsverfahren [SKMM03] liegt eine abstands-basierte Rekonstruktionsfunktion zu Grunde. Der verwendete Rekonstruktionsalgorithmus schätzt zu einem Eingangsgrößenvektor \mathbf{x}_j einen Ausgangsgrößenwert \hat{y}_j auf der Basis aller übrigen Datenpunkte, wobei möglichst $y_j = \hat{y}_j$ gelten soll. Dieses Prinzip wird auch als *leave one out cross-validation* bezeichnet. Der Ausgangsgrößenwert y_i eines Datenpunktes geht um so stärker in die Rekonstruktion ein, je kleiner der Abstand zwischen den Eingangsgrößenvektoren \mathbf{x}_j und \mathbf{x}_i ist, was über eine geeignete Gewichtungsfunktion erreicht wird. Der Abstand wird dabei jeweils in dem Raum bestimmt, der von den in diesem Fall betrachteten Eingangsgrößen aufgespannt wird. Dieser Ansatz ist eher für den Fall vergleichsweise weniger Datenpunkte praktikabel, da der Rekonstruktionsalgorithmus die Ordnung $O(N^2)$ aufweist. Das zweite Selektionsverfahren [SNK01a, SNK01b] arbeitet ebenfalls rekonstruktionsbasiert, allerdings werden hier für die Rekonstruktion eines Datenpunktes nicht alle übrigen Datenpunkte verwendet, sondern nur solche, die in dessen Nachbarschaft liegen. Als einander benachbart werden in diesem Fall solche Datenpunkte angesehen, die in einer gemeinsamen Zelle liegen. Unter einer Zelle wird dabei ein achsparalleler Hyperquader im n -dimensionalen Raum der betrachteten Eingangsgrößen verstanden. Der Wertebereich einer Eingangsgröße X_i wird durch k_i frei wählbare Stützstellen in $(k_i + 1)$ überlappungsfreie Intervalle unterteilt¹. Durch die Gesamtheit aller Stützstellen wird der von den n Eingangsgrößen aufgespannte Raum in $m = \prod_{i=1}^n (k_i + 1)$ überlappungsfreie, achsparallele Hyperquader segmentiert. Ein Hyperquader entspricht im Prinzip dem Teilvolumen, das von der Prämisse einer vollständigen Regel, so wie sie im Fuzzy-ROSA-Verfahren ausgelegt ist, abgedeckt wird. Dieses Zellenmodell weist jedoch zwei wesentliche Unterschiede gegenüber einem Fuzzy-Modell auf: erstens werden keine graduellen Zugehörigkeiten zu einem Hyperquader verwendet, und zweitens werden nur vollständige und keine generalisierenden Hyperquader betrachtet. Das Auffinden mit Datenpunkten besetzter Zellen erfolgt mittels einer effizienten baumorientierten Suche. Für diejenigen Datenpunkte, die in einer gemeinsamen Zelle liegen, wird zum Beispiel im Fall eines Approximationsproblems der mittlere Ausgangsgrößenwert \bar{y} dieser Datenpunkte berechnet und für deren Rekonstruktion $\hat{y}_i = \bar{y}$ verwendet². Für einen Satz von Eingangsgrößen und zugehöriger Stützstellen lässt sich so ein Rekonstruktionsfehler bezüglich aller Datenpunkte berechnen und für eine evolutionäre Optimierung der Stützstellen

¹Die untere und die obere Grenze des Wertebereichs einer Eingangsgröße X_i werden nicht als Stützstellen aufgefasst, da sie nicht frei platzierbar sind. Diese Grenzen werden zusätzlich zu den k_i Stützstellen verwendet, um die $(k_i + 1)$ Intervalle festzulegen.

²Im Falle eines Klassifikationsproblems wird für \hat{y}_i entsprechend der Wert der am häufigsten auftretenden Klasse in einer Zelle genutzt.

nutzen. Abbildung 5.1 zeigt beispielhaft die Vorzüge einer Zelleneinteilung durch evolutionär optimierte Stützstellen gegenüber zum Beispiel äquidistant angeordneten Stützstellen anhand eines Klassifikationsproblems.

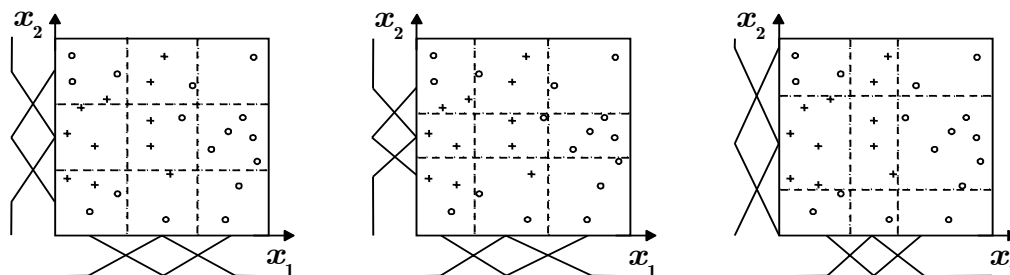


Abbildung 5.1: Beispiele für unterschiedliche Zelleneinteilungen und daraus generierten Zugehörigkeitsfunktionen: äquidistant (links), äquifrequent (Mitte) und optimiert (rechts).

In diesem Beispiel mit zwei Eingangsgrößen X_1 und X_2 sind nur zwei Klassen vorhanden, in Abbildung 5.1 dargestellt durch Kreise und Kreuze. In einer Zelle sollten möglichst nur Datenpunkte, die zu einer Klasse gehören, enthalten sein, damit der Rekonstruktionsfehler minimal wird. Es ist zu erkennen, dass hier sowohl bei einer äquidistanten als auch bei einer äquifrequenten Platzierung der Stützstellen die angestrebte Trennung nicht vollständig gelingt. Im Gegensatz dazu werden bei den evolutionär optimierten Stützstellen die Datenpunkte in der gewünschten Weise voneinander separiert. Aus den optimierten Stützstellen der letztlich selektierten Eingangsgrößen können nachgeschaltet automatisch linguistische Werte und entsprechende Zugehörigkeitsfunktionen hierzu für die Regelsuche mit dem Fuzzy-ROSA-Verfahren generiert werden [SNK01a]. Dies geschieht in der Art, dass sich die dabei erzeugten Zugehörigkeitsfunktionen in einem vom Anwender vorgegebenen Maße überlappen und so eine der wesentlichen Eigenschaften eines Fuzzy-Modells gegeben ist. In Abbildung 5.1 ist dies ebenfalls beispielhaft dargestellt. Der gewählte Ansatz des Selektionsverfahrens (Segmentierung in überlappungsfreie, achsparallele Hyperquader) ist dem des eingesetzten Modellierungsverfahrens (Regelsuche in überlappenden, achsparallelen Hyperquadern) sehr ähnlich und führt daher oftmals zu signifikanten Verbesserungen gegenüber den Resultaten, die mit Standard-Partitionierungen erzielt werden [SNK01a].

Kapitel 6

Neue Strategieelemente für die Generierung von Mamdani-Regeln

Wie in Abschnitt 4.3 ausgeführt, stehen im Fuzzy-ROSA-Verfahren verschiedene Strategien für die Generierung relevanter Mamdani-Regeln zur Auswahl, um unterschiedliche Zielsetzungen realisieren zu können. Im Folgenden werden zunächst die in früheren Arbeiten entwickelten Strategien getrennt und nach den in Abschnitt 4.3 beschriebenen Klassen genauer vorgestellt. Daran anschließend wird in Abschnitt 6.3 ein in dieser Arbeit neu entwickeltes Test- und Bewertungsverfahren vorgestellt, das auf den Fall kleiner und/oder stark gestörter Datensätze abzielt [SKK04]. In Abschnitt 6.4 wird ein spezielles Konzept für die Regelgenerierung vorgestellt, das auf eine möglichst gleichberechtigte Behandlung positiver und negativer Regeln abzielt [KKSS02b].

6.1 Test- und Bewertungsverfahren basierend auf Datenhäufigkeiten

Die unbedingte Wahrscheinlichkeit $p(C)$ für das Auftreten der Konklusion C einer Hypothese ist unbekannt und wird, basierend auf den vorhandenen Daten, wie folgt geschätzt:

$$\hat{p}(C) = \frac{1}{D} \cdot \sum_{q=1}^D \mu_{Y_j}(y_q) \quad . \quad (6.1)$$

Die bedingte Wahrscheinlichkeit $p(C|P)$ für das Auftreten der Konklusion C einer Hypothese unter der Voraussetzung, dass die in der Prämisse P definierte Bedingung vorliegt,

ist unbekannt und wird, basierend auf den vorhandenen Daten, wie folgt geschätzt:

$$\hat{p}(C|P) = \frac{1}{D_{sup}} \cdot \sum_{q=1}^D \mu_{Y_j}(y_q) \wedge \mu_P(\mathbf{x}_q) \quad . \quad (6.2)$$

Da die wahren Wahrscheinlichkeiten nicht bekannt sind und daher anhand der vorhandenen Daten (Stichproben) geschätzt werden müssen, ist es angebracht, zu den geschätzten Werten entsprechende Konfidenzintervalle [HEK95, Kro99] anzugeben. Der wahre, aber unbekannt Wert sollte innerhalb des berechneten Konfidenzintervalls liegen. Dies lässt sich jedoch nicht mit absoluter Sicherheit gewährleisten. Darum wird zu einem Konfidenzintervall stets eine bestimmte Irrtumswahrscheinlichkeit α angegeben. Typischerweise wählt man $\alpha = 0.01$, also eine Irrtumswahrscheinlichkeit von 1%. Dies bedeutet anschaulich, dass in 99% aller Stichproben das jeweilige Konfidenzintervall den wahren Wert enthält und in 1% nicht. Oftmals wird dies auch als Wahrscheinlichkeit, dass der wahre Wert innerhalb des Konfidenzintervalls liegt, bezeichnet. Die Größe des Konfidenzintervalls richtet sich nach der geforderten Irrtumswahrscheinlichkeit und dem Umfang der Stichprobe. Betrachtet man den Fall einer Stichprobe konstanten Umfangs und einer variablen Irrtumswahrscheinlichkeit, so wird das Konfidenzintervall umso größer, je kleiner die geforderte Irrtumswahrscheinlichkeit gewählt wird. Mit anderen Worten: Man erreicht die gewünschte Sicherheit dadurch, dass man das Konfidenzintervall groß genug wählt. Betrachtet man hingegen den Fall einer Stichprobe variablen Umfangs und einer konstanten Irrtumswahrscheinlichkeit, so wird das Konfidenzintervall umso größer, je kleiner der Umfang der Stichprobe ist. Mit anderen Worten: Je größer die Stichprobe ist, desto gesicherter ist der geschätzte Wert und desto kleiner kann das Konfidenzintervall ausfallen.

Zu einem geschätzten Wert kann ein einseitiges oder ein zweiseitiges Konfidenzintervall angegeben werden. Ein zweiseitiges Konfidenzintervall liegt in der Regel symmetrisch um den geschätzten Wert. Die Wahrscheinlichkeit, dass der wahre Wert oberhalb der oberen Grenze des Intervalls liegt, soll gleich groß sein wie die Wahrscheinlichkeit, dass der wahre Wert unterhalb der unteren Grenze des Intervalls liegt. Bei einem einseitigen Konfidenzintervall wird eine der beiden Grenzen entsprechend zu $+\infty$ bzw. $-\infty$ gesetzt. In diesem Fall ist es nur von Interesse, ob der wahre Wert unterhalb bzw. oberhalb eines bestimmten Grenzwertes liegt. Lautet die interessierende Fragestellung beispielsweise, wie hoch der durchschnittliche Benzinpreis ist, so muss zu dem mittleren Benzinpreis, der anhand der zu Grunde liegenden Stichprobe von Benzinpreisen bestimmt worden ist, ein zweiseitiges Konfidenzintervall angegeben werden. Lautet die Fragestellung aber, ob der durchschnittliche Benzinpreis oberhalb einer bestimmten Marke liegt, so muss zu dem Mittelwert ein einseitiges Konfidenzintervall angegeben werden.

Bezogen auf die im Fuzzy-ROSA-Verfahren betrachteten geschätzten relativen Häufigkeiten $\hat{p}(C)$ und $\hat{p}(C|P)$ sind ohne Beachtung von Konfidenzintervallen zunächst drei grundsätzliche Fälle zu unterscheiden. Erstens kann $\hat{p}(C|P) > \hat{p}(C)$ gelten. Die Konklusion tritt also unter der Bedingung, dass die in der Prämisse spezifizierte Eingangssituation vorliegt, häufiger auf als in dem Fall, dass keine spezielle Eingangssituation vorliegt. Es gibt offenbar

einen kausalen Zusammenhang zwischen dem Vorliegen einer speziellen Eingangssituation und dem Auftreten einer bestimmten Ausgangssituation. Dieser Zusammenhang kann in der bekannten Form einer positiven Regel gemäß Gl. (3.5) formuliert werden. Zweitens kann $\hat{p}(C|P) = \hat{p}(C)$ gelten. Das Auftreten der Konklusion ist im unbedingten Fall und im bedingten Fall gleich häufig. Da sich kein Unterschied zeigt, gibt es keinen Zusammenhang zwischen Prämisse und Konklusion und es leitet sich folglich keine Regel daraus ab. Drittens kann $\hat{p}(C|P) < \hat{p}(C)$ gelten. Die Konklusion tritt also unter der Bedingung, dass die in der Prämisse spezifizierete Eingangssituation vorliegt, weniger häufig auf als in dem Fall, dass keine spezielle Eingangssituation vorliegt. Dieser Zusammenhang kann in Form einer negativen Regel gemäß Gl. (8.2) formuliert werden.

Aus den oben genannten Gründen ist es sinnvoll, bei dieser Untersuchung Konfidenzintervalle zu beachten. Abgeleitet aus den zuvor beschriebenen Fallunterscheidungen stellt sich die Frage, ob die entsprechenden Konfidenzintervalle disjunkt sind oder nicht. Aus diesem Grund werden einseitige Konfidenzintervalle verwendet, da für eine mögliche Überlappung zweier Konfidenzintervalle nur die obere Grenze des einen und die untere Grenze des anderen Konfidenzintervalls relevant sind. Für die unbedingte relative Häufigkeit ergeben sich die einseitigen Konfidenzintervalle

$$I_l^\alpha(C) = [0, \hat{p}_l^\alpha(C)] \quad \text{und} \quad I_u^\alpha(C) = [\hat{p}_u^\alpha(C), 1] \quad (6.3)$$

und für die bedingte relative Häufigkeit entsprechend die einseitigen Konfidenzintervalle

$$I_l^\alpha(C|P) = [0, \hat{p}_l^\alpha(C|P)] \quad \text{und} \quad I_u^\alpha(C|P) = [\hat{p}_u^\alpha(C|P), 1] \quad (6.4)$$

Übertragen auf den Fall, dass $\hat{p}(C|P) > \hat{p}(C)$ gilt, ist zu prüfen, ob $I_l^\alpha(C)$ und $I_u^\alpha(C|P)$ disjunkt sind. Falls dies zutrifft, kann eine positive Regel formuliert werden. Für den Fall, dass $\hat{p}(C|P) < \hat{p}(C)$ gilt, ist entsprechend zu prüfen, ob $I_u^\alpha(C)$ und $I_l^\alpha(C|P)$ disjunkt sind. Ist dies der Fall, kann eine negative Regel formuliert werden. Abbildung 6.1 veranschaulicht die unterschiedlichen Fälle.

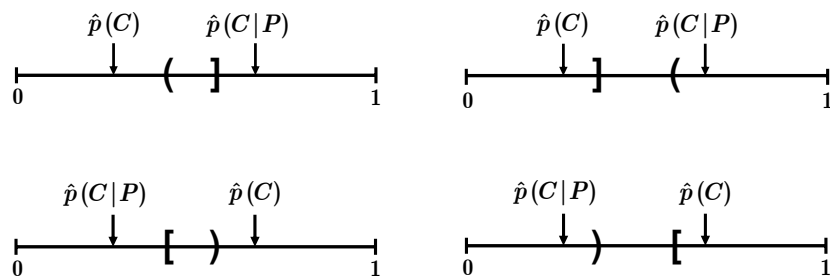


Abbildung 6.1: Beispiele für disjunkte bzw. nicht disjunkte Konfidenzintervalle: keine positive Regel (oben links), positive Regel (oben rechts), keine negative Regel (unten links) und negative Regel (unten rechts).

Im Fuzzy-ROSA-Verfahren stehen eine Reihe von Test- und Bewertungsverfahren zur Verfügung, die auf der Betrachtung der hier beschriebenen relativen Datenhäufigkeiten basieren [JS98b, Kro99, KT01]. Sie unterscheiden sich zum Teil aber darin, inwieweit auf Konfidenzintervalle sowohl für den Test einer Hypothese als auch für die gegebenenfalls vorzunehmende Bewertung einer Regel zurückgegriffen wird. Eine detaillierte Vorstellung dieser Test- und Bewertungsverfahren findet sich in Anhang A.

6.2 Test- und Bewertungsverfahren basierend auf Datenverteilungen

Im Gegensatz zu den in Abschnitt 6.1 bzw. in Anhang A vorgestellten Strategien orientieren sich diese Verfahren nicht an den relativen Häufigkeiten, sondern an den Verteilungen der Daten. Der Einsatz dieser Verfahren wird durch die Tatsache motiviert, dass die geschätzten relativen Häufigkeiten nach Gl. (6.1) und Gl. (6.2) abhängig sind von der zuvor vorgenommenen Partitionierung (Granularisierung) der Ausgangsgröße. Wenn eine Hypothese geprüft wird, kann es sein, dass diejenigen Datenpunkte, die die Prämisse erfüllen, eine vergleichsweise große Streuung im Hinblick auf ihre Ausgangsgrößenwerte aufweisen. Mit anderen Worten: Die Support-Datenpunkte einer Prämisse können mehrere linguistische Ausgangsgrößenwerte abdecken. In der Konklusion der Hypothese tritt jedoch nur ein einziger linguistischer Wert auf, ebenso bei der Schätzung der relativen Häufigkeiten. Bezogen auf den skizzierten Fall einer starken Streuung der Ausgangsgrößenwerte der Support-Datenpunkte kann dies dazu führen, dass die geschätzte bedingte relative Häufigkeit $\hat{p}(C|P)$ einen niedrigen Wert annimmt und daher keine positive Regel generiert wird. Eine vergleichsweise starke Streuung in den Ausgangsgrößenwerten der Support-Datenpunkte tritt oftmals bei gestörten bzw. widersprüchlichen Datensätzen auf oder auch, wenn sich die Zusammenhänge zwischen den Eingangsgrößen und der Ausgangsgröße nur schwach abzeichnen. Vor diesem Hintergrund sind die in den nachfolgenden Abschnitten vorgestellten Test- und Bewertungsverfahren entworfen worden, bei denen die Partitionierung der Ausgangsgröße nicht beachtet wird. Diese Verfahren generieren dabei aber ausschließlich positive Regeln. Negative Regeln können mit diesen Verfahren nicht erzeugt werden. Im Folgenden wird zunächst auf grundlegende Prinzipien und Begriffe des statistischen Testens eingegangen (siehe hierzu beispielsweise [Vog89]).

Statistisches Testen

Ein statistischer Test dient zur Überprüfung einer statistischen Hypothese und ihrer Signifikanz. Man kann prüfen, ob bestimmte Verhältnisse in Stichprobendaten (zum Beispiel Unterschiede in den Mittelwerten zweier Stichproben) auf Zufall zurückzuführen sind oder nicht. Man formuliert dabei zunächst eine Nullhypothese H_0 (Beispiel: *Mittelwerte unter-*

scheiden sich nicht) und eine Alternativhypothese H_1 (Beispiel: *Mittelwerte unterscheiden sich*). Aus den Stichproben wird eine Testgröße gebildet und der kritische Bereich für die Testgröße zu dem spezifizierten Signifikanzniveau festgelegt. Liegt die Testgröße nicht innerhalb des kritischen Bereichs, so wird die Nullhypothese beibehalten, andernfalls zugunsten der Alternativhypothese verworfen.

Dabei können zwei unterschiedliche Fehler auftreten. Als Fehler erster Art wird die Fehlentscheidung zugunsten der Alternativhypothese bezeichnet. Mit anderen Worten: Man verwirft die Nullhypothese, obwohl diese eigentlich zutrifft. Die Wahrscheinlichkeit, mit der man diesen Irrtum begeht, wird mit α bezeichnet. Als Fehler zweiter Art bezeichnet man die Fehlentscheidung zugunsten der Nullhypothese. Mit anderen Worten: Man verwirft die Nullhypothese nicht, obwohl diese eigentlich nicht zutrifft. Die Wahrscheinlichkeit, mit der man diesen Irrtum begeht, wird mit β bezeichnet. Bei vielen Testverfahren ist die Irrtumswahrscheinlichkeit α frei wählbar und das Vorliegen der Alternativhypothese ist durch das Testverfahren mit der Wahrscheinlichkeit $(1 - \alpha)$ abgesichert. Die Irrtumswahrscheinlichkeit β der Nullhypothese ist meist nicht bekannt und von verschiedenen Faktoren abhängig. Die Wahrscheinlichkeit $(1 - \beta)$ wird auch als Teststärke bezeichnet. Das primäre Ziel des statistischen Testens besteht darin, die Nullhypothese zugunsten der statistisch signifikanten Alternativhypothese zurückzuweisen. Will man also beispielsweise prüfen, ob sich die Mittelwerte zweier Stichproben unterscheiden, so stellt man die Nullhypothese *Mittelwerte unterscheiden sich nicht* auf und versucht, diese zugunsten der Alternativhypothese *Mittelwerte unterscheiden sich* zurückzuweisen.

Ein wichtiger Aspekt des statistischen Testens ist die Anzahl der zu beachtenden Stichproben. Im Ein-Stichprobenfall kann beispielsweise die Fragestellung von Interesse sein, ob der Mittelwert der Stichprobe oberhalb eines bestimmten Wertes liegt oder nicht. Ein Beispiel für einen solchen Fall ist die Frage, ob der Benzinpreis im Mittel oberhalb einer bestimmten Schwelle liegt. Im Zwei-Stichprobenfall (bzw. im allgemeinen k -Stichprobenfall) kann beispielsweise die Fragestellung von Interesse sein, ob sich die Mittelwerte der Stichproben signifikant voneinander unterscheiden oder nicht. Ein weiterer wichtiger Aspekt ist, ob im Zwei- und k -Stichprobenfall die Stichproben voneinander abhängig oder unabhängig sind. Voneinander abhängige Stichproben werden häufig auch als *verbundene* Stichproben bezeichnet. Ein Beispiel hierfür ist die Fragestellung, ob sich bei einer Gruppe von Patienten das Auftreten allergischer Reaktionen vor und nach einer Behandlung im Mittel signifikant unterscheidet.

Für die unterschiedlichen Fälle, die sich nach der Anzahl der zu beachtenden Stichproben und nach der Art der Abhängigkeiten zwischen den Stichproben ergeben können, sind unterschiedliche Testverfahren etabliert. Im Rahmen der Regelgenerierung innerhalb des Fuzzy-ROSA-Verfahrens handelt es sich um einen Zwei-Stichprobenfall: Die eine Stichprobe \mathcal{S}_A bilden diejenigen Datenpunkte bzw. deren Ausgangsgrößenwerte, die im unbedingten Fall vorliegen. Es handelt sich dabei folglich um alle Datenpunkte des zu Grunde liegenden Lerndatensatzes bzw. deren Ausgangsgrößenwerte. Die andere Stichprobe \mathcal{S}_B bilden dieje-

nigen Datenpunkte des Lerndatensatzes bzw. deren Ausgangsgrößenwerte, die im bedingten Fall auftreten, also die Prämisse der zu prüfenden Hypothese erfüllen. Die Stichproben sind voneinander unabhängig.

Mittelwertbasiertes Test- und Bewertungsverfahren

Das Mittelwertbasierte Test- und Bewertungsverfahren [JS98a, Jes99, Jes00b] (englisch: Mean Value Based) ist im Hinblick auf die Regelgenerierung für den Fall konzipiert worden, dass sich Zusammenhänge in den Daten vergleichsweise schwach abzeichnen. Das Verfahren verfolgt primär das Ziel, durch die generierten Regeln eine hohe Approximationsgüte hinsichtlich der zu Grunde liegenden Daten zu erreichen. Es werden zunächst für die unbedingte Verteilung der Mittelwert \bar{y} und die Standardabweichung σ der Ausgangsgrößenwerte berechnet:

$$\bar{y} = \frac{1}{D} \cdot \sum_{q=1}^D y_q \quad \text{und} \quad \sigma = \sqrt{\frac{\sum_{q=1}^D (y_q - \bar{y})^2}{D - 1}} \quad . \quad (6.5)$$

Gleiches wird für die bedingte Verteilung durchgeführt:

$$\bar{y}_P = \frac{1}{D_{sup}} \cdot \sum_{q=1}^D \mu_P(\mathbf{x}_q) \cdot y_q \quad \text{und} \quad \sigma_P = \sqrt{\frac{\sum_{q=1}^D \mu_P(\mathbf{x}_q) \cdot (y_q - \bar{y}_P)^2}{D_{sup} - 1}} \quad . \quad (6.6)$$

Mit Hilfe des Mittelwertbasierten Tests wird untersucht, ob sich die Mittelwerte der unbedingten und der bedingten Verteilung signifikant voneinander unterscheiden. Abbildung 6.2 veranschaulicht drei typische Szenarien, die bei dieser Fragestellung auftreten.

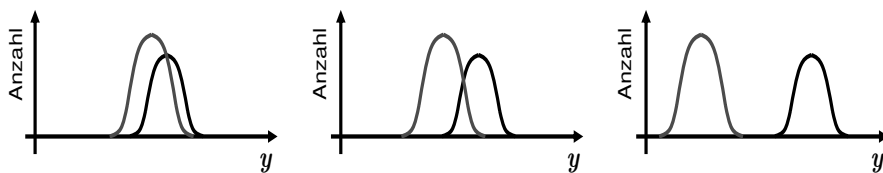


Abbildung 6.2: Beispiele für stark überlappende (links), wenig überlappende (Mitte) und nicht überlappende (rechts) Verteilungen.

Im ersten Fall (Abbildung 6.2, links) überlappen sich die unbedingte und die bedingte Verteilung so stark, dass offenbar kein signifikanter Unterschied in den Mittelwerten vorhanden

ist und daher keine Regel generiert werden sollte. Im dritten Fall (Abbildung 6.2, rechts) überlappen sich die Verteilungen gar nicht, und da offenbar ein signifikanter Unterschied vorliegt, sollte in diesem Fall eine Regel generiert werden. Häufig ist die Situation jedoch weniger eindeutig, wie hier im zweiten Fall dargestellt (Abbildung 6.2, Mitte).

Als zu überprüfende Nullhypothese H_0 wird angenommen, dass die Mittelwerte gleich sind. Die Alternativhypothese H_1 ist folglich, dass die Mittelwerte nicht gleich sind. Zur Überprüfung wird der statistische t -Test [Gos08, Har95, HEK95] eingesetzt. Der Wert t_n der Testgröße berechnet sich für die zu untersuchenden Verteilungen wie folgt:

$$t_n = \left| \frac{\bar{y}_P - \bar{y}}{\sigma_P / \sqrt{D_{sup}}} \right| . \quad (6.7)$$

Die Nullhypothese H_0 kann für eine gegebene Irrtumswahrscheinlichkeit α verworfen werden, falls $t_n > t(D_{sup}, \alpha)$ gilt. In diesem Fall unterscheiden sich die Mittelwerte signifikant voneinander und die Hypothese wird zur Regel erklärt. Der kritische Wert $t(D_{sup}, \alpha)$ kann dabei aus Tabellen entnommen oder geeignet approximiert werden. Der Bewertungsindex für eine mit diesem Verfahren generierte Regel wird folgendermaßen festgelegt:

$$J_{MVB} = e^{-\sigma_P / \sigma} . \quad (6.8)$$

Da sich die zu prüfende Hypothese (Regel) nicht auf einen bestimmten linguistischen Ausgangsgrößenwert bezieht, muss die Konklusion der generierten Regel geeignet gewählt werden. Für den Fall eines hier als gegeben angenommenen Fuzzy-Informationssystems kann der mittlere Ausgangsgrößenwert der bedingten Verteilung höchstens zu zwei linguistischen Ausgangsgrößenwerten in einem Grade größer Null zugehörig sein. Alle linguistischen Ausgangsgrößenwerte, denen \bar{y}_P nicht zugehörig ist, scheiden als Konklusion grundsätzlich aus. Im Fuzzy-ROSA-Verfahren wird so vorgegangen, dass die Konklusion der Regel so gewählt wird, dass für den Mittelwert \bar{y}_P der bedingten Verteilung der Grad der Zugehörigkeit maximal wird: $\mu(\bar{y}_P = S_j) \rightarrow \max$.

Varianzbasiertes Test- und Bewertungsverfahren

Das Varianzbasierte Test- und Bewertungsverfahren [Jes00a] (englisch: Variance Value Based) arbeitet ähnlich wie das Mittelwertbasierte Test- und Bewertungsverfahren. In diesem Fall wird jedoch geprüft, ob die Varianz der bedingten Verteilung die der unbedingten Verteilung signifikant unterschreitet. Dies wird mit Hilfe des statistischen F -Test [Har95, HEK95] geprüft. Der Wert der verwendeten Prüfgröße berechnet sich für dieses Verfahren zu:

$$F_n = \frac{\sigma_P^2}{\sigma^2} . \quad (6.9)$$

Falls $F_n < F_K(D, D_{sup}, \alpha)$ gilt, wird die Hypothese zur Regel erklärt. Dabei ist F_K der kritische Wert, für den die Nullhypothese einer nicht kleineren Varianz der bedingten Verteilung zum Konfidenzniveau α zugunsten der Alternativhypothese zurückzuweisen ist.

Dieser Wert kann Tabellen entnommen oder geeignet approximiert werden. Die Relevanz einer mit diesem Verfahren generierten Regel wird gemäß des Bewertungsindex J_{VVB} festgelegt:

$$J_{VVB} = 1 - \sigma_P / \sigma \quad . \quad (6.10)$$

Bezüglich der Festlegung der Konklusion einer Regel gelten dieselben Überlegungen wie bei dem Mittelwertbasierten Test- und Bewertungsverfahren. Es wird hier ebenfalls der Mittelwert der bedingten Verteilung für die Bestimmung des am besten geeigneten linguistischen Ausgangsgrößenwertes herangezogen.

6.3 Medianbasiertes Test- und Bewertungsverfahren

Die in Abschnitt 6.2 vorgestellten Test- und Bewertungsverfahren sind im Hinblick auf die Handhabung gestörter Daten bzw. die Aufdeckung nur schwach ausgeprägter Zusammenhänge in den Daten entwickelt worden. Ihr Vorteil gegenüber den in Abschnitt 6.1 beschriebenen Verfahren liegt darin, dass sie, abgesehen von der Festlegung der Konklusion einer Regel, unabhängig von der Partitionierung der Ausgangsgröße arbeiten und daher auch vergleichsweise schwach ausgeprägte Zusammenhänge aufdecken können. Sie sind daher auch weniger anfällig gegenüber einer ungünstig gewählten Partitionierung der Ausgangsgröße. Sowohl das Mittelwertbasierte Verfahren als auch das Varianzbasierte Verfahren haben jedoch auch gewisse Schwachstellen im Hinblick auf die Anwendung zur Regelgenerierung. Beide Ansätze verwenden parametrische Testverfahren, das heißt, es werden Hypothesen über die Parameter (Mittelwert und Standardabweichung) der Verteilungen geprüft. Die Verfahren gehen daher von der Annahme aus, dass die dem Test zu Grunde liegenden Daten näherungsweise normalverteilt sind und zu einer kontinuierlichen Größe erhoben wurden. Die Annahme normalverteilter Daten ist oftmals durchaus gerechtfertigt, insbesondere dann, wenn vergleichsweise viele Datenpunkte vorliegen. Gerade im Fall von nur wenigen Datenpunkten ist diese Annahme über die Verteilung allerdings kritisch zu sehen, denn die Verletzung dieser Annahme – durch Ausreißer innerhalb der Daten – hat Einschränkungen auf die Gültigkeit der Testergebnisse.

Nicht-parametrische Testverfahren machen keine Annahmen über die Verteilungen und werden daher häufig auch als *verteilungsfreie* Verfahren bezeichnet. Im Fall nominaler oder ordinaler Größen lassen sich Verteilungen nicht zutreffend durch Mittelwert und Standardabweichung charakterisieren. Für den Fall, dass zwei Stichproben zu einer kontinuierlichen Größe erfasst werden, keine speziellen Annahmen über die Verteilungen der Daten gemacht werden und voneinander unabhängige Stichproben vorliegen, ist der Wilcoxon-Rangsummen-Test [Wil45, Har95] konzipiert worden. Dieser ist äquivalent zu dem Mann-Whitney-U-Test. Mit Hilfe dieses Tests wird geprüft, ob sich die Mediane zweier Stichproben signifikant voneinander unterscheiden. Der Median wird auch als Zentralwert bezeichnet, da 50% der Werte einer Datenstichprobe unterhalb und 50% der Werte oberhalb des Me-

dians liegen. Hieraus leitet sich die im Rahmen dieser Arbeit verwendete Abkürzung CVB (englisch: Central Value Based) für dieses Verfahren ab.

Wie in Abschnitt 6.2 einleitend erläutert, handelt es sich im Rahmen der Regelgenerierung um einen Zwei-Stichprobenfall, bei dem die Stichproben \mathcal{S}_A und \mathcal{S}_B voneinander unabhängig sind. Die Stichprobe \mathcal{S}_A mit dem Umfang N_A beinhaltet die Ausgangsgrößenwerte aller Datenpunkte des zu Grunde liegenden Lerndatensatzes. Die Stichprobe \mathcal{S}_B mit dem Umfang $N_B \leq N_A$ umfasst die Ausgangsgrößenwerte derjenigen Datenpunkte, die die Prämisse der zu prüfenden Hypothese (Regel) erfüllen. In der Regel gilt $N_B < N_A$, da typischerweise nicht alle Datenpunkte des Lerndatensatzes die Prämisse der zu prüfenden Hypothese in einem Grade größer Null erfüllen. Die beiden Stichproben werden für den Test zunächst in einer gepoolten Probe \mathcal{S}_P mit dem Umfang $N_P = N_A + N_B$ zusammengefasst. Zu jedem Element der gepoolten Probe wird dabei anhand einer Markierung festgehalten, aus welcher der beiden Stichproben es stammt. Die Elemente der gepoolten Probe werden dann in aufsteigender Reihenfolge sortiert. Jedem Element der sortierten gepoolten Probe wird eine Rangnummer R zugewiesen, beginnend bei dem kleinsten Element mit $R = 1$ und kontinuierlich fortlaufend für alle weiteren Elemente bis hin zu $R = N_P$ für das größte Element. Danach ist zu prüfen, ob Werte innerhalb der sortierten gepoolten Probe mehrfach auftreten, ob also so genannte Verbundwerte (englisch: *ties*) vorliegen. Ist dies der Fall, so würden ein und demselben Wert mehrere unterschiedliche Rangnummern zugeordnet, was nicht plausibel ist. Solchen Verbundwerten wird dadurch begegnet, dass die Rangnummern über solche Verbundwerte aufsummiert und gemittelt werden und dann jedem entsprechenden Element die gemittelte Rangnummer zugewiesen wird. Aus der aufbereiteten sortierten gepoolten Probe werden die Rangsummen W_A und W_B über die Stichproben \mathcal{S}_A und \mathcal{S}_B gebildet:

$$W_A = \sum_{R_i \in \mathcal{S}_A} R_i \quad \text{und} \quad W_B = \sum_{R_j \in \mathcal{S}_B} R_j \quad , \quad (6.11)$$

deren Summe

$$W_A + W_B = \frac{N_P(N_P + 1)}{2} \quad (6.12)$$

gleich der Rangsumme über die gepoolte Probe \mathcal{S}_P ist. Wenn in der Stichprobe \mathcal{S}_B beispielsweise überwiegend niedrigere Werte als in der Stichprobe \mathcal{S}_A enthalten sind, so sollte sich dies darin widerspiegeln, dass in der Stichprobe \mathcal{S}_B insgesamt niedrigere Rangnummern vertreten sind als in Stichprobe \mathcal{S}_A . Mit anderen Worten: In der Stichprobe \mathcal{S}_B würde man eine niedrigere Summe über alle Rangnummern erwarten als in der Stichprobe \mathcal{S}_A . Wenn also die Nullhypothese H_0 : *Mediane sind gleich* gültig sein sollte, dann sollten die Rangsummen der beiden Stichproben (bzw. die mittleren Rangnummern) in etwa gleich sein. Dabei ist zu bedenken, dass die Stichproben unterschiedlich groß sein können.

Für jedes Element der kleineren Stichprobe \mathcal{S}_B kann bestimmt werden, wie viele Elemente in der Stichprobe \mathcal{S}_A eine höhere Rangnummer aufweisen als dieses Element. Die Summe

dieser Rangunterschreitungen über alle Elemente der Stichprobe \mathcal{S}_B ergibt sich zu:

$$U_B = (N_A \cdot N_B) + \frac{N_B(N_B + 1)}{2} - W_B \quad . \quad (6.13)$$

Gleiches kann für die Elemente der Stichprobe \mathcal{S}_A , bezogen auf die Elemente in der Stichprobe \mathcal{S}_B , durchgeführt werden:

$$U_A = (N_A \cdot N_B) + \frac{N_A(N_A + 1)}{2} - W_A \quad . \quad (6.14)$$

Man kann zeigen, dass

$$U_A + U_B = N_A \cdot N_B \quad (6.15)$$

gilt. Man bestimmt den kleineren der beiden Werte $U = \min(U_A, U_B)$ und verwendet diesen als Testgröße. Für kleine Stichprobengrößen kann die kritische Größe U_K für das geforderte Signifikanzniveau α aus Tabellen entnommen werden. Falls $U < U_K$ gilt, kann die Nullhypothese zugunsten der Alternativhypothese zurückgewiesen werden. Für insgesamt große Stichproben ist die Testgröße U näherungsweise normalverteilt und kann geeignet approximiert werden. Aus der verwendeten Testgröße leitet sich der Begriff U -Test ab.

Äquivalent zu dem Mann-Whitney- U -Test ist der Wilcoxon-Rangsummentest. Für insgesamt kleine Stichproben kann die Wahrscheinlichkeit $P(W_B)$, dass sich die Rangsumme W_B zufällig so ergibt, wie folgt exakt berechnet werden: Man stellt alle möglichen Kombinationen auf, N_B Elemente ohne Wiederholung aus den $N_P = (N_A + N_B)$ Elementen der gepoolten Probe zu ziehen. Die Anzahl der möglichen Kombinationen beträgt:

$$k = \frac{N_A!}{(N_A - N_B)!N_B!} \quad . \quad (6.16)$$

Mit anderen Worten: Man erzeugt aus der gepoolten Probe \mathcal{S}_P unterschiedliche neue Stichproben $\mathcal{S}_{\tilde{B}}$ mit jeweils N_B Elementen. Für jede der k Kombinationen wird die Rangsumme $W_{\tilde{B}}$ berechnet und mit dem Wert W_B verglichen. Die Anzahl der Kombinationen, für die $W_{\tilde{B}} < W_B$ gilt, wird mit k_L bezeichnet, die Anzahl der Kombinationen, für die $W_{\tilde{B}} \geq W_B$ gilt, wird entsprechend mit k_U bezeichnet. Die Wahrscheinlichkeit berechnet sich zu

$$P(W_B) = 2 \cdot \min\{k_L/k, k_U/k\} \quad . \quad (6.17)$$

Der Faktor 2 berücksichtigt, dass es sich um eine zweiseitige Fragestellung handelt. Für insgesamt große Stichproben nähert sich die Verteilung von W_B einer Normalverteilung mit folgendem Mittelwert μ und Standardabweichung σ :

$$\mu = \frac{N_B \cdot (N_P + 1)}{2} \quad \text{und} \quad \sigma = \frac{N_A \cdot N_B \cdot (N_P + 1)}{12} \quad . \quad (6.18)$$

Die Testgröße z ergibt sich durch Standardisierung von W_B zu

$$z = \frac{W_B - \mu}{\sigma} \quad . \quad (6.19)$$

Die Testgröße z ist normalverteilt mit dem Mittelwert $\mu = 0$ und der Standardabweichung $\sigma = 1$ (Standard-Normalverteilung). Die Wahrscheinlichkeit P , dass ein einzelner beobachteter Wert x einer (μ, σ) -normalverteilten Zufallsgröße X innerhalb des Intervalls $[-\infty, x]$ liegt, berechnet sich allgemein zu [BSMM99]:

$$P(X \leq x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt \quad . \quad (6.20)$$

Für den hier betrachteten Fall der Prüfgröße z und deren Standard-Normalverteilung kann die Wahrscheinlichkeit daher wie folgt berechnet werden:

$$P(Z \leq z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{t^2}{2}} dt \quad . \quad (6.21)$$

Für die Berechnung der Wahrscheinlichkeit existieren geeignete Approximationsfunktionen. Falls $2 \cdot P(Z \leq z) < \alpha$ gilt, kann die Nullhypothese zugunsten der Alternativhypothese zurückgewiesen werden. Die Mediane der beiden Stichproben unterscheiden sich signifikant voneinander und die zu prüfende Hypothese wird zu einer Regel erklärt.

Wie bei dem Mittelwertbasierten und dem Varianzbasierten Test- und Bewertungsverfahren muss die Konklusion einer mit dem Medianbasierten Verfahren generierten Regel geeignet gewählt werden. Hierfür wird vom Ansatz her im Prinzip wie bei den beiden genannten Verfahren vorgegangen, es wird jedoch nicht der Mittelwert, sondern der Median der bedingten Verteilung betrachtet. Aus den Ausgangsgrößenwerten der regelstützenden Datenpunkte wird der gewichtete Median $y_{\xi_{0.5P}}$ gebildet. Der Ausgangsgrößenwert y_i eines Datenpunktes erhält dabei den zugehörigen Erfülltheitsgrad der Prämisse als Gewicht $w_i = \mu_P(\mathbf{x}_i)$. Die Berechnung des gewichteten Median einer Datenstichprobe wird in Abschnitt 7.5 detailliert beschrieben.

Die Bewertung der Relevanz einer generierten Regel orientiert sich ebenfalls an dem Prinzip, das bei dem Mittelwertbasierten und dem Varianzbasierten Test- und Bewertungsverfahren angewandt wird. Bei den beiden genannten Verfahren wird die Relevanz einer Regel umso höher, je kleiner das Verhältnis σ_P/σ der bedingten zu der unbedingten Standardabweichung ist. Es wird bei dem Medianbasierten Verfahren jedoch nicht die Standardabweichung, sondern die *median absolute deviation* (MAD) ς bzw. ς_P als statistisches Maß für die Streuung herangezogen. Dabei werden über die Elemente der Stichprobe \mathcal{S}_A bzw. \mathcal{S}_B die einzelnen betraglichen Abweichungen zu dem Median der jeweiligen Stichprobe berechnet und in den Proben \mathcal{S}_{A2} bzw. \mathcal{S}_{B2} gespeichert. Für die Stichprobe \mathcal{S}_B wird die Abweichung dabei mit dem jeweiligen Erfülltheitsgrad der Regelprämisse gewichtet:

$$\mathcal{S}_{A2} = \{|y_{\xi_{0.5}} - y_j| \mid y_j \in \mathcal{S}_A\} \quad \text{und} \quad \mathcal{S}_{B2} = \{\mu_P(\mathbf{x}_j) \cdot |y_{\xi_{0.5P}} - y_j| \mid y_j \in \mathcal{S}_B\} \quad . \quad (6.22)$$

Die MAD ist der Median der betragsmäßigen Abweichungen:

$$\varsigma = \text{median}(\mathcal{S}_{A2}) \quad \text{und} \quad \varsigma_P = \text{median}(\mathcal{S}_{B2}) \quad . \quad (6.23)$$

Die Relevanz einer generierten Regel wird folgendermaßen festgelegt:

$$J_{CVB} = e^{-sP/s} \quad . \quad (6.24)$$

Die Relevanz einer Regel ist damit umso höher, je kleiner die MAD der bedingten Verteilung im Verhältnis zu der unbedingten Verteilung ist.

6.4 Regelgenerierung nach dem Konzept des Inkrementellen Relevanzindex

Die in den Abschnitten 6.1 und 6.2 vorgestellten Test- und Bewertungsverfahren weisen gewisse Schwachstellen im Hinblick auf die Erzeugung negativer Regeln auf [Kie97a]. Im Sinne einer möglichst gleichberechtigten Behandlung positiver und negativer Regeln ist von KIENDL der Inkrementelle Relevanzindex für die Regelgenerierung vorgeschlagen worden [Kie00, KKSS02b]. Hierbei handelt es sich nicht um ein Test- und Bewertungsverfahren, sondern hierunter ist vielmehr ein spezielles Konzept für die Regelgenerierung zu verstehen.

Die Regelsuche verläuft hierbei grundsätzlich so wie in anderen Fällen auch, jedoch werden zusätzlich zu den anhand eines Test- und Bewertungsverfahrens generierten Regeln noch weitere Regeln der Regelbasis hinzugefügt. Das Konzept des Inkrementellen Relevanzindex basiert auf der Verwendung von Regeln, die sich in ihrer Bedeutung in drei Kategorien aufteilen lassen: universelle Regeln, herkömmliche Regeln und komplementäre Regeln. Universelle Regeln haben folgende Form:

$$\text{WENN } \langle I=1 \rangle \text{ DANN } \langle C \rangle \quad . \quad (6.25)$$

Da die Prämisse einer solchen Regel stets erfüllt ist, wird eine solche Regel unabhängig von dem Eingangsgrößenvektor \mathbf{x} immer aktiviert. Zu jedem der S linguistischen Ausgangsgrößenwerte wird genau eine universelle Regel erzeugt. Das Glaubensmaß einer universellen Regel richtet sich primär nach zwei Aspekten, zum einen danach, welche Fläche F_i von der Zugehörigkeitsfunktion des linguistischen Wertes eingeschlossen wird, auf den sich die Konklusion der universellen Regel bezieht, zum anderen danach, wie oft der linguistische Wert in der Konklusion der universellen Regel durch die zu Grunde liegenden Lerndatenpunkte aktiviert wird. Die von einer Zugehörigkeitsfunktion eingeschlossene Fläche wird mit F_i bezeichnet. Die Summe der Teilflächen aller Zugehörigkeitsfunktionen ist $F_G = \sum F_i$. Die Aktivierung der Konklusion einer universellen Regel durch die Lerndatenpunkte ist gleich der relativen unbedingten Häufigkeit $\hat{p}(y = S_i)$. Ferner wird ein globaler Parameter $0 \leq \lambda \leq 1$ eingeführt, über den gesteuert wird, wie stark universelle Regeln grundsätzlich zu berücksichtigen sind. Es wird zunächst ein vorübergehendes Glaubensmaß für eine universelle Regel wie folgt festgelegt:

$$\tilde{\rho}_i = \hat{p}(y = S_i) \frac{F_i}{F_G} \quad . \quad (6.26)$$

Eine universelle Regel wird als umso relevanter eingestuft, je stärker ihre Konklusion durch die Lerndatenpunkte aktiviert wird und je größer die eingeschlossene Fläche unter der entsprechenden Zugehörigkeitsfunktion ist. Es wird für die endgültigen Glaubensmaße der universellen Regeln ferner gefordert, dass $\sum \rho_i = \lambda$ gilt. Um dies zu gewährleisten, wird ein Normierungsfaktor q eingeführt, der sich zu $q = 1 / \sum \tilde{\rho}_i$ ergibt. Damit ergibt sich das endgültige Glaubensmaß einer universellen Regel zu

$$\rho_i = \lambda q \tilde{\rho}_i \quad . \quad (6.27)$$

Das Verhältnis der Glaubensmaße der universellen Regeln zueinander ist damit gerade so gewählt, dass sich für den Fall, dass ausschließlich die universellen Regeln aktiviert sind, bei der Defuzzifizierung der so erzeugten ausgangsseitigen Zugehörigkeitsfunktion¹ nach der Schwerpunkt-Methode der mittlere Ausgangsgrößenwert der Lerndatenpunkte einstellt. Da die universellen Regeln immer aktiviert sind, ist dieser Wert gleichzeitig auch ein Basis-Ausgabewert des Mamdani-Systems. Alle weiteren Regeln, ob positiv oder negativ, wirken inkrementell und erzeugen Abweichungen von diesem Basis-Ausgabewert, was bei deren Generierung und Bewertung zu beachten ist.

Bei den herkömmlichen Regeln handelt es sich um diejenigen Regeln, die im Zuge der Regelsuche anhand eines Test- und Bewertungsverfahrens als relevante Regeln generiert werden. Hierbei kann es sich sowohl um positive als auch um negative Regeln handeln. Im Prinzip kann jedes der in Abschnitt 6.2 und in Anhang A vorgestellten Test- und Bewertungsverfahren für die Regelgenerierung eingesetzt werden. Speziell für den Einsatz in Verbindung mit dem Konzept des Inkrementellen Relevanzindex zur Regelgenerierung sind aber die inkrementelle Trefferquote J_{ITQ} und die konfidente inkrementelle Trefferquote J_{KITQ} konzipiert worden (siehe Anhang A). Zu jeder erzeugten herkömmlichen Regel werden zusätzlich $(S - 1)$ komplementäre Regeln erzeugt, das heißt, zu einer herkömmlichen positiven Regel entsprechend negative Regeln bzw. im Fall einer herkömmlichen negativen Regel dann entsprechend positive Regeln. Die Relevanzgrade $\rho_{i,j}$ der komplementären Regeln werden dabei so festgelegt, dass zum einen $\rho_i = \sum_{j=1}^{S-1} \rho_{i,j}$ gilt und zum anderen die Verhältnisse der Relevanzgrade $\rho_{i,j}$ zueinander bzw. zu ρ_i denen entsprechen, die sich für die universellen Regeln für $\lambda = 1$ nach Gl. (6.27) ergeben. Aufgrund dieses speziellen Prinzips der Regelgenerierung ist die Anzahl der insgesamt erzeugten Regeln meist vergleichsweise hoch.

Es sei an dieser Stelle angemerkt, dass die Einführung des Basis-Ausgabewertes den Vorteil mit sich bringt, dass nur noch in Ausnahmefällen auf den Default-Wert zurückgegriffen werden muss und sich so eine höhere Flexibilität ergibt. Zum einen muss in dem Fall auf den Default-Wert zurückgegriffen werden, wenn in der Regelbasis keine negativen Regeln vorhanden sind und keine positiven Regeln durch den Eingangsgrößenvektor \mathbf{x} aktiviert werden und $\lambda = 0$ ist. In diesem Fall sind die universellen Regeln deaktiviert und es gibt keinerlei andere Empfehlungen durch positive Regeln. Zum anderen kann dies für den Fall

¹Als ODER-Operator werde dabei die algebraische Summe verwendet.

gelten, dass in der Regelbasis negative Regeln vorhanden sind und $\lambda > 0$ ist. In diesem Fall sind die universellen Regeln zwar aktiviert, jedoch wird der Basis-Ausgabewert durch die negativen Regeln verboten². Der Vorteil des Basis-Ausgabewertes ist, dass er sich problemabhängig einstellt. Dies lässt sich im Prinzip zwar auch problemlos auf den Default-Wert übertragen, aber die Verwendung zweier Werte kann gewisse Vorteile haben. Zieht man das Beispiel eines chemischen Prozesses heran, so kann eine mögliche Strategie zur Prozesskontrolle darin bestehen, eine konstante Energiezufuhr (Basis-Wert) vorzunehmen, solange keine Regeln eine Abweichung hiervon empfehlen. Gerät der Prozess, aus welchen Gründen auch immer, aber in einen kritischen Zustand, so sollte die Energiezufuhr durch Sicherheitsmaßnahmen (negative Regeln) ausgesetzt werden. Die Unterbrechung der Energiezufuhr kann über den Default-Wert Null realisiert werden.

²Dies ist auch auf den Fall übertragbar, dass zusätzlich zu den universellen Regeln noch positive Regeln wirken, aber aufgrund der negativen Regeln das resultierende ausgangsseitige Empfehlungsgebirge leer ist.

Kapitel 7

Neue Strategieelemente für Takagi-Sugeno-Kang-Systeme

Wie in Abschnitt 4.3 skizziert, kann ein mit dem Fuzzy-ROSA-Verfahren generiertes Mamdani-Modell nachgelagert in ein korrespondierendes TSK-Modell transformiert werden. Ziel der Transformation ist es, die Approximationsgenauigkeit des Modells signifikant zu erhöhen und dabei dessen Interpretierbarkeit zu erhalten. In Abschnitt 7.1 wird die prinzipielle Vorgehensweise bei der Transformation genauer erläutert. Die bisher eingesetzten Strategieelemente [Kra02] für die Transformation von Mamdani-Regeln in TSK-Regeln weisen gewisse Schwachstellen im Hinblick auf zwei Aspekte auf, die in industriellen Anwendungen von großer Bedeutung sind. Dies betrifft zum einen den Fall, dass vergleichsweise wenige Datenpunkte und vergleichsweise viele Eingangsgrößen vorliegen. Zum anderen ist es der Fall, dass innerhalb des zu Grunde liegenden Datensatzes unter Umständen fragwürdige Daten enthalten sind. In Abschnitt 7.2 werden diese Aspekte detaillierter diskutiert. Zur Überwindung der genannten Schwachstellen wird das Fuzzy-ROSA-Verfahren um neue Strategieelemente [SKK04, SKKS05] erweitert, die in den Abschnitten 7.3 und 7.4 bzw. in Abschnitt 7.5 vorgestellt werden.

7.1 Transformation einer Mamdani-Regelbasis in eine TSK-Regelbasis

Um eine mit dem Fuzzy-ROSA-Verfahren erzeugte Mamdani-Regelbasis in eine korrespondierende TSK-Regelbasis zu transformieren, wird jede Mamdani-Regel individuell in eine entsprechende TSK-Regel umgewandelt. Die Transformation einer Mamdani-Regel in eine TSK-Regel gliedert sich in zwei wesentliche Teilschritte: Die Auswahl der zu beachtenden Lerndatenpunkte und die darauf aufsetzende Bestimmung der freien Parameter in der

Konklusion der TSK-Regel. Diese Teilschritte werden in den nachfolgenden Abschnitten genauer beschrieben.

Einschränkung der Transformation auf positive Mamdani-Regeln

Wie in Abschnitt 4.3 erwähnt, unterstützt das Fuzzy-ROSA-Verfahren die Verwendung negativer Mamdani-Regeln. Eine Transformation von Regeln dieser Art in korrespondierende TSK-Regeln ist jedoch problematisch und nicht unmittelbar möglich. Weitergehende Ausführungen hierzu folgen in Abschnitt 8.2. Die nachfolgenden Abschnitte beziehen sich daher ausschließlich auf die Transformation positiver Mamdani-Regeln in positive TSK-Regeln.

Ansätze für die Konklusion einer TSK-Regel

Für die Konklusion $f(\mathbf{x})$ einer TSK-Regel kommen prinzipiell beliebig komplexe Funktionsansätze in Frage. Es können grundsätzlich zwei verschiedene Klassen von Ansätzen unterschieden werden: solche Ansätze, in denen die freien Parameter linear in $f(\mathbf{x})$ eingehen, und solche Ansätze, in denen diese nichtlinear eingehen. Der zentrale Unterschied zwischen den beiden Klassen liegt in dem Aufwand, der mit der Bestimmung der freien Parameter verbunden ist. Im Fall linearer Abhängigkeiten lassen sich die Parameter vergleichsweise einfach bestimmen, wohingegen im Fall nichtlinearer Abhängigkeiten aufwendigere Verfahren eingesetzt werden müssen. Derartige Verfahren arbeiten in der Regel ausgehend von einer Initiallösung iterativ, bis das Verfahren konvergiert oder abbricht. Der Aufwand für die Bestimmung der Parameter ist entsprechend hoch. Ferner ergeben sich in diesem Fall sehr viele Entwurfsfreiheiten für den Ansatz der Konklusion. Diese Freiheiten erlauben zwar prinzipiell eine noch stärkere Verbesserung der Approximationsgenauigkeit, andererseits besteht aber auch die Gefahr, dass sich durch einen ungünstig gewählten Ansatz diese Freiheiten zum Nachteil auswirken. Daher werden im Rahmen dieser Arbeit ausschließlich lineare Abhängigkeiten in der Konklusion $f(\mathbf{x})$ betrachtet. Dadurch wird zwar an Flexibilität eingebüßt, aber die Gefahr einer Fehleinstellung kleingehalten.

Aus der Literatur sind das *zero-order* TSK-Modell und das *first-order* TSK-Modell als gängige Ansätze bekannt. Bei dem ersten ist $f(\mathbf{x}) = a_0$, also ein konstanter Ausgangsgrößenwert. Bei dem zweiten ist $f(\mathbf{x}) = a_0 + \sum_{j=1}^n a_j \cdot x_j$, also eine lineare Funktion der n Eingangsgrößen. Verallgemeinert ist die Konklusion $f(\mathbf{x}) = \sum_{j=0}^m a_j \cdot g_j(\mathbf{x})$ eine Linearkombination verschiedener Elementarfunktionen $g_j(\mathbf{x})$, wobei diese sowohl lineare als auch nichtlineare Funktionen der Eingangsgrößen sein können. Aus dieser allgemeinen Formulierung leiten sich entsprechend weitere Ansätze für $f(\mathbf{x})$ ab, beispielsweise basierend auf Polynomen oder eine multilineare Funktion [Kra02].

Bei der Wahl des Ansatzes für die Konklusion sind verschiedene grundlegende Aspekte zu bedenken. Zum einen ist die Gefahr einer Überanpassung (Overfitting) zu berücksichtigen.

Bei sehr komplexen Funktionsansätzen besteht die Gefahr, dass die zu Grunde liegenden Datenpunkte des Lerndatensatzes zwar sehr gut approximiert werden, jedoch die Generalisierungsfähigkeit der einzelnen Regeln und damit letztlich auch die des gesamten Modells verloren geht. Die Folge kann sein, dass das Modell auf bis dahin unbekanntem Daten (Validierungsdaten) sehr fragwürdige Prognosen liefert. Der zweite Aspekt betrifft die Anzahl der Parameter in der Konklusion sowie die damit verbundene Frage, wie diese bestimmt werden können. Bei einem zero-order TSK-Modell muss unabhängig von der Anzahl n der Eingangsgrößen stets nur ein einziger Parameter a_0 bestimmt werden, bei einem first-order TSK-Modell sind es $(n + 1)$ Parameter. Wählt man einen komplexeren Ansatz für $f(\mathbf{x})$, so steigt die Anzahl der Parameter weiter an. Bei einer multilinearen Funktion sind beispielsweise $(2^n - 1)$ freie Parameter zu bestimmen. Es ist dabei zu bedenken, dass sich mit zunehmender Dimension n des Modellierungsproblems in vielen praktischen Anwendungen zwei gegenläufige Tendenzen ausbilden. Einerseits steigt die Anzahl freier Parameter in der Konklusion an, andererseits wird eine Regel tendenziell von vergleichsweise wenigen Datenpunkten gestützt, da der Raum der Eingangsgrößen in hochdimensionalen Fällen meist nur schwach durch die vorhandenen Datenpunkte abgedeckt wird. Die spärliche Abdeckung des Raumes durch Datenpunkte wird oft auch als *Curse Of Dimensionality* (COD-Problematik) bezeichnet¹.

Es sei an dieser Stelle angemerkt, dass es natürlich grundsätzlich auch möglich ist, für jede TSK-Regel individuell einen bestimmten Ansatz für die Konklusion zu wählen. Der Ansatz wird im Rahmen dieser Arbeit zwar nicht untersucht, er kann für zukünftige Arbeiten jedoch von Interesse sein. Die Transformation kann dabei in der Art geschehen, dass vorab eine handhabbare Menge an Standard-Konklusionen² definiert wird und für die aktuell zu transformierende Regel anhand charakteristischer Merkmale der regelstützenden Datenpunkte die hierfür am besten geeignete Standard-Konklusion ausgewählt und die darin enthaltenen freien Parameter bestimmt werden. Die Auswahl der am besten geeigneten Standard-Konklusion darf sich dabei jedoch nicht allein nach der jeweils erzielten Approximationsgenauigkeit richten, da sonst sehr wahrscheinlich tendenziell komplexere Funktionsansätze für die Konklusionen ausgewählt werden, die eine entsprechend hohe Approximationsgenauigkeit erzielen. Damit ist die Gesamtanzahl an Parametern entsprechend groß und das Modell tendiert zu einer Überanpassung. Es muss vielmehr eine multi-kriterielle Bewertung erfolgen, die beispielsweise einen Kompromiss zwischen der erzielten Approximationsgenauigkeit und der Anzahl der hierfür benötigten Parameter anstrebt. Der Vorteil dieser Vorgehensweise liegt in der großen Flexibilität, da der Aufwand (also letztlich die Anzahl freier Parameter in der Konklusion) bedarfsorientiert gewählt werden kann. Außerdem lassen sich auch solche Standard-Konklusionen vorsehen, die aus nicht-linearen Elementarfunktionen (zum Beispiel Exponentialfunktionen) aufgebaut sind und daher für die Modellierung entsprechend gearteter Zusammenhänge grundsätzlich besser geeignet sind als zum Beispiel eine Hyperebene. Bei *einfach zu modellierenden Zusam-*

¹Der Begriff geht auf BELLMAN [Bel61] zurück, der sich mit der Schätzung von Wahrscheinlichkeitsdichten in hochdimensionalen Räumen anhand von Stichproben beschäftigte.

²Mit anderen Worten: Verschiedene Schablonen oder Vorlagen (englisch: Templates).

menhängen wählt man einfachere Ansätze bzw. weniger Parameter in der Konklusion, bei *schwer zu modellierenden Zusammenhängen* entsprechend komplexere Ansätze bzw. mehr Parameter in der Konklusion. In Summe über alle Regeln ist der Gesamtaufwand (also die Gesamtanzahl an Parametern) daher nicht zwangsläufig größer als in dem Fall, dass alle Regeln den gleichen Ansatz für die Konklusion verwenden. Die Nachteile liegen zum einen in dem erhöhten Aufwand (die Analyse der regelstützenden Datenpunkte) und zum anderen in der geringeren Transparenz des resultierenden TSK-Modells (die Konklusionen der unterschiedlichen Regeln sind nur eingeschränkt vergleichbar).

Transformation generalisierender Mamdani-Regeln

Wie in Abschnitt 4.3 erläutert, unterstützt das Fuzzy-ROSA-Verfahren generalisierende Regeln. Im Gegensatz zu einer vollständigen Regel kommen in der Prämisse einer generalisierenden Regel nicht alle zu beachtenden Eingangsgrößen vor. Daher könnte für die Transformation von Regeln dieser Art auch der Ansatz verfolgt werden, die Parameter derjenigen Eingangsgrößen, die nicht in der Prämisse enthalten sind, fest auf $a_j = 0$ zu setzen. Auf diese Weise würde sich die Anzahl der freien Parameter reduzieren, was insbesondere dann von Vorteil sein könnte, wenn nur vergleichsweise wenige Lerndatenpunkte vorhanden sind. Vorangegangene Untersuchungen [Kra02] haben jedoch gezeigt, dass auch für generalisierende Regeln stets alle Eingangsgrößen beachtet werden sollten, da auch in den nicht in der Prämisse enthaltenen Eingangsgrößen oftmals Trends in den Daten vorhanden sind, deren Berücksichtigung zu einer Verbesserung der Approximationsgenauigkeit führt.

Auswahl der zu beachtenden Lerndatenpunkte

Nach geeigneter Wahl des Ansatzes für $f(\mathbf{x})$ müssen für jede Mamdani-Regel die freien Parameter in der Konklusion $f(\mathbf{x})$ der korrespondierenden TSK-Regel bestimmt werden. Dies geschieht auf Grundlage derjenigen Datenpunkte (oder einer Teilmenge hiervon) des zu Grunde liegenden Lerndatensatzes, die die aktuell zu transformierende Mamdani-Regel stützen, also deren Prämisse in einem Grade größer Null erfüllen. Die Auswahl der zu beachtenden Lerndatenpunkte soll möglichst gewährleisten, dass die Konklusion der zu Grunde liegenden Mamdani-Regel und die Konklusion der so erzeugten TSK-Regel ineinander überführbar bleiben.

Der zu Grunde liegende Lerndatensatz $\mathcal{D}_L = \{\mathbf{d}_1, \dots, \mathbf{d}_D\}$ bestehe aus D Datenpunkten \mathbf{d}_i , wobei ein Datenpunkt durch $\mathbf{d}_i = (x_{i,1}, \dots, x_{i,n}, y_{i,1})$ beschrieben wird. Von den D Datenpunkten erfüllen D_P Datenpunkte die Prämisse P der zu transformierenden Regel jeweils im Grade $\mu_P(\mathbf{x}) > 0$. In vielen Fällen werden jedoch nicht alle D_P Datenpunkte auch die Konklusion C der Regel im Grade $\mu_{Y_C}(y) > 0$ aktivieren. Für die Anzahl der Datenpunkte, die sowohl die Prämisse der Regel erfüllen als auch die spezifizierte Konklusion

aktivieren, gilt somit $D_{P \wedge C} \leq D_P$. Bei der Transformation einer Mamdani-Regel in eine TSK-Regel sollten in der Regel nur die $D_{P \wedge C}$ Datenpunkte für die Bestimmung der freien Parameter in $f(\mathbf{x})$ verwendet werden, damit die Gefahr einer ungewollten Verschiebung der resultierenden Funktion aus dem Wertebereich der Konklusion C möglichst gering ist [Kra02]. Man geht davon aus, dass gilt:

$$\mu_P(\mathbf{x}) > 0 \Rightarrow \mu_{Y_C}(f(\mathbf{x})) > 0 \quad . \quad (7.1)$$

Mit anderen Worten: Wenn ein Eingangsgrößenvektor \mathbf{x} die Prämisse der Regel in einem Grade größer Null erfüllt, so sollte der zu diesem Eingangsgrößenvektor durch die TSK-Regel modellierte bzw. prognostizierte Ausgangsgrößenwert $\hat{y} = f(\mathbf{x})$ die Konklusion der zu Grunde liegenden Mamdani-Regel ebenfalls in einem Grade größer Null aktivieren, also innerhalb des Wertebereichs des linguistischen Wertes liegen, auf den sich die Elementaraussage in der Konklusion der Mamdani-Regel bezieht. In diesem Fall bleiben die Mamdani-Regel und die TSK-Regel ineinander überführbar. Dieser Aspekt ist insbesondere im Hinblick auf die Prognose für solche Daten zu bedenken, die nicht in den Lerndaten enthalten sind (Validierungsdaten). Wie das folgende einfache Beispiel mit nur einer Eingangsgröße illustriert, kann aber auch bei einer entsprechenden Auswahl der Lerndatenpunkte für die Transformation einer Regel unter Umständen eine unerwünschte Verschiebung aus dem angestrebten Wertebereich auftreten.

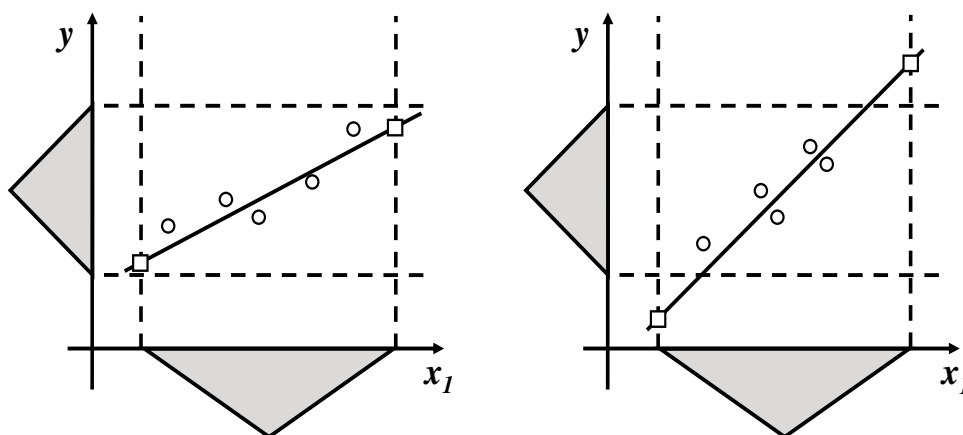


Abbildung 7.1: Beispiele zur Extrapolation aus dem angestrebten Wertebereich bei TSK-Regeln: keine Verschiebung (links) und eventuell ungünstige Verschiebung (rechts).

Abbildung 7.1 zeigt zwei unterschiedliche Szenarien für die Transformation einer Mamdani-Regel in eine TSK-Regel. Dabei sind die regelstützenden Datenpunkte des Lerndatensatzes jeweils durch Kreise und die dreieckförmigen Zugehörigkeitsfunktionen für die in diesem Beispiel relevanten linguistischen Werte sowohl der Eingangs- als auch der Ausgangsgröße

grau dargestellt. Die freien Parameter in der Konklusion $f_i(\mathbf{x}) = a_0 + a_1 \cdot x_1$ der TSK-Regel werden nach dem Least-Squares-Ansatz (siehe Abschnitt 7.2.2) bestimmt und die resultierende Ausgleichsgerade ist jeweils als durchgezogene Linie dargestellt. Die oben erwähnten Validierungsdaten bzw. deren Eingangsgrößenwerte und die zugehörigen prognostizierten Ausgangsgrößenwerte sind in den Grafiken in Abbildung 7.1 beispielhaft durch Quadrate dargestellt. Im ersten Fall liegen die von $f_i(\mathbf{x})$ zu den Validierungsdaten prognostizierten Ausgangsgrößenwerte innerhalb des angestrebten Wertebereichs, im zweiten Fall liegen sie außerhalb. Die Empfehlung der TSK-Regel deckt sich dann in einigen Fällen also nicht mehr mit der der zu Grunde liegenden Mamdani-Regel. Diese unter Umständen ungewollten Verschiebungen lassen sich jedoch leicht dadurch vermeiden, dass die Ausgabe der Konklusion einer TSK-Regel auf einen entsprechenden Wertebereich begrenzt wird. Diese Begrenzung wird im Fuzzy-ROSA-Verfahren standardmäßig vorgenommen.

Bestimmung der freien Parameter in der Konklusion einer TSK-Regel

Im Fuzzy-ROSA-Verfahren wird die Einschränkung getroffen, dass die Konklusion einer TSK-Regel eine Funktion $f(\mathbf{x})$ sein muss, die linear in den Parameter a_j ist. Es ist also $f(\mathbf{x})$ allgemein eine Linearkombination verschiedener Elementarfunktionen $g_j(\mathbf{x})$ und kann als $f(\mathbf{x}) = \sum_{j=0}^m a_j \cdot g_j(\mathbf{x})$ geschrieben werden. Dabei können die einzelnen Elementarfunktionen sowohl lineare als auch nichtlineare Funktionen der Eingangsgrößen sein. Mit $\mathbf{g}^T = (g_0(\mathbf{x}), \dots, g_m(\mathbf{x}))$ und $\mathbf{a}^T = (a_0, \dots, a_m)$ kann auch in vektorieller Form $f(\mathbf{x}) = \mathbf{g}^T \mathbf{a}$ geschrieben werden. Für einen gegebenen Datenpunkt $\mathbf{d}_j = (\mathbf{x}_j, y_j)$ wird der für diesen Eingangsgrößenvektor \mathbf{x}_j gemessene Ausgangsgrößenwert y_j durch $f(\mathbf{x}_j) = \hat{y}_j$ modelliert, wobei der Approximationsfehler $e_j = y_j - \hat{y}_j$ möglichst gering sein sollte. Dies ist der bekannte Ansatz für ein lineares Modell

$$\mathbf{y} = \mathbf{G} \cdot \mathbf{a} + \mathbf{e} \quad , \quad (7.2)$$

das den wahren, aber unbekanntem Zusammenhang zwischen den Eingangsgrößen und der Ausgangsgröße nachbildet. Auf Grundlage der ausgewählten Datenpunkte, die die zu transformierende Regel stützen, wird das lineare Gleichungssystem

$$\mathbf{G} \cdot \mathbf{a} = \mathbf{y} \quad (7.3)$$

aufgestellt, das das zu lösende Problem für die Transformation einer Mamdani-Regel in eine TSK-Regel beschreibt. In Abhängigkeit davon, ob es sich um ein eindeutig lösbares, oder um ein unter- oder überbestimmtes lineares Gleichungssystem handelt, kommen unterschiedliche Lösungsstrategien für die Bestimmung einer speziellen Lösung zum Einsatz.

7.2 Bisher angewandte Strategien zur Parameterbestimmung

Das lineare Gleichungssystem nach Gl. (7.3) kann unterbestimmt oder eindeutig lösbar oder überbestimmt sein. Der Fall eines eindeutig lösbaren linearen Gleichungssystems ist unkritisch. Die Lösung berechnet sich in diesem Fall im Prinzip einfach zu

$$\mathbf{a} = \mathbf{G}^{-1} \cdot \mathbf{y} \quad , \quad (7.4)$$

wobei die Berechnung auch in einer anderen Form als durch Invertieren der Matrix \mathbf{G} erfolgen kann. Die beiden anderen Fällen sind hingegen komplexer, da sich in diesen Fällen verschiedene Optionen für das weitere Vorgehen anbieten. Nachfolgend werden diese beiden Fälle genauer diskutiert. Dabei wird zunächst die jeweilige Strategie vorgestellt, die zu Beginn dieser Arbeit im Fuzzy-ROSA-Verfahren für die Behandlung eines solchen Falles angewandt wurde. Daran anschließend werden jeweils mögliche Alternativen hierzu diskutiert. Die letztlich realisierten neuen Lösungsstrategien werden in den Abschnitten 7.3 und 7.4 vorgestellt.

Im Fuzzy-ROSA-Verfahren wird standardmäßig der Ansatz $f(\mathbf{x}) = a_0 + \sum_{j=1}^n a_j \cdot x_j$ für die Konklusion einer TSK-Regel verwendet. Es ergeben sich entsprechend $\mathbf{g}^T = (1, x_1, \dots, x_n)$ und $\mathbf{a}^T = (a_0, a_1, \dots, a_n)$. Dies ist primär für den Fall eines unterbestimmten linearen Gleichungssystems von Bedeutung, wie in Abschnitt 7.2.1 deutlich wird.

7.2.1 Unterbestimmte lineare Gleichungssysteme

Das lineare Gleichungssystem $\mathbf{G} \cdot \mathbf{a} = \mathbf{y}$ ist unterbestimmt, wenn die $m \times n$ Matrix \mathbf{G} mehr Spalten als Zeilen aufweist ($m < n$) oder die Matrix \mathbf{G} mindestens so viele Zeilen wie Spalten aufweist ($m \geq n$), aber nicht den vollen Rang hat ($\text{rang}(\mathbf{G}) < m$). Das lineare Gleichungssystem hat in diesem Fall keine eindeutige Lösung. Der Fall eines unterbestimmten linearen Gleichungssystems tritt bei der Transformation einer Mamdani-Regel in eine TSK-Regel typischerweise in zwei Fällen auf. Zum einen, wenn der Lerndatensatz insgesamt nur sehr wenige Datenpunkte beinhaltet. Zum anderen, wenn die Anzahl der zu bestimmenden Parameter in der Konklusion der TSK-Regel aufgrund der Anzahl an Eingangsgrößen sehr groß ist. Insbesondere der zweite Fall ist zu bedenken, da mit zunehmender Dimension des Problems nicht nur mehr Parameter bestimmt werden müssen, sondern der von den Eingangsgrößen aufgespannte Raum tendenziell immer schwächer durch Datenpunkte abgedeckt wird. Eine zu transformierende Regel wird also tendenziell durch eher wenige Datenpunkte gestützt, gleichzeitig müssen aber vergleichsweise viele Parameter bestimmt werden. Diese zwei gegenläufigen Tendenzen können dazu führen, dass gerade bei hochdimensionalen Problemen verstärkt der Fall eines unterbestimmten linearen Gleichungssystems auftritt.

Im Fall eines unterbestimmten linearen Gleichungssystems ist bisher auf eine einfach und schnell zu berechnende Mittelwert-Prognose zurückgegriffen worden. Dabei werden $a_0 = \bar{y}_P$ und $a_1 = \dots = a_n = 0$ gesetzt, wobei \bar{y}_P der mittlere Ausgangsgrößenwert der regelstützenden Lerndatenpunkte ist. Dies entspricht in der Wirkung einem zero-order TSK-Modell, also einem lokal konstanten Modell. Es ist ersichtlich, dass so Freiheitsgrade gegenüber dem angestrebten first-order TSK-Modell verloren gehen, wodurch die Genauigkeit der Approximation entsprechend eingeschränkt ist. Außerdem ist zu bedenken, dass diese Vorgehensweise so nur dann anwendbar ist, wenn der Ansatz für die Konklusion der TSK-Regel zu $f(\mathbf{x}) = a_0 + \sum_{j=1}^n a_j \cdot x_j$ gewählt wird. Im allgemeinen Fall, dass $f(\mathbf{x}) = \sum_{j=0}^m a_j \cdot g_j(\mathbf{x})$ ist, kommt dem Parameter a_0 unter Umständen eine ganz andere Bedeutung zu. Zu einem unterbestimmten linearen Gleichungssystem existiert keine eindeutige Lösung. Ein gängiger Ansatz für die Bestimmung einer speziellen Lösung wird im Folgenden vorgestellt.

Lösung mittels Pseudo-Inverser

Da die Matrix \mathbf{G} nicht quadratisch ist, existiert keine Inverse \mathbf{G}^{-1} dieser Matrix. Es kann jedoch eine Pseudo-Inverse \mathbf{G}^+ angegeben werden, welche die gleiche Dimension hat wie \mathbf{G}^T [Moo20, Pen55]. Es gilt $\mathbf{G} \cdot \mathbf{G}^+ \cdot \mathbf{G} = \mathbf{G}$ und $\mathbf{G}^+ \cdot \mathbf{G} \cdot \mathbf{G}^+ = \mathbf{G}^+$, ferner sind $\mathbf{G} \cdot \mathbf{G}^+$ und $\mathbf{G}^+ \cdot \mathbf{G}$ hermitesch. Mit der Pseudo-Inversen kann eine spezielle Lösung $\mathbf{a}_{PI} = \mathbf{G}^+ \cdot \mathbf{y}$ berechnet werden, die sich dadurch auszeichnet, dass sie die kleinste Norm $\|\mathbf{a}\|_2$ aller möglichen Lösungen aufweist.

7.2.2 Überbestimmte lineare Gleichungssysteme

Das lineare Gleichungssystem $\mathbf{G} \cdot \mathbf{a} = \mathbf{y}$ ist überbestimmt, wenn die $m \times n$ Matrix \mathbf{G} mehr Zeilen als Spalten aufweist ($m > n$) und den vollen Rang hat ($\text{rang}(\mathbf{G}) = n$). Das lineare Gleichungssystem hat in diesem Fall keine eindeutige Lösung. Allgemein formuliert besteht die Aufgabe darin, eine Lösung zu finden, die einen *möglichst guten Kompromiss* in der Approximation der zu Grunde liegenden Datenpunkte liefert. Es existieren eine Reihe unterschiedlicher Strategien hierzu, allen ist jedoch der prinzipielle Ansatz gemeinsam: Man definiert eine geeignete Gütefunktion Q , die die resultierenden Approximationsfehler berücksichtigt und sucht denjenigen Parametersatz \mathbf{a} , der die Gütefunktion minimiert.

Ordinary Least-Squares (OLS)

Eine gängige – und bisher auch im Fuzzy-ROSA-Verfahren verfolgte – Vorgehensweise besteht darin, diejenige Lösung zu $\mathbf{G} \cdot \mathbf{a} = \mathbf{y}$ zu bestimmen, die die Summe der quadratischen Fehlerwerte $Q_{SSE} = \sum_{j=1}^m (y_j - \hat{y}_j)^2$ minimiert. Wird Gl. (7.2) nach dem Fehlervektor \mathbf{e}

aufgelöst, so lässt sich auch schreiben

$$\begin{aligned} \text{grad}(Q_{SSE}(\mathbf{a})) &= \mathbf{0} \\ \Leftrightarrow \text{grad}(\mathbf{e}^T \cdot \mathbf{e}) &= \mathbf{0} \\ \Leftrightarrow \text{grad}((\mathbf{y} - \mathbf{G} \cdot \mathbf{a})^T (\mathbf{y} - \mathbf{G} \cdot \mathbf{a})) &= \mathbf{0} \end{aligned} \quad (7.5)$$

woraus sich die Berechnung der in diesem Sinne optimalen Lösung zu

$$\mathbf{a}_{SSE} = (\mathbf{G}^T \cdot \mathbf{G})^{-1} \cdot \mathbf{G}^T \cdot \mathbf{y} \quad (7.6)$$

herleiten lässt. Diese Lösung wird oftmals auch als *least-squares fit* bezeichnet. Aufgrund der Differenzierbarkeit der Gütefunktion Q_{SSE} kann die in diesem Sinne optimale Lösung vergleichsweise einfach berechnet werden. Ein entsprechender Algorithmus hierfür wird beispielsweise in [PTVF99] vorgestellt. Die Gütefunktion Q_{SSE} zeichnet sich durch zwei wesentliche Eigenschaften aus. Zum einen werden alle Datenpunkte gleichberechtigt behandelt, zum anderen werden große Approximationsfehler wesentlich stärker bestraft als kleine. Aufgrund der genannten Eigenschaften können einzelne Datenpunkte einen unter Umständen unerwünscht starken Einfluss auf die resultierende Lösung haben.

Least Median Of Squares (LMS)

Dieses Verfahren wurde von ROUSSEEUW vorgeschlagen [Rou84]. Das Optimierungsproblem besteht in diesem Fall darin, denjenigen Parametersatz \mathbf{a}_{LMS} zu finden der

$$\min_{\mathbf{a}} \text{median}\{(y_i - \mathbf{g}^T(\mathbf{x}_i) \cdot \mathbf{a})^2\} \quad (7.7)$$

liefert. Approximiert man die Datenpunkte durch eine Hyperebene, so bedeutet dies anschaulich, dass man solche Hyperebenen sucht, bei denen 50% der Datenpunkte (bzw. deren Ausgangsgrößenwerte) oberhalb und 50% unterhalb der Hyperebene liegen. Aus der Menge aller Hyperebenen, die diese Anforderung erfüllen, wird diejenige ausgewählt, die den kleinsten Abstand zu den nächstgelegenen Datenpunkten aufweist. Der Vorteil dieses Ansatzes liegt in dem hohen Bruchpunkt, das heißt der Anteil an Ausreißern in den zu approximierenden Daten kann sehr hoch sein. Der Nachteil liegt darin, dass die Gütefunktion viele lokale Minima aufweist und die Bestimmung der optimalen Lösung entsprechend aufwendig ist.

Quantile Regression (QR)

Die Anfälligkeit gegenüber einzelnen Daten tritt bei Verwendung anderer Gütefunktionen ebenfalls weniger stark auf, zum Beispiel wenn die Summe der absoluten Fehlerwerte $Q_{SAE} = \sum_{j=1}^m |y_j - \hat{y}_j|$ verwendet wird. Für diese Gütefunktion ergibt sich eine optimale

Lösung \mathbf{a}_{SAE} , für die in der Regel $\mathbf{a}_{SAE} \neq \mathbf{a}_{SSE}$ gilt. Dieser Ansatz wird in der Literatur als *least absolute error regression* (LAE), *least absolute deviation regression* (LAD) oder auch verkürzt als *L₁ regression* bezeichnet [Abd80]. Er stellt einen Spezialfall der so genannten *quantile regression* [KB78, Koe05] dar. Das Prinzip besteht dabei darin, eine Gütefunktion folgender Art zu verwenden:

$$Q_{QR} = \sum_{y_j \geq \hat{y}_j} \alpha |y_j - \hat{y}_j| + \sum_{y_j < \hat{y}_j} (1 - \alpha) |y_j - \hat{y}_j| \quad . \quad (7.8)$$

Mit anderen Worten: Die betraglichen Approximationsfehler zu denjenigen Datenpunkten mit Ausgangsgrößenwerten y_j , die oberhalb der approximierenden Hyperebene³ liegen, werden mit α gewichtet, die der Datenpunkte mit unterhalb der Hyperebene liegenden Ausgangsgrößenwerten entsprechend mit $(1 - \alpha)$, wobei $\alpha \in [0, 1]$ gilt. Es ist ersichtlich, dass sich für $\alpha = 0.5$ die oben genannte Gütefunktion Q_{SAE} ergibt, lediglich mit einem Vorfaktor versehen. Da das Quantil $\alpha = 0.5$ dem Median entspricht, wird dieser Spezialfall auch als *median regression* bezeichnet. Da die Gütefunktion nach Gl. (7.8) im Ursprung nicht differenzierbar ist, kann die Lösung nicht explizit berechnet werden. Durch eine geeignete Umformung wird die Suche nach der Lösung in Form eines Problems der *Linearen Programmierung* geschrieben. Zur Lösung dieses Problems wird der bekannte Simplex-Algorithmus eingesetzt bzw. modifizierte Varianten hiervon.

Quantile werden in der deskriptiven Statistik zur Beschreibung von Verteilungen eingesetzt. Sie erlauben eine Aussage darüber, wie gleichmäßig sich die Werte innerhalb einer Stichprobe verteilen. Der Grundgedanke bei der Betrachtung von Quantilen besteht darin, nicht nur ein einziges Merkmal (beispielsweise den Mittelwert) einer Verteilung zu betrachten, sondern eine größere Menge an Merkmalen, die einen *besseren Gesamteindruck* über die zu Grunde liegende Verteilung vermitteln. Übertragen auf die Regression liegt es nahe, die Lösungen zu verschiedenen Quantilen α_i zu berechnen und diese miteinander zu vergleichen. Diese Zielsetzung ist eine andere als die der Least-Squares Regression, die eine einzige Lösung hervorbringt. Für die Transformation von Regeln ist der Ansatz der *quantile regression* nur bedingt geeignet, da zu einer Mamdani-Regel nur eine korrespondierende TSK-Regel erzeugt werden soll und nicht eine Menge von TSK-Regeln.

7.3 Parameterbestimmung über QR-Zerlegung mit Spaltenpivotierung

Ein Ansatz zur Bestimmung einer speziellen Lösung eines unterbestimmten linearen Gleichungssystems basiert auf einer QR-Zerlegung mit Spaltenpivotierung der Matrix \mathbf{G} . Eine

³Dies bezieht sich auf den Fall, dass $f(\mathbf{x}) = a_0 + \sum_{j=1}^n a_j \cdot x_j$ ist. Für den allgemeinen Fall, dass $f(\mathbf{x}) = \sum_{j=0}^m a_j \cdot g_j(\mathbf{x})$ ist, ergibt sich allgemein eine Hyperfläche.

$m \times n$ Matrix \mathbf{G} kann als das Produkt zweier Matrizen $\mathbf{G} = \mathbf{Q} \cdot \mathbf{R}$ geschrieben werden, wobei \mathbf{Q} eine orthogonale $m \times m$ Matrix und \mathbf{R} eine $m \times n$ Matrix in Form einer oberen Dreiecksmatrix ist. Wird bei dieser Zerlegung eine Spaltenpivotierung vorgenommen, so werden die entsprechenden Pivotierungen in der $n \times n$ Permutationsmatrix \mathbf{P} nachgehalten. Zu Beginn der QR-Zerlegung ist die Matrix \mathbf{P} eine Einheitsmatrix, deren Spalten im Laufe der QR-Zerlegung vertauscht werden. Es gilt entsprechend $\mathbf{G} \cdot \mathbf{P} = \mathbf{Q} \cdot \mathbf{R}$. Im Gegensatz zu beispielsweise der LU-Zerlegung benötigt die QR-Zerlegung nicht zwangsläufig eine Pivotierung, diese kann aber nützlich sein, um zu erkennen, ob die Matrix singulär oder rangdefizitär ist. Bei jedem Schritt der Faktorisierung wird diejenige Spalte der verbliebenen noch unfaktorierten Matrix \mathbf{G} , welche die größte Norm aufweist, für die Pivotierung genutzt. So ist gesichert, dass die Diagonalelemente von \mathbf{R} in betraglich absteigender Reihenfolge auftreten und lineare Abhängigkeiten zwischen den Spalten durch eine Analyse dieser Diagonalelemente erkannt werden können. Eine entsprechende QR-Faktorisierung kann beispielsweise durch den Einsatz von Householder-Matrizen oder Givens-Matrizen erreicht werden.

Das lineare Gleichungssystem $\mathbf{G} \cdot \mathbf{a} = \mathbf{y}$ lässt sich nach der QR-Zerlegung umschreiben:

$$\begin{aligned} \mathbf{G} \cdot \mathbf{a} &= \mathbf{y} \\ \Leftrightarrow \mathbf{Q} \cdot \mathbf{R} \cdot \mathbf{a} &= \mathbf{y} \\ \Leftrightarrow \mathbf{R} \cdot \mathbf{a} &= \mathbf{Q}^{-1} \cdot \mathbf{y} \\ \Leftrightarrow \mathbf{R} \cdot \mathbf{a} &= \mathbf{Q}^T \cdot \mathbf{y} \quad . \end{aligned} \tag{7.9}$$

Die Matrix \mathbf{R} ist dabei eine obere Dreiecksmatrix. Durch einen Substitutions-Algorithmus kann so leicht eine Lösung \mathbf{a}_{QR} zu Gl. (7.9) berechnet werden. Wird die QR-Zerlegung mit Spaltenpivotierung durchgeführt, so gilt entsprechend

$$\begin{aligned} \mathbf{G} \cdot \mathbf{a} &= \mathbf{y} \\ \Leftrightarrow \mathbf{Q} \cdot \mathbf{R} \cdot \mathbf{P}^{-1} \cdot \mathbf{a} &= \mathbf{y} \\ \Leftrightarrow \mathbf{R} \cdot \mathbf{P}^{-1} \cdot \mathbf{a} &= \mathbf{Q}^{-1} \cdot \mathbf{y} \\ \Leftrightarrow \mathbf{R} \cdot \mathbf{P}^{-1} \cdot \mathbf{a} &= \mathbf{Q}^T \cdot \mathbf{y} \quad . \end{aligned} \tag{7.10}$$

Es ist ersichtlich, dass sich eine Lösung $\mathbf{a}_{QRP} = \mathbf{P} \cdot \mathbf{a}_{QR}$ für Gl. (7.10) leicht aus der zu Gl. (7.9) gefundenen Lösung \mathbf{a}_{QR} herleiten lässt. Die Lösung \mathbf{a}_{QRP} zeichnet sich dadurch aus, dass sie höchstens $k = \text{rang}(\mathbf{G})$ von Null verschiedene Komponenten besitzt. Sie unterscheidet sich in der Regel von der Lösung \mathbf{a}_{PI} , die sich bei Verwendung der Pseudo-Inversen ergibt.

7.4 Parameterbestimmung über Iteratively Reweighted Least-Squares

In Abschnitt 7.2.2 werden verschiedene Ansätze zur Lösung überbestimmter linearer Gleichungssysteme vorgestellt und diskutiert. Nachfolgend wird genauer erläutert, welche Gründe den Einsatz eines *Iteratively Reweighted Least-Squares* (IRLS) [HW77, Hub81, RL87] Verfahrens motivieren. Details zu diesem Ansatz werden in Abschnitt 7.4.2 beschrieben und der prinzipielle Vorteil gegenüber anderen Ansätzen in Abschnitt 7.4.3 anhand eines Anwendungsbeispiels verdeutlicht.

7.4.1 Motivation

In Abschnitt 7.2.2 wird der bisher im Fuzzy-ROSA-Verfahren verfolgte OLS-Ansatz zur Parameterbestimmung vorgestellt. Die vergleichsweise starke Anfälligkeit dieses Ansatzes gegenüber einzelnen Datenpunkten ist häufig unvorteilhaft, insbesondere auch deshalb, weil es sich bei entsprechenden Datenpunkten oftmals um potenziell fehlerhafte Daten handelt, die für die Parameterbestimmung besser ignoriert werden sollten. Das Auftreten solcher Daten kann in industriellen Anwendungen aus verschiedensten Gründen nie ausgeschlossen werden. Dies kann beispielsweise die Folge von defekten Sensoren bei der Datenerfassung oder einer fehlerhaften manuellen Dateneingabe sein. Eine Analyse der Daten unter diesem Aspekt ist insbesondere bei hochdimensionalen Problemen eine nicht triviale Aufgabe und es kann nicht davon ausgegangen werden, dass Anwender eine solche Analyse und Aufbereitung ihrer Daten vor der Modellierung durchführen. Eine bessere Robustheit gegenüber solchen Daten bietet der in Abschnitt 7.2.2 vorgestellte Ansatz der *quantile regression* zur Parameterbestimmung, der jedoch auf die Bestimmung einer Menge von Lösungen abzielt und daher für das hier verfolgte Ziel der Regeltransformation eher ungeeignet ist. Ähnlich wie der genannte Ansatz verfolgt auch der als Iteratively Reweighted Least-Squares bezeichnete Ansatz die Idee, die einzelnen Approximationsfehler zu gewichten und so eine Erhöhung der Robustheit gegenüber fragwürdigen Daten zu erzielen. Die Gewichtung geschieht bei dem IRLS-Ansatz jedoch in einer anderen Art und Weise.

Es ist zunächst zu diskutieren, wann bzw. warum Daten überhaupt fragwürdig erscheinen. Häufig werden so genannte Ausreißer, also Daten, die außerhalb der als normal angesehenen Wertebereiche liegen, als fragwürdig oder fehlerhaft angesehen. Beispielsweise kann ein Datensatz zunächst daraufhin untersucht werden, ob er Datenpunkte enthält, deren Ausgangsgrößenwerte außerhalb des Intervalls $[\bar{y} - \lambda \cdot \sigma, \bar{y} + \lambda \cdot \sigma]$ liegen. Dabei ist \bar{y} der mittlere Ausgangsgrößenwert über alle Datenpunkte, σ ist die Standardabweichung der Ausgangsgrößenwerte und λ ist ein frei wählbarer Parameter⁴. Eine andere Möglichkeit ist die visuelle Untersuchung der Ausgangsgrößenwerte (beispielsweise mit Hilfe von Box-Whisker-Plots,

⁴Ein hierfür häufig gewählter Wert ist $\lambda = 3$.

siehe Anhang B). Bei der Betrachtung eines gesamten Datensatzes eignen sich diese Ansätze, um *globale* Ausreißer zu erkennen. Bei der Transformation einer Mamdani-Regel in eine TSK-Regel ist aber immer nur ein Teil eines Datensatzes – die regelstützenden Datenpunkte – von Interesse. Es stellt sich daher im Rahmen der Regeltransformation vielmehr die Aufgabe, solche Daten zu erkennen, die im Kontext der gerade betrachteten Gruppe von Datenpunkten als *lokale* Ausreißer anzusehen sind. Hierfür könnten im Prinzip die beiden oben genannten Ansätze verfolgt werden, die dann jeweils nur auf einem Teil des gesamten Datensatzes arbeiten. Wie jedoch das Beispiel aus Abschnitt 7.4.3 zeigt, greift die alleinige Betrachtung der Ausgangsgrößenwerte aber unter Umständen zu kurz. Weitere Aspekte wie beispielsweise die Lage der Datenpunkte zueinander im Raum der Eingangsgrößen sind ebenfalls von Bedeutung.

7.4.2 Verfahrensweise

Die grundlegende Idee des Verfahrens besteht darin, zunächst eine Initiaillösung zu bestimmen und die resultierenden Approximationsfehler zu analysieren. Darauf aufbauend werden den einzelnen Datenpunkten bzw. deren Approximationsfehlern Gewichte zugeordnet. Anhand der so gewichteten Daten wird die Lösung erneut berechnet und analysiert. Dieser Vorgang wiederholt sich solange, bis sich die Lösung nicht mehr signifikant verändert. Der nachfolgende Pseudo-Code zeigt den grundsätzlichen Ablauf des Verfahrens.

procedure $\mathbf{a} = \text{IRLS}(\mathbf{G}, \mathbf{y}, \text{MaxNumIter}, \text{WeightingFunction})$

```

1:  $\mathbf{a} = \text{Ordinary Least-Squares Regression}(\mathbf{G}, \mathbf{y})$ 
2:  $\text{Flag} = 1$ 
3:  $\text{NumIter} = 0$ 
4: while ( $\text{NumIter} < \text{MaxNumIter}$  AND  $\text{Flag} == 1$ ) do
5:    $\hat{\mathbf{y}} = \mathbf{G} \cdot \mathbf{a}$ 
6:    $\mathbf{r} = \mathbf{y} - \hat{\mathbf{y}}$ 
7:    $\tilde{\mathbf{r}} = \text{Adjust Residuals}(\mathbf{r}, \mathbf{h})$ 
8:    $\mathbf{w} = \text{Evaluate Weighting Function}(\tilde{\mathbf{r}})$ 
9:    $\tilde{\mathbf{a}} = \text{Weighted Least-Squares Regression}(\mathbf{G}, \mathbf{y}, \mathbf{w})$ 
10:   $\text{Flag} = \text{Compare Solutions}(\mathbf{a}, \tilde{\mathbf{a}})$ 
11:   $\mathbf{a} = \tilde{\mathbf{a}}$ 
12:   $\text{NumIter} = \text{NumIter} + 1$ 
13: end while

```

Es wird zunächst die Initiaillösung \mathbf{a} mit Hilfe des OLS-Ansatzes zu \mathbf{G} und \mathbf{y} bestimmt. Mit Hilfe der so gefundenen Lösung werden die gemessenen Ausgangsgrößenwerte \mathbf{y} durch $\hat{\mathbf{y}} = \mathbf{G} \cdot \mathbf{a}$ approximiert. Die Residuen $\mathbf{r} = \mathbf{y} - \hat{\mathbf{y}}$ werden bestimmt und geeignet individuell gewichtet. Hierfür wird der Vektor \mathbf{h} verwendet, in dem zu jedem Datenpunkt ein zugehöriger konstanter Wert vermerkt ist, der berücksichtigt, wie groß der Einfluss – der auch als Hebel (englisch: leverage) bezeichnet wird – eines jeden Datenpunktes unter Beachtung

aller Datenpunkte auf die Regression ist. Aus den gewichteten Residuen $\tilde{\mathbf{r}}$ wird mittels einer Gewichtungsfunktion der Vektor \mathbf{w} gebildet, um die einzelnen Approximationsfehler zu gewichten. Entsprechend wird eine neue Lösung $\tilde{\mathbf{a}}$ nach dem Weighted Least-Squares-Ansatzes bestimmt. Es wird der Vektor mit den komponentenweisen Beträgen der Differenz $\Delta\mathbf{a} = |\mathbf{a} - \tilde{\mathbf{a}}|$ zwischen dieser neuen und der bisherigen Lösung gebildet. Falls eine oder mehrere Komponenten des Differenzvektors $\Delta\mathbf{a}$ signifikant von Null verschieden sind, wird das Signal zum weiteren Durchlaufen der Schleife beibehalten, andernfalls gelöscht, sodass keine weitere Iteration erfolgt. Die Lösung $\tilde{\mathbf{a}}$ wird als neue Lösung $\mathbf{a} = \tilde{\mathbf{a}}$ übernommen. Sofern die vorgesehene maximal zulässige Anzahl an Iterationen noch nicht ausgeschöpft und das Signal für weitere Iterationen gesetzt ist, wird die Schleife erneut durchlaufen. In der Regel ist es ausreichend, einen Wert von $MaxNumIter = 20$ zu wählen. Weitergehende Betrachtungen im Hinblick auf die Konvergenz des Verfahrens finden sich in [Rya97]. Nachfolgend werden einzelne Aspekte des Verfahrens genauer diskutiert.

Hebelmaß (Leverage)

Diesem Maß liegt die Überlegung zu Grunde, dass der Einfluss eines einzelnen Datenpunktes auf die Regression davon abhängig ist, wie sich die Gesamtheit der Datenpunkte anordnet. Dieser Zusammenhang lässt sich an einem einfachen Beispiel anschaulich illustrieren. Die Aufgabe bestehe darin, die freien Parameter in der Konklusion $f(\mathbf{x}) = a_0 + a_1 \cdot x_1$ einer TSK-Regel zu bestimmen. Abbildung 7.2 zeigt zwei mögliche Szenarien hierfür.

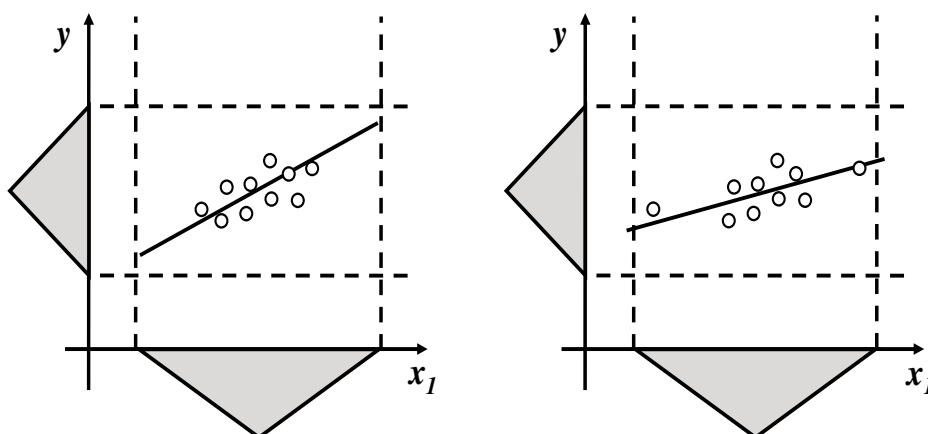


Abbildung 7.2: Beispiel-Szenarien für die Bestimmung von Leverage-Werten: keine (links) bzw. zwei (rechts) high-leverage Datenpunkte.

Die zu transformierende Regel wird in beiden Fällen von jeweils zehn Datenpunkten gestützt, die sowohl die Regelprämisse erfüllen als auch die Regelkonklusion aktivieren. Die

Ausgangsgrößenwerte der regelstützenden Datenpunkte sind in den beiden Szenarien identisch, allerdings unterscheiden sich die Eingangsgrößenwerte der regelstützenden Datenpunkte (siehe auch Tabellen 7.1 und 7.2). Im ersten Szenario (Abbildung 7.2, links) liegen die Eingangsgrößenwerte der regelstützenden Datenpunkte recht dicht beieinander und sind symmetrisch um ihren Mittelwert $\bar{x}_1 = 9.5$ verteilt. Im zweiten Szenario (Abbildung 7.2, rechts) hingegen liegen die Eingangsgrößenwerte der meisten regelstützenden Datenpunkte zwar ebenfalls dicht beieinander, allerdings gibt es zwei Datenpunkte, die aus dieser Gruppe deutlich heraus fallen. Auch hier sind die Datenpunkte symmetrisch um ihren Mittelwert $\bar{x}_1 = 9.5$ verteilt. In Abbildung 7.2 ist für beide Szenarien jeweils zusätzlich die entsprechende Ausgleichsgerade dargestellt, die sich nach dem herkömmlichen Least-Squares Ansatz (siehe Abschnitt 7.2.2) ergibt. Es zeigt sich, dass die Ausgleichsgerade im zweiten Szenario deutlich in Richtung der beiden Datenpunkte gezogen wird, die sich von der Gruppe der übrigen Datenpunkte absetzen.

Die relativ geringen Abweichungen der Datenpunkte zu dem Mittelwert im ersten Szenario spiegeln sich in den zugehörigen Hebelwerten h_i in Tabelle 7.1 wieder. Die Hebelwerte sind die Diagonalelemente der so genannten *hat matrix* \mathbf{H} , die sich gemäß Gl. (7.12) ergibt und auf deren Berechnung später eingegangen wird.

x_1	5	6	7	8	9	10	11	12	13	14
h	0.34	0.24	0.17	0.12	0.10	0.10	0.12	0.17	0.24	0.34

Tabelle 7.1: Hebelwerte der Datenpunkte für das Beispiel aus Abbildung 7.2 (links).

Im zweiten Szenario sind die Eingangsgrößenwerte ebenfalls symmetrisch um ihren Mittelwert verteilt, aber zwei Datenpunkte liegen deutlich außerhalb der Gruppe der übrigen Datenpunkte. Die relativ großen Abweichungen der Eingangsgrößenwerte dieser zwei Datenpunkte zu dem Mittelwert schlagen sich in entsprechend hohen Hebelwerten in Tabelle 7.2 nieder. Die Hebelwerte der übrigen Datenpunkte sind entsprechend geringer als im ersten Szenario.

x_1	2	6	7	8	9	10	11	12	13	17
h	0.46	0.17	0.14	0.11	0.10	0.10	0.11	0.14	0.17	0.46

Tabelle 7.2: Hebelwerte der Datenpunkte für das Beispiel aus Abbildung 7.2 (rechts).

Als Analogie kann die aus der Mechanik bekannte Hebelwirkung herangezogen werden: Das durch eine Kraft verursachte Drehmoment wird umso größer, je weiter der Angriffspunkt der Kraft von der Drehachse entfernt ist. Übertragen auf das hier vorliegende Problem der Regression kann vereinfacht gesagt werden: Der mittlere Eingangsgrößenwert $\bar{x}_1 = 9.5$ aller Datenpunkte kann als Drehachse angesehen werden und je weiter der Eingangsgrößenwert

eines Datenpunktes von dem Mittelwert abweicht, desto größer ist seine Hebelwirkung im Hinblick auf die Regression.

Allgemein werden die Hebelwerte \mathbf{h} der Datenpunkte anhand der Matrix \mathbf{G} berechnet. Durch Einsetzen der Lösung aus Gl. (7.6) ergibt sich:

$$\begin{aligned} \mathbf{G} \cdot \mathbf{a} &= \hat{\mathbf{y}} \\ \mathbf{G} \cdot (\mathbf{G}^T \cdot \mathbf{G})^{-1} \cdot \mathbf{G}^T \cdot \mathbf{y} &= \hat{\mathbf{y}} \end{aligned} \quad (7.11)$$

Die Hebelwerte sind gleich den Elementen der Hauptdiagonalen der Matrix \mathbf{H} :

$$\mathbf{H} = \mathbf{G} \cdot (\mathbf{G}^T \cdot \mathbf{G})^{-1} \cdot \mathbf{G}^T \quad (7.12)$$

Die Matrix \mathbf{H} wird auch als *hat matrix* bezeichnet, da sich mit Hilfe dieser Matrix die modellierten Ausgangsgrößenwerte über $\hat{\mathbf{y}} = \mathbf{H} \cdot \mathbf{y}$ berechnen lassen. Es ist ersichtlich, dass die Elemente der Hauptdiagonalen den Einfluss der einzelnen Datenpunkte widerspiegeln.

Eine Faustregel zur Erkennung von Datenpunkten mit sehr starken Hebelwerten ist der Vergleich der berechneten Hebelwerte mit dem Referenzwert $h_{ref} = (2n)/m$. Dabei ist n die Anzahl an Spalten der Matrix \mathbf{G} und entspricht damit der Anzahl an freien Parametern in der Konklusion einer TSK-Regel, m ist die Anzahl an Zeilen der Matrix \mathbf{G} und entspricht damit der Anzahl an regelstützenden Datenpunkten. Datenpunkte, die einen Hebelwert $h_i > h_{ref}$ aufweisen, können als *high leverage* Datenpunkte angesehen werden. In Abbildung 7.3 ist dieser Referenzwert für unterschiedliche Werte von m und n dargestellt.

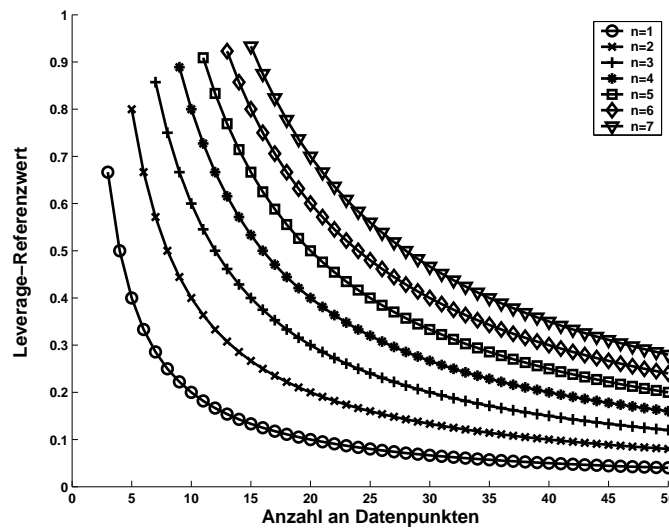


Abbildung 7.3: Leverage-Referenzwert für unterschiedliche Anzahl m an Datenpunkten bei unterschiedlicher Anzahl n an Parametern in der Konklusion einer TSK-Regel.

In dem Beispiel aus Abbildung 7.3 wird vorausgesetzt, dass $m > 2n$ ist und daher $h_{ref} < 1$ gilt. Bei konstanter Anzahl n an freien Parametern und zunehmender Anzahl m an re-

gestützten Datenpunkten wird der Referenzwert zunehmend kleiner. Dies ist auch anschaulich nachvollziehbar, da mit zunehmender Anzahl an regelstützenden Datenpunkten insgesamt der Einfluss eines einzelnen Datenpunktes auf die Regression tendenziell immer geringer werden sollte. Für $m = 10$ und $n = 2$ (siehe Beispiele aus Abbildung 7.2) ergibt sich so der Leverage-Referenzwert von $h_{ref} = 0.4$. Vergleicht man diesen Referenzwert mit den Werten aus den Tabellen 7.1 und 7.2, so zeigt sich, dass im ersten Szenario keine high-leverage Datenpunkte vorhanden sind, dafür aber im zweiten Szenario. Dies sind gerade die beiden Datenpunkte, die sich von der Gruppe der übrigen Datenpunkte abgrenzen. Ein weiteres Maß zur Erkennung von Datenpunkten mit großer Hebelwirkung ist die so genannte *cook's distance* [CW82].

Anpassung der Residuen

Aus dem Hebelwert eines Datenpunktes wird der Anpassungsfaktor für das zugehörige Residuum dieses Datenpunktes gebildet. Je größer der Hebelwert eines Datenpunktes ist, desto größer sollte der entsprechende Anpassungsfaktor sein. In [DO89] wird folgender Ansatz vorgeschlagen:

$$s_i = 1/\sqrt{1 - h_i} \quad . \quad (7.13)$$

Dabei wird vorausgesetzt, dass $h_i < 1$ gilt (siehe auch Abbildung 7.3). Unter Verwendung dieses Ansatzes ergeben sich die in Tabelle 7.3 angegebenen Anpassungsfaktoren.

h	0.01	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	0.99
s	1.01	1.05	1.12	1.20	1.29	1.41	1.58	1.83	2.24	3.16	4.47	10.00

Tabelle 7.3: Anpassungsfaktoren s nach Gl. (7.13) für ausgewählte Werte von h .

Für Hebelwerte $h \leq 0.5$ erfolgt eine eher geringe Verstärkung, für $0.5 < h \leq 0.8$ nimmt die Verstärkung deutlich zu und für $0.8 < h$ steigt die Verstärkung stark an. Jedes Residuum wird mit dem zugehörigen Anpassungsfaktor multipliziert:

$$\check{r}_i = r_i \cdot s_i \quad . \quad (7.14)$$

Um aus den so angepassten Residuen die Gewichtungswerte berechnen zu können, werden diese geeignet standardisiert:

$$\tilde{r}_i = \check{r}_i / (\hat{\sigma}_r \cdot \eta) \quad . \quad (7.15)$$

Dabei ist $\hat{\sigma}_r$ die geschätzte Standardabweichung der angepassten Residuen \check{r} und η ist ein Tuning-Faktor, der in Abhängigkeit der eingesetzten Gewichtungsfunktion (siehe nachfolgenden Abschnitt) gewählt wird. Details hierzu finden sich in [DO89].

Gewichtungsfunktionen

Auf Grundlage der angepassten Residuen werden Gewichtungswerte berechnet. Die Approximationsfehler der einzelnen Datenpunkte werden entsprechend gewichtet und eine Weighted-Least-Squares Regression durchgeführt (siehe nachfolgenden Abschnitt). Verschiedene Ansätze für Gewichtungsfunktionen [Hub81] sind etabliert, die hier jedoch nicht alle vorgestellt werden. Es ist ersichtlich, dass sich durch die Gewichtungsfunktion

$$w_i = 1 \quad (7.16)$$

die herkömmliche Least-Squares Regression (OLS) realisieren lässt. Die Gewichtungsfunktion nach HUBER ist wie folgt definiert:

$$w_i = \begin{cases} 1 & \text{falls } |\tilde{r}_i| < 1 \quad , \\ 1/|\tilde{r}_i| & \text{sonst} \quad . \end{cases} \quad (7.17)$$

Durch die so genannte *bisquare* Gewichtungsfunktion nach TUKEY

$$w_i = \begin{cases} (1 - \tilde{r}_i^2)^2 & \text{falls } |\tilde{r}_i| < 1 \quad , \\ 0 & \text{sonst} \end{cases} \quad (7.18)$$

werden Datenpunkte mit einem großen Residuum so gewichtet, dass sie zu der Gütefunktion keinen Beitrag liefern und daher bei der Bestimmung der Lösung nicht von Bedeutung sind. Diese Gewichtungsfunktion wird in dieser Arbeit eingesetzt (siehe Abschnitt 9.4). Allen Ansätzen ist gemeinsam, dass für Datenpunkte, die ein betragsmäßig großes Residuum aufweisen, das zugehörige Gewicht tendenziell immer geringer wird. Dies sind zum einen solche Datenpunkte, die sehr schlecht approximiert werden, aber nur einen vergleichsweise kleinen Hebelwert aufweisen. Zum anderen sind dies solche Datenpunkte, die mäßig approximiert werden, aber einen vergleichsweise großen Hebelwert aufweisen.

Parameterbestimmung über Weighted Least-Squares

Innerhalb der Schleife (siehe den Pseudo-Code des IRLS-Ansatzes) wird die Gütefunktion $Q_{SSEW} = \sum_{j=1}^m w_j \cdot (y_j - \hat{y}_j)^2$ verwendet, bei der die einzelnen Approximationsfehler gemäß ihres Gewichts berücksichtigt werden. In diesem Fall lässt sich

$$\begin{aligned} \text{grad}(Q_{SSEW}(\mathbf{a})) &= \mathbf{0} \\ \Leftrightarrow \text{grad}(\mathbf{e}^T \cdot \mathbf{W} \cdot \mathbf{e}) &= \mathbf{0} \\ \Leftrightarrow \text{grad}((\mathbf{y} - \mathbf{G} \cdot \mathbf{a})^T \cdot \mathbf{W} \cdot (\mathbf{y} - \mathbf{G} \cdot \mathbf{a})) &= \mathbf{0} \end{aligned} \quad (7.19)$$

schreiben (vergleiche auch Gl. (7.5)), wobei \mathbf{W} eine $m \times m$ Diagonalmatrix mit den entsprechenden Gewichten als Hauptdiagonalelementen ist. Die in diesem Sinne optimale Lösung ist dann entsprechend

$$\mathbf{a}_{SSEW} = (\mathbf{G}^T \cdot \mathbf{W} \cdot \mathbf{G})^{-1} \cdot \mathbf{G}^T \cdot \mathbf{W} \cdot \mathbf{y} \quad . \quad (7.20)$$

Abbruchkriterien

Die innerhalb eines Schleifendurchlaufs über den Weighted Least-Squares Ansatz gefundene Lösung $\tilde{\mathbf{a}}$ wird mit der bisherigen Lösung \mathbf{a} komponentenweise verglichen. Bei einer als signifikant anzusehenden Veränderung einer oder mehrerer Komponenten (beispielsweise wenn $|a_i - \tilde{a}_i| > 10^{-6}$ ist) wird die Schleife erneut durchlaufen, sofern die maximal zulässige Anzahl an Iterationen noch nicht ausgeschöpft ist. Falls keine signifikante Veränderung der Lösung mehr auftritt oder die maximal zulässige Anzahl an Iterationen erreicht ist, bricht das Verfahren ab und die letzte berechnete Lösung wird zurückgegeben.

7.4.3 Anwendungsbeispiel

Der prinzipielle Vorteil bei der Parameterbestimmung über den IRLS-Ansatz gegenüber dem OLS-Ansatz im Sinne einer erhöhten Robustheit gegenüber potenziell fehlerhaften Daten wird hier an einem einfachen Beispiel illustriert. Es handelt sich um ein Modellierungsproblem mit nur einer Eingangsgröße X_1 . Eine zuvor erzeugte Mamdani-Regel wird in eine korrespondierende TSK-Regel transformiert. Die Mamdani-Regel wird von zehn Datenpunkten gestützt, als Konklusion für die TSK-Regel wird der Standard-Funktionsansatz einer Hyperebene verwendet.

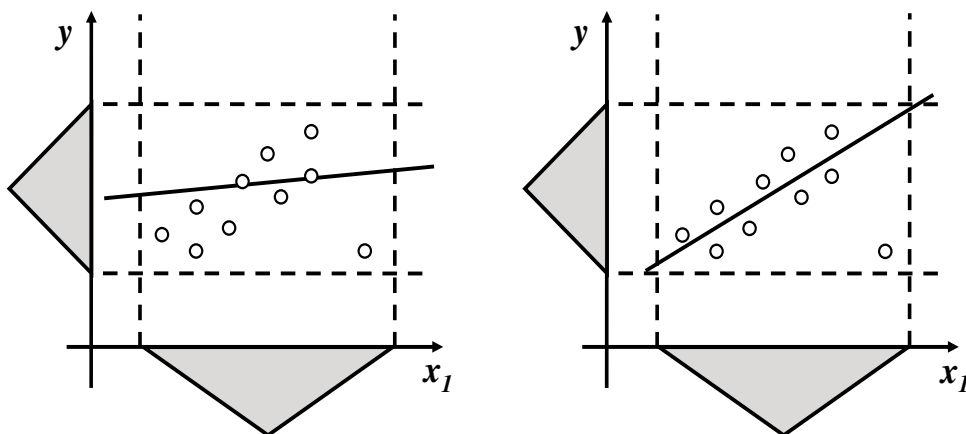


Abbildung 7.4: Beispiel für die Bestimmung der Ausgleichsgeraden über den OLS-Ansatz (links) und den IRLS-Ansatz (rechts).

In Abbildung 7.4 sind die zehn regelstützenden Datenpunkte durch Kreise und die dreieckförmigen Zugehörigkeitsfunktionen für die in diesem Beispiel relevanten linguistischen Werte sowohl der Eingangs- als auch der Ausgangsgröße grau dargestellt. Die freien Parameter in der Konklusion $f(\mathbf{x}) = a_0 + a_1 \cdot x_1$ der TSK-Regel werden einmal über den

OLS-Ansatz (linke Grafik) und einmal über den IRLS-Ansatz (rechte Grafik) bestimmt. Die resultierende Ausgleichsgerade ist jeweils als durchgezogene Linie dargestellt. Im Fall der Parameterbestimmung über den OLS-Ansatz wird die Ausgleichsgerade in Richtung des Datenpunktes \mathbf{d}_{10} gezogen, der im Kontext aller übrigen regelstützenden Datenpunkte eher fragwürdig erscheint. Im Gegensatz dazu wird dieser Datenpunkt bei der Parameterbestimmung über den IRLS-Ansatz so stark abgewichtet, dass er für die resultierende Lösung keine Bedeutung mehr hat. Das Ergebnis entspricht im Prinzip dem Ergebnis für den Fall, dass nur die regelstützenden Datenpunkte $\mathbf{d}_1, \dots, \mathbf{d}_9$ vorhanden sind, die einen nahezu linearen Zusammenhang aufweisen. Im Hinblick auf die erzielten Approximationsfehler – mittlerer absoluter Fehler (MAE) und mittlerer quadratischer Fehler (MSE) – bzgl. aller Datenpunkte zeigt sich, dass die Lösung über den OLS-Ansatz bzgl. des MSE besser ist als die Lösung über den IRLS-Ansatz (um ca. 40%), dafür ist die Lösung über den IRLS-Ansatz bzgl. des MAE besser als die Lösung über den OLS-Ansatz (um ca. 30%). Welche der beiden Lösungen als insgesamt besser anzusehen ist, hängt davon ab, wie der Datenpunkt \mathbf{d}_{10} eingestuft wird. Betrachtet man diesen Datenpunkt im Kontext aller Datenpunkte als Ausreißer bzw. als Störung, dann erscheint die Lösung über den IRLS-Ansatz wesentlich sinnvoller als die Lösung über den OLS-Ansatz, da sie die übrigen Datenpunkte wesentlich genauer, vor allem aber auch in der Tendenz besser nachbildet.

Im Hinblick auf die Modellierungsergebnisse, die im Rahmen der systematischen Exploration der Einstellmöglichkeiten diskutiert werden (siehe Abschnitte 9.3 und 9.4), sei darauf hingewiesen, dass, bezogen auf *eine* Regel, der mittlere quadratische Approximationsfehler (MSE) der Lösung, die durch den OLS-Ansatz gefunden wird, optimal im Hinblick auf dieses Fehlermaß ist. Die über den IRLS-Ansatz gefundene Lösung wird nahezu immer einen größeren MSE-Wert liefern, wie auch das obige Beispiel deutlich zeigt. Das Fuzzy-ROSA-Verfahren unterstützt (wie in Abschnitt 4.3 ausgeführt) die Verwendung generalisierender Regeln, daher werden durch einen Eingangsgrößenvektor \mathbf{x} sehr häufig mehrere Regeln gleichzeitig aktiviert. Der Ausgabewert eines TSK-Systems wird zum Beispiel durch Überlagerung der einzelnen Empfehlungen der Regeln gemäß Gl. (7.21) bestimmt. Es kann daher der Fall eintreten, dass sich bei Verwendung des IRLS-Ansatzes über alle Regeln ein insgesamt kleinerer MSE-Wert ergibt als bei dem OLS-Ansatz. Umgekehrt kann aber auch bei Verwendung des OLS-Ansatzes der mittlere absolute Approximationsfehler (MAE) über alle Regeln kleiner sein als bei Verwendung des IRLS-Ansatzes.

7.5 Berechnung des Ausgabewertes über den gewichteten Median

Bei der Berechnung des Ausgabewertes eines TSK-Systems wird typischerweise so vorgegangen, dass die Empfehlungen der einzelnen TSK-Regeln mit dem jeweiligen Glaubensgrad und dem Erfülltheitsgrad der Regelprämisse gewichtet werden und eine entsprechende

Mittelwertbildung

$$y_D(\mathbf{x}) = \frac{\sum_{i=1}^R \rho_i \cdot \mu_{P_i}(\mathbf{x}) \cdot f_i(\mathbf{x})}{\sum_{i=1}^R \rho_i \cdot \mu_{P_i}(\mathbf{x})} \quad (7.21)$$

über alle aktivierten Regeln durchgeführt wird. Wie bereits im Zusammenhang mit dem Mittelwertbasierten Regeltest erläutert, ist der Mittelwert vergleichsweise anfällig gegenüber Ausreißern. Wenige unter Umständen unplausible Einzelempfehlungen $f_i(\mathbf{x})$ können daher einen unerwünscht starken Einfluss auf den Ausgabewert $y_D(\mathbf{x})$ haben. Eine alternative Strategie besteht darin, statt des gewichteten Mittelwertes den gewichteten Median aller Einzelempfehlungen als Ausgabewert zu verwenden. Hierfür werden die Einzelempfehlungen zunächst in aufsteigender Reihenfolge sortiert, wobei zu jeder Einzelempfehlung $f_i(\mathbf{x})$ das zugehörige Gewicht $w_i = \rho_i \cdot \mu_{P_i}(\mathbf{x})$ vermerkt wird. Der gewichtete Median ist diejenige Einzelempfehlung $f_k(\mathbf{x})$, für die die beiden Ungleichungen

$$\sum_{f_i(\mathbf{x}) < f_k(\mathbf{x})} w_i < \frac{1}{2} \sum_{i=1}^R w_i \quad \text{und} \quad \sum_{f_i(\mathbf{x}) > f_k(\mathbf{x})} w_i > \frac{1}{2} \sum_{i=1}^R w_i \quad (7.22)$$

erfüllt sind⁵. Falls ein solcher Wert nicht existiert, muss derjenige Wert $f_j(\mathbf{x})$ bestimmt werden, für den die Ungleichung

$$\sum_{f_i(\mathbf{x}) \leq f_j(\mathbf{x})} w_i < \frac{1}{2} \sum_{i=1}^R w_i \quad (7.23)$$

gerade noch erfüllt ist. Aus diesem Wert und dem nächst größeren Wert $f_{j+1}(\mathbf{x})$ wird dann durch eine Mittelwertbildung der gewichtete Median gebildet. Der so berechnete Ausgabewert des TSK-Systems ist deutlich robuster gegenüber extremen und daher unter Umständen fragwürdigen Einzelempfehlungen, wie das nachfolgende Beispiel in Abbildung 7.5 illustriert.

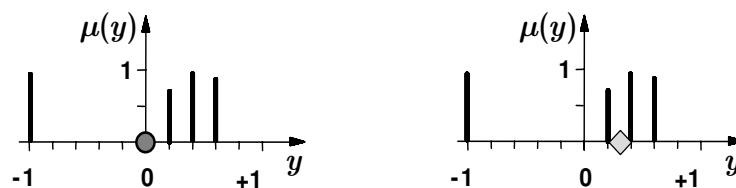


Abbildung 7.5: Beispiel für die Ausgabeberechnung eines TSK-Systems nach dem gewichteten Mittelwert (links) und nach dem gewichteten Median (rechts) der Einzelempfehlungen.

⁵Hierbei wird vorausgesetzt, dass ein Wert nicht mehrfach auftritt. Falls ein Wert mehrfach auftritt, werden alle Einträge dieses Wertes zu einem Eintrag zusammengefasst, dem als Gewicht die Summe aller vorherigen Einzelgewichte zugewiesen wird.

Dargestellt sind die Empfehlungen der einzelnen TSK-Regeln in Form von ausgangsseitigen Singletons und die jeweiligen Ausgabewerte bei Verwendung des gewichteten Mittelwertes (Abbildung 7.5, links) und des gewichteten Median (Abbildung 7.5, rechts).

Es sei an dieser Stelle angemerkt, dass die Berechnung des Ausgabewertes bei einem TSK-System und die Defuzzifizierung bei einem Mamdani-System mit ausgangsseitigen Singletons starke Parallelen aufweisen (siehe auch das vorherige Beispiel). So entspricht die Defuzzifizierung mit der COS-Methode bei solchen Mamdani-Systemen im Prinzip der gewichteten Mittelwertbildung nach Gl. (7.21) für TSK-Systeme. Daher lässt sich der hier für TSK-Systeme beschriebene Ansatz, den gewichteten Median als Ausgabewert zu verwenden, prinzipiell auch auf entsprechend ausgelegte Mamdani-Systeme übertragen. Umgekehrt lassen sich aber auch Strategien wie beispielsweise das Inferenzfilter, die für die Defuzzifizierung bei Mamdani-Systemen entwickelt wurden, für die Ausgabeberechnung in TSK-Systemen einsetzen.

Kapitel 8

Hybrides Fuzzy-System

Wie in Abschnitt 7.1 erwähnt, ist die Transformation von Mamdani-Regeln in TSK-Regeln auf positive Mamdani-Regeln beschränkt. Die typische Vorgehensweise bei der Modellierung mit dem Fuzzy-ROSA-Verfahren – die Erstellung eines Mamdani-Systems und die anschließende Transformation in ein TSK-System – ist im Fall des Vorliegens negativer Regeln daher so nicht möglich. Dies ist ein Defizit des Verfahrens, da das Potenzial negativer Regeln so nicht in TSK-Systemen genutzt werden kann. Es ist aber auch deshalb ein Manko, da es so schwerer fällt, industrielle Anwender für den prinzipiellen Nutzen negativer Regeln zu sensibilisieren. Industrielle Anwender bevorzugen in sich geschlossene Lösungsansätze und stehen Teillösungen eher skeptisch gegenüber. Negative Regeln werden nur dann vollständig akzeptiert, wenn sich durch deren Beachtung für Anwender Vorteile ergeben und sie dabei weiterhin die ihnen bereits bekannte und vertraute Vorgehensweise zur Erstellung von Modellen anwenden können. Die beiden genannten Punkte motivieren daher die Suche nach einer Lösung, die die beschriebene Lücke zwischen negativen Regeln und TSK-Systemen schließt bzw. zumindest überbrückt.

Nachfolgend wird in Abschnitt 8.1 zunächst erläutert, was die Verwendung negativer Regeln grundsätzlich motiviert, wie negative Regeln aufgebaut sind und wie diese in Mamdani-Systemen genutzt und verarbeitet werden. In Abschnitt 8.2 wird darauf eingegangen, worin die Schwierigkeiten bei der angestrebten Transformation negativer Mamdani-Regeln in TSK-Regeln bestehen und es werden mögliche Lösungsansätze für dieses Problem diskutiert. Dabei zeigt sich, dass die in Abschnitt 7.1 beschriebene Transformation für positive Mamdani-Regeln nur sehr bedingt auf negative Mamdani-Regeln übertragbar ist. Daher wird in Abschnitt 8.3 ein Ansatz für ein hybrides Fuzzy-System [SKK05] vorgestellt, mit dem negative Mamdani-Regeln gewinnbringend in Verbindung mit TSK-Systemen genutzt werden können. Dieser Ansatz zielt nicht darauf ab, eine Transformation der negativen Mamdani-Regeln in wie auch immer geartete negative TSK-Regeln durchzuführen, sondern verfolgt vielmehr das Ziel, die negativen Mamdani-Regeln dazu zu verwenden, eine situationsbezogene Selektion geeigneter positiver TSK-Regeln vorzunehmen. Abschnitt 8.4

zeigt den prinzipiellen Nutzen des hybriden Fuzzy-Systems an einem Anwendungsbeispiel.

8.1 Mamdani-Systeme mit negativen Regeln

Aus der Theorie ist bekannt, dass sich durch entsprechend ausgelegte Mamdani-Systeme, die Regeln der Art

$$\text{WENN } \langle \textit{Prämisse} \rangle \text{ DANN } \langle \textit{Konklusion} \rangle \quad (8.1)$$

verwenden, universelle Approximatoren realisieren lassen [Kos94, CD96]. Regeln dieser Art drücken Empfehlungen aus und werden im Folgenden daher auch als *positive Regeln* bezeichnet. In verschiedenen Bereichen des alltäglichen Lebens, aber auch in der Natur, finden sich jedoch Beispiele, in denen Anweisungen, Wissen oder Erfahrungen in Form von Warnungen oder Verboten formuliert wird. In Analogie zu positiven Regeln, die Empfehlungen ausdrücken, können daher auch *negative Regeln* formuliert werden, die entsprechend Warnungen oder Verbote repräsentieren. Negative Mamdani-Regeln sind von der Art

$$\text{WENN } \langle \textit{Prämisse} \rangle \text{ DANN NICHT } \langle \textit{Konklusion} \rangle \quad (8.2)$$

und wurden von KIENDL für die Nutzung in Fuzzy-Systemen vorgeschlagen [Kie93a, Kie93b, KK94, Kie99]. Der prinzipielle Nutzen negativer Regeln wurde bereits an verschiedenen Anwendungen gezeigt [KFR95, SKS95, Kie97a, Sch99]. Um die Motivation für die Verwendung negativer Regeln in Fuzzy-Systemen zu verdeutlichen, werden nachfolgend einige Beispiele aus verschiedenen anderen Feldern vorgestellt, in denen sich solche Zusammenhänge zeigen.

Im alltäglichen Leben trifft man Verbote bzw. negative Regeln in unterschiedlichsten Fällen an. Ein besonders anschauliches Beispiel sind Verkehrsregeln. Hier finden sich Überhol-, Halte- und diverse andere Verbote, die zusammen einen Großteil des Gesamtregelwerks bilden. Ein weiteres Beispiel mit einem starken Praxisbezug ist die Art und Weise, wie Prozessbediener (zum Beispiel Anlagenführer) das ihnen bekannte Wissen bzw. ihre Handlungsmechanismen erklären. Zur Vermeidung kritischer Prozesszustände dürfen in bestimmten Situationen gewisse Handlungen gerade nicht durchgeführt werden. Wenn beispielsweise der Druck in einem chemischen Reaktor zu hoch zu werden droht, darf in keinem Falle eine weitere Energiezufuhr erfolgen. Dieses Verbot liefert keinen alternativen Handlungsvorschlag. Falls also keine weiteren Handlungsempfehlungen (positive Regeln) für diese Situation vorliegen, kann es am günstigsten sein, den Prozess sich selbst zu überlassen, damit er sich möglichst normalisiert. Auch in der Biologie findet sich ein Beispiel, das auf negativen Regeln beruht. Es ist seit längerer Zeit bekannt, dass Ameisen über Duftstoffe (Pheromone) miteinander kommunizieren können. Ameisenkolonien nutzen diese Möglichkeit überaus effektiv für ihre Futtersuche. Einzelne Ameisen, die eine ergiebige Futterquelle gefunden haben, markieren den Weg dorthin mit entsprechenden Duftmarken, damit dieser Weg als lohnend angesehen und von vielen weiteren Ameisen genutzt wird. Dieses der

Natur entlehnte Prinzip der *Ant Colony Optimization* [CDM91, DMC96] wird mittlerweile auch höchst erfolgreich in technischen Anwendungen eingesetzt, beispielsweise für die Routenplanung in komplexen Versand- oder Rechnernetzen. Jüngst haben Forschungen gezeigt, dass bestimmte Ameisenarten zusätzlich aber auch solche Duftmarken verwenden, die vor bestimmten Wegen warnen [RJHR05]. Dies sind solche Wege, die zu keiner Futterquelle führen und daher von den übrigen Ameisen möglichst gemieden werden sollten. Die Verbindung zu negativen Regeln ist offensichtlich. Die genannten Beispiele zeigen, dass das Prinzip negativer Regeln in verschiedensten Bereichen anzutreffen ist und daher die Nutzung dieses Prinzips auch in Fuzzy-Systemen ein großes Potenzial verspricht. Wie einleitend erwähnt, lassen sich auch ohne negative Regeln theoretisch beliebige Kennfelder durch Fuzzy-Systeme realisieren. Der Zugewinn durch die Verwendung negativer Regeln liegt aber darin, dass diese eine sehr kompakte Darstellung bzw. Formulierung entsprechender Zusammenhänge erlauben und damit für die Kompaktheit einer Regelbasis und damit letztlich auch für die Transparenz eines Fuzzy-Systems von Vorteil sein können. Nachfolgend wird zunächst erläutert, wie negative Regeln in Mamdani-Systemen genutzt bzw. verarbeitet werden. Es werden einzelne Teilaufgaben genauer vorgestellt, da sie für das hybride Fuzzy-System (siehe Abschnitt 8.3) von Bedeutung sind.

Zweisträngiges Konzept zur Verarbeitung negativer Regeln

Für die Verarbeitung von positiven und negativen Mamdani-Regeln wird ein zweisträngiges Konzept nach KIENDL [Kie97a] verwendet, das in Abbildung 8.1 dargestellt ist.

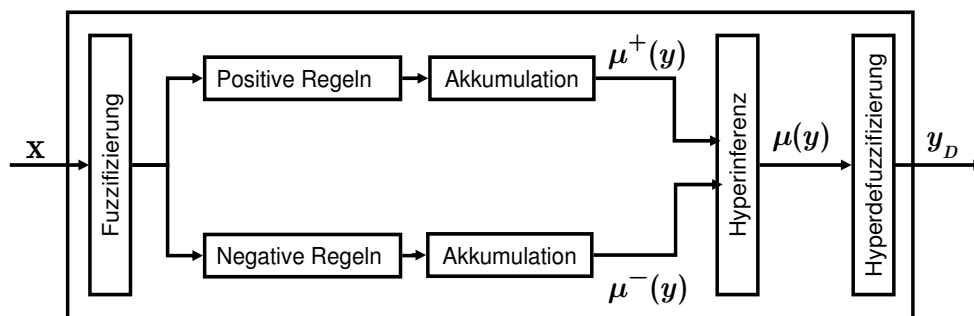


Abbildung 8.1: Prinzip eines zweisträngigen Mamdani-Systems.

Die Regelbasis, die sowohl positive als auch negative Regeln enthält, wird zunächst in zwei Teil-Regelbasen (Stränge) aufgeteilt, die jeweils nur positive bzw. nur negative Regeln enthalten. Für beide Stränge wird jeweils die Inferenz gemäß Gl. (3.11) durchgeführt.

Der Vorteil dieser Vorgehensweise liegt unter anderem darin, dass für die beiden Stränge gegebenenfalls unterschiedliche Operatoren für die UND- bzw. ODER-Verknüpfungen gewählt werden können. Es ergeben sich so zwei ausgangsseitige Zugehörigkeitsfunktionen: $\mu^+(y)$ für die Empfehlungen der positiven Regeln und $\mu^-(y)$ für die Warnungen der negativen Regeln¹. Diese beiden ausgangsseitigen Zugehörigkeitsfunktionen werden durch die Hyperinferenz zu dem resultierenden Empfehlungsgebirge $\mu(y)$ vereinigt. Die Aufgabe der Hyperdefuzzifizierung besteht darin, aus dem resultierenden Empfehlungsgebirge $\mu(y)$ einen sinnvollen scharfen Ausgangsgrößenwert y_D des Mamdani-Systems zu erzeugen.

Hyperinferenzstrategien

Das Fuzzy-ROSA-Verfahren bietet drei unterschiedliche Strategien [Kie97a] für die Vereinigung von $\mu^+(y)$ und $\mu^-(y)$ zu $\mu(y)$. Bei Verwendung der Strong-Veto-Hyperinferenz wird die ausgangsseitige Zugehörigkeitsfunktion $\mu^-(y)$ der negativen Regeln unmittelbar als Verbot interpretiert:

$$\mu(y) = \begin{cases} 0, & \text{falls } \mu^-(y) > 0, \\ \mu^+(y) & \text{sonst} \end{cases} . \quad (8.3)$$

Im Gegensatz dazu wird bei der Weak-Veto-Hyperinferenz die ausgangsseitige Zugehörigkeitsfunktion $\mu^-(y)$ der negativen Regeln erst dann als Verbot interpretiert, wenn der Grad der Warnung gleich oder größer dem Grad der Empfehlung ist:

$$\mu(y) = \begin{cases} 0, & \text{falls } \mu^-(y) \geq \mu^+(y), \\ \mu^+(y) & \text{sonst} \end{cases} . \quad (8.4)$$

Die Fuzzy-Veto-Hyperinferenz bildet einen Kompromiss zwischen den beiden zuvor genannten Strategien:

$$\mu(y) = \mu^+(y) \wedge \neg\mu^-(y) . \quad (8.5)$$

Das Resultat ist in diesem Fall abhängig von der Wahl des UND-Operators. Die Funktionsweise der Hyperinferenz soll an einem einfachen Beispiel verdeutlicht werden. Abbildung 8.2 zeigt beispielhaft die ausgangsseitigen Empfehlungsgebirge $\mu^+(y)$ und $\mu^-(y)$, die sich nach Auswertung der Teil-Regelbasen ergeben.

¹Es wird hier und im Folgenden statt $\mu(\mathbf{x}, y)$ verkürzt $\mu(y)$ geschrieben.

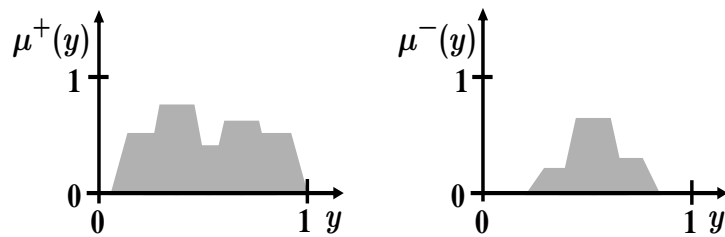


Abbildung 8.2: Beispiel für ausgangsseitige Empfehlungsgebirge $\mu^+(y)$ (links) und $\mu^-(y)$ (rechts).

Diese Empfehlungsgebirge werden durch Hyperinferenz zu dem resultierenden ausgangsseitigen Empfehlungsgebirge $\mu(y)$ verrechnet. Abbildung 8.3 zeigt die Ergebnisse bei Anwendung der Strong-Veto-Hyperinferenz und der Weak-Veto-Hyperinferenz für das obige Beispiel.

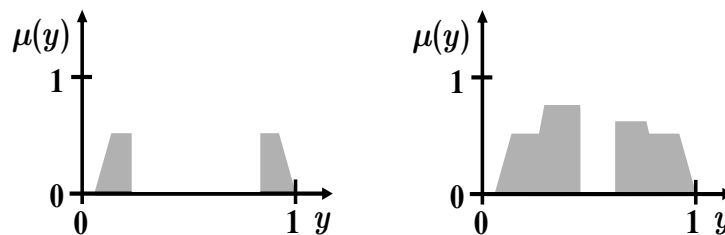


Abbildung 8.3: Resultierendes Empfehlungsgebirge $\mu(y)$ für das Beispiel aus Abbildung 8.2 bei Strong-Veto-Hyperinferenz (links) und bei Weak-Veto-Hyperinferenz (rechts).

Hyperdefuzzifizierungsstrategien

Die Aufgabe der Hyperdefuzzifizierung besteht darin, aus dem resultierenden Empfehlungsgebirge $\mu(y)$ einen sinnvollen Ausgangsgrößenwert y_D zu erzeugen, für den in keinem Falle $\mu(y_D) = 0$ gelten sollte. Dabei ist zu bedenken, dass die Hyperinferenz für einen Ausgangsgrößenwert y_i aus qualitativ unterschiedlichen Gründen zu $\mu(y_i) = 0$ führen kann. Erstens kann $\mu^+(y_i) = 0$ und $\mu^-(y_i) = 0$ gelten. Der Ausgangsgrößenwert y_i wird dann weder empfohlen noch verboten. Zweitens kann $\mu^+(y_i) > 0$ und $\mu^-(y_i) > 0$ gelten. Der Ausgangsgrößenwert y_i wird dann sowohl empfohlen als auch verboten. Drittens kann $\mu^+(y_i) = 0$

und $\mu^-(y_i) > 0$ gelten. Der Ausgangsgrößenwert y_i wird dann nicht empfohlen, sondern ausschließlich verboten. Für die Bestimmung des scharfen Ausgangsgrößenwertes y_D eines Mamdani-Systems im Fall positiver und negativer Regeln sind verschiedene Strategien entwickelt worden. Standardmäßig wird das resultierende Empfehlungsgebirge $\mu(y)$ zunächst in Teilempfehlungsgebirge $\mu_i(y)$ zerlegt, die jeweils in den Intervallen erklärt sind, in denen $\mu(y) > 0$ ist. Jedem Teilempfehlungsgebirge wird ein Gewicht g_i zugeordnet, das beispielsweise die eingeschlossene Fläche unter $\mu_i(y)$ oder ein anderes charakteristisches Merkmal berücksichtigt. Das Teilempfehlungsgebirge $\mu_j(y)$ mit dem größten Gewicht g_j wird ausgewählt und auf herkömmliche Weise defuzzifiziert (zum Beispiel mit der Schwerpunkt-Methode oder der Maximum-Methode). Abbildung 8.4 greift das Beispiel aus Abbildung 8.2 bzw. Abbildung 8.3 auf: Aus dem resultierenden Empfehlungsgebirge $\mu(y)$, das sich durch die Anwendung der Weak-Veto-Hyperinferenz ergibt (siehe Abbildung 8.3, rechts), wird das Teilempfehlungsgebirge $\mu_1(y)$ isoliert, das die größte Fläche umschließt. Abbildung 8.4 zeigt das Ergebnis der Defuzzifizierung für die Maximum-Methode und die Schwerpunkt-Methode. Da das Gewicht g_i hier die eingeschlossene Fläche unter $\mu_i(y)$ beschreibt, entspricht letztere Strategie im Prinzip der Wirkungsweise der auch für herkömmliche Mamdani-Systeme bekannten *Center Of Largest Area*-Defuzzifizierung.

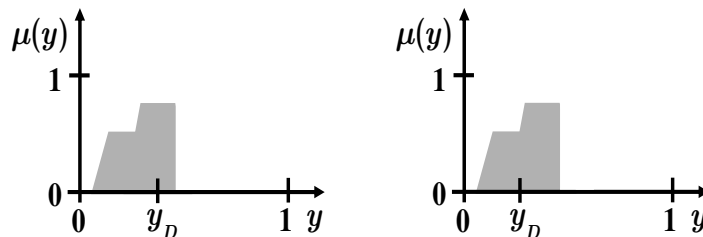


Abbildung 8.4: Scharfer Ausgangsgrößenwert y_D nach MOM-Defuzzifizierung (links) und COG-Defuzzifizierung (rechts) für das Beispiel aus Abbildung 8.3.

8.2 Zur Übertragbarkeit des Prinzips negativer Regeln auf TSK-Systeme

Gesucht wird nach Möglichkeiten zur Übertragung des in Abschnitt 8.1 beschriebenen Prinzips negativer Mamdani-Regeln auf TSK-Systeme. Die Aufgabe negativer Regeln – unabhängig welchen Typs – besteht darin, bei einem gegebenen Eingangsgrößenvektor \mathbf{x} vor bestimmten Ausgangsgrößenwerten y zu warnen bzw. diese zu verbieten. Nachfolgend wird auf die Schwierigkeiten bei der angestrebten Übertragung eingegangen.

Der Ansatz, negative Mamdani-Regeln in herkömmliche positive TSK-Regel transformieren und die Ausgabe des TSK-Systems in der bekannten Weise berechnen zu wollen, ist aus offensichtlichen Gründen ungeeignet. Das Problem liegt darin, dass die Konklusionen entsprechender Regeln, deren Intention das Verbot bestimmter Ausgangsgrößenwerte ist, in Abhängigkeit von den Konklusionen der positiven Regeln zu wählen sind, die Empfehlungen abgeben. Es muss zunächst festgestellt werden, welche positiven Regeln durch den Eingangsgrößenvektor \mathbf{x} aktiviert werden und welchen Ausgabewert diese Regeln liefern. Erst dann können die Konklusionen der ursprünglich negativen Regeln so gewählt werden, dass sich bei deren Hinzunahme ein bestimmter Ausgabewert gerade nicht einstellt. Der Beitrag der ursprünglich negativen Regeln zu dem Ausgabewert des TSK-Systems muss im Prinzip auf alle aktivierten Regeln dieser Art aufgeteilt werden. Es ergibt sich so eine Vielzahl von Entwurfsfreiheiten, wodurch wiederum die Transparenz des TSK-Systems stark beeinträchtigt wird.

In Anlehnung an das zweisträngige Konzept für Mamdani-Systeme kommt folgender mögliche Ansatz in Frage: Man transformiert positive wie negative Mamdani-Regeln in der bekannten Weise und vermerkt zu den so erzeugten TSK-Regeln, ob sie als positiv oder negativ anzusehen sind. Die Empfehlungen bzw. Warnungen $f_i(\mathbf{x})$ der einzelnen TSK-Regeln können auch als ausgangsseitige Singletons (siehe Abschnitt 3.3) interpretiert werden. Es ergeben sich so zwei ausgangsseitige Zugehörigkeitsfunktionen $\mu^+(y)$ und $\mu^-(y)$ in Form von Singletons, auf die die Hyperinferenz in der bekannten Weise angewandt werden kann, um die resultierende ausgangsseitige Zugehörigkeitsfunktion $\mu(y)$ zu erhalten. Dieser Ansatz hat einen gravierenden Nachteil: Bei der Aktivierung einer derartigen negativen TSK-Regel wird nur ein bestimmter Ausgangsgrößenwert verboten, nicht aber – wie es die zu Grunde liegende negative Mamdani-Regel tut und wie es sinnvoll ist – auch benachbarte Ausgangsgrößenwerte. Das grundlegende Problem besteht darin, dass mit der Transformation einer positiven Mamdani-Regel in eine korrespondierende TSK-Regel eine Erhöhung der Approximationsgenauigkeit angestrebt wird, sich dieser Gedanke aber nicht analog auf die Transformation einer negativen Mamdani-Regel übertragen lässt.

8.3 Ansatz für ein hybrides Fuzzy-System

Zur Nutzung negativer Mamdani-Regeln in Verbindung mit TSK-Systemen wird ein hybrides Fuzzy-System [SKK05, SKKS06] vorgeschlagen, das in Abbildung 8.5 schematisch dargestellt ist. Ziel ist es, die Vorteile von negativen Mamdani-Regeln (kompakte und transparente Formulierung von Warnungen bzw. Verboten) und die Vorteile von TSK-Regeln (hohe Approximationsgenauigkeit) innerhalb eines Systems nutzen zu können. Anwender sollen dieses Fuzzy-System in der gleichen Weise erstellen und einsetzen können wie die ihnen bereits bekannten Mamdani-Systeme und TSK-Systeme. Das in dieser Arbeit konzipierte hybride Fuzzy-System beinhaltet sowohl ein Mamdani-System als auch ein TSK-System. Die Regelbasis des Mamdani-Systems setzt sich aus positiven und negativen Regeln zu-

sammen. Die Regelbasis des TSK-Systems beinhaltet hingegen nur positive Regeln, die durch Transformation aus den positiven Regeln des Mamdani-Systems hervorgehen. Für einen gegebenen Eingangsgrößenvektor \mathbf{x} wird zunächst die Mamdani-Regelbasis nach dem zweisträngigen Konzept (siehe Abschnitt 8.1) ausgewertet. Hieraus ergeben sich die ausgangsseitigen Empfehlungsgebirge $\mu^+(y)$ für die Empfehlungen der positiven Regeln und $\mu^-(y)$ für die Warnungen oder Verbote der negativen Regeln. Nach Durchführung der Hyperinferenz ergibt sich für das Mamdani-System das resultierende Empfehlungsgebirge $\mu(y)$, das aus mehreren Teilempfehlungsgebirgen $\mu_i(y)$ bestehen kann, die jeweils in den Intervallen $[y_{i,min}, y_{i,max}]$ erklärt sind, in denen $\mu(y) > 0$ ist. Aus $\mu(y)$ wird das am stärksten gewichtete Teilempfehlungsgebirge $\mu_j(y)$ und das zugehörige Intervall $[y_{j,min}, y_{j,max}]$, in dem dieses definiert ist, bestimmt. Dieses Intervall ist, bezogen auf den gegebenen Eingangsgrößenvektor \mathbf{x} , der prädestinierte Wertebereich – das heißt zulässig im Sinne der Verbote und am stärksten empfohlen im Sinne der Empfehlungen – für den scharfen Ausgangsgrößenwert y_D . Der abschließende Schritt der Hyperdefuzzifizierung wird für das Mamdani-System jedoch ausgesetzt. Für die Berechnung des Ausgangsgrößenwertes y_D des hybriden Fuzzy-Systems wird das TSK-System herangezogen. Der anhand des Mamdani-Systems bestimmte prädestinierte Wertebereich für den scharfen Ausgangsgrößenwert wird in das TSK-System eingekoppelt und dort für eine situationsbezogene Regelselektion genutzt. Aus denjenigen TSK-Regeln, die durch den Eingangsgrößenvektor \mathbf{x} aktiviert werden, werden nur diejenigen TSK-Regeln, für die $y_{j,min} \leq f_i(\mathbf{x}) \leq y_{j,max}$ gilt, ausgewählt und für die Berechnung des Ausgabewertes herangezogen. Die Berechnung des scharfen Ausgabewertes anhand der so selektierten TSK-Regeln erfolgt dann in der gewohnten Weise durch eine gewichtete Mittelwert- oder Medianbildung (siehe Gl. (3.13) und Gl. (7.22)). Da alle Empfehlungen der so selektierten positiven TSK-Regeln innerhalb des gewünschten Wertebereichs liegen, gilt dies auch für den daraus berechneten Ausgabewert y_D des TSK-Systems, der gleichzeitig auch der scharfe Ausgangsgrößenwert des hybriden Fuzzy-Systems ist.

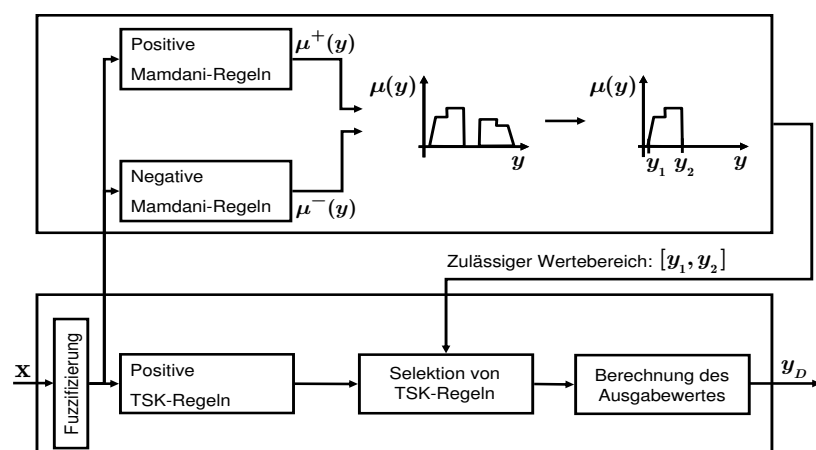


Abbildung 8.5: Schematische Darstellung eines hybriden Fuzzy-Systems.

8.4 Anwendungsbeispiel

Der prinzipielle Nutzen des hybriden Fuzzy-Systems wird im Folgenden an einem Beispiel illustriert. Es wird dabei der hypothetische Fall betrachtet, dass die Aktionen eines Prozessbedieners beobachtet und durch ein Fuzzy-Modell nachgebildet werden. Der Eingriff des Experten in den Prozess wird im Folgenden mit Y bezeichnet und richtet sich nach zwei Prozessgrößen X_1 und X_2 . Dabei entspricht der Wert $x_1 = 0$ bzw. $x_2 = 0$ dem gewünschten Sollwert einer Prozessgröße. Abweichungen von den gewünschten Sollwerten werden durch geeignete Eingriffe in den Prozess korrigiert. Durch Beobachtung des Experten wird zunächst ein Datensatz gewonnen, der die Grundlage für die nachfolgende Modellierung bildet. Durch Interpolation wird aus den so erhobenen Messdaten ein Kennfeld erzeugt, das die Bedienstrategie des Experten repräsentiert (siehe Abbildung 8.6). Anhand des Kennfeldes zeigt sich, dass der Experte bei gleichzeitig betragsmäßig kleinen Werten beider Prozessgrößen mit einem betragsmäßig geringen Eingriff reagiert. Für $-1.5 \leq x_1 \leq 1.5$ ist der Eingriff des Experten unabhängig von dem jeweiligen Wert für x_2 ebenfalls vergleichsweise gering und für $-1.5 \leq x_2 \leq 1.5$ nahezu unabhängig von dem jeweiligen Wert für x_1 . Die Steigung des Kennfeldes in x_2 -Richtung ist in diesem Fall aber deutlich stärker als die in x_1 -Richtung im zuvor beschriebenen Fall.

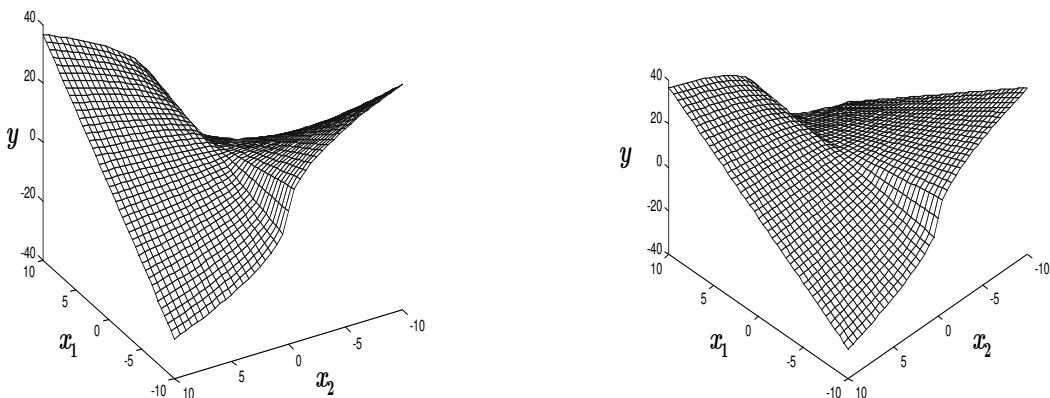


Abbildung 8.6: Aus den durch Beobachtung des Experten gewonnenen Daten interpoliertes Kennfeld unter verschiedenen Blickwinkeln.

Die wesentlichen Eigenschaften des Kennfeldes sollen sich in dem zugehörigen Kennfeld eines wie auch immer gearteten Fuzzy-Modells zeigen. Anhand der erhobenen Messdaten wird ein solches Fuzzy-Modell erstellt; gleichzeitig sollen aber auch Modifikationen hieran ermöglicht werden. Für das Regelungsverhalten ist es oftmals von Vorteil, bei kleineren Regelabweichungen gar nicht in einen Prozess einzugreifen, um so ein unruhiges Prozessverhalten zu vermeiden. Eine Maßnahme, um dies zu erreichen, ist das Einfügen einer so genannten toten Zone in einen Kennfeldfeldregler, beispielsweise eines Zweipunktreglers.

Das von dem Fuzzy-Modell erzeugte Kennfeld soll daher in der Art von dem ursprünglichen Kennfeld abweichen, dass es eine tote Zone für betragsmäßig kleine Sollwertabweichungen aufweist.

Für die datenbasierte Fuzzy-Modellierung werden für jede Eingangs- und die Ausgangsgröße jeweils fünf linguistische Werte vorgesehen. Die Zugehörigkeitsfunktionen hierzu werden äquidistant über den jeweiligen Wertebereich platziert und überlappen sich in einem gewissen Grad. Es handelt sich dabei um eine gewöhnliche Standard-Partitionierung, da hier nicht in erster Linie die Genauigkeit der Nachbildung, sondern die prinzipielle Leistungsfähigkeit negativer Regeln in Verbindung mit TSK-Systemen von Interesse ist. In Abbildung 8.7 ist die Partitionierung der Eingangsgrößen X_1 und X_2 dargestellt. Zusätzlich sind die Ausgangsgrößenwerte der Datenpunkte in Abhängigkeit von dem Wert der jeweiligen Eingangsgröße dargestellt und gemäß des Ausgangsgrößenwertes farblich codiert. Es ist zu erkennen, dass bei alleiniger Betrachtung der Eingangsgröße X_1 die Ausgangsgrößenwerte der Datenpunkte, die die Prämisse einer entsprechenden generalisierenden Hypothese in einem Grade größer Null erfüllen, über einen großen Wertebereich streuen. Folglich werden bei der Regelsuche keine entsprechenden generalisierenden Regel erzeugt. Gleiches gilt für generalisierende Hypothesen, die sich auf die Eingangsgröße X_2 beziehen.

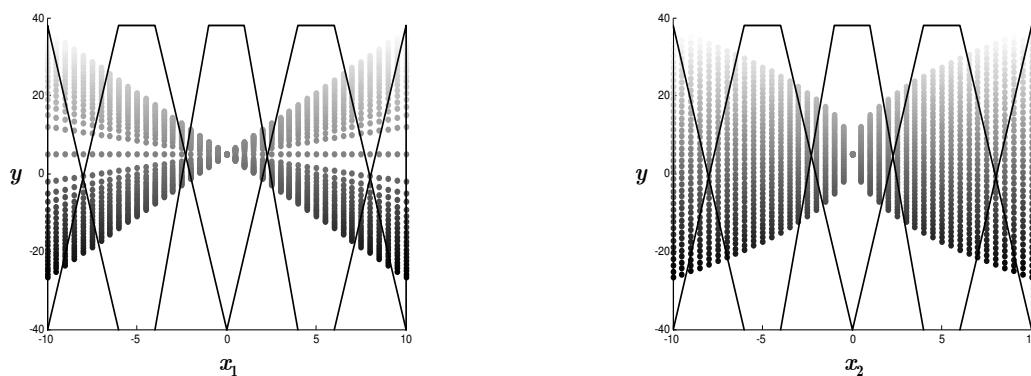


Abbildung 8.7: Partitionierungen der Eingangsgrößen X_1 (links) und X_2 (rechts).

Werden für dieses Beispiel beide Eingangsgrößen in der Prämisse betrachtet, so handelt es sich um vollständige Hypothesen bzw. Regeln. Bei der gewählten Partitionierung sind 25 vollständige Hypothesen zu prüfen, die in Abbildung 8.8 schematisch dargestellt sind. Zur besseren Übersicht wird auf die Wiedergabe der Überlappungen der Zugehörigkeitsfunktionen verzichtet, stattdessen werden die Schnittpunkte der Überlappungen als Stützstellen für die Darstellung verwendet. Für die meisten dieser 25 Hypothesen gelingt es, so eine Gruppe von Datenpunkten zu separieren, die sich von der Gesamtheit der Datenpunkte signifikant unterscheidet, beispielsweise im Hinblick auf den mittleren Ausgangsgrößenwert. Daher werden in 16 der 25 Fälle Regeln generiert. Es zeigt sich aber auch, dass für neun Hypothesen die Ausgangsgrößenwerte der entsprechend relevanten Datenpunkte ähnlich

stark streuen wie für die generalisierenden Hypothesen. Dies sind die Hypothesen, in denen mindestens ein Eingangsgrößenwert einen betragsmäßig kleinen Wert annimmt. Diese Hypothesen werden daher bei der Regelsuche zurückgewiesen. Um auch für diese Bereiche Regeln zu erzeugen, müsste die Partitionierung hier entsprechend feiner gewählt werden.

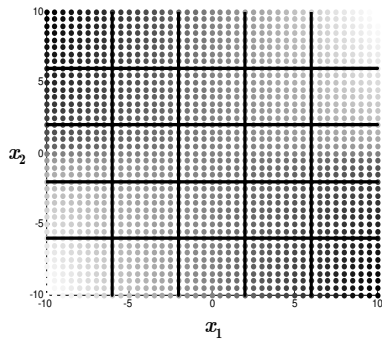


Abbildung 8.8: Schematische Darstellung der Suchraumstrukturierung für vollständige Regeln.

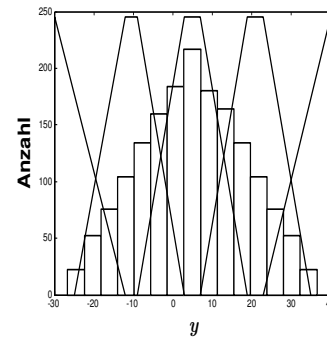


Abbildung 8.9: Partitionierung der Ausgangsgröße Y .

Für die oben beschriebene Partitionierung wird ein Fuzzy-Modell vom Typ Mamdani mit nur positiven Regeln erzeugt, dessen Regelbasis 16 vollständige Regeln aufweist. Falls durch einen Eingangsgrößenvektor \mathbf{x} keine Regel aktiviert wird, wird durch eine Default-Regel der mittlere Ausgangsgrößenwert $\hat{y} = 5$ der zu Grunde liegenden Datenpunkte empfohlen. Das resultierende Kennfeld des Mamdani-Modells ist in Abbildung 8.10 (links) dargestellt. Es weist ein sehr stufenartiges Verhalten auf und entspricht nur sehr bedingt dem gewünschten Kennfeld aus Abbildung 8.6. Um die Genauigkeit der Approximation zu verbessern, wird das Mamdani-Modell in ein korrespondierendes TSK-Modell transformiert. Dessen Kennfeld, dargestellt in Abbildung 8.10 (rechts), nähert sich in den äußeren Bereichen dem ursprünglichen Kennfeld recht genau an. Deutlich zu erkennen sind auch hier die Bereiche, in denen die oben erwähnte Default-Regel greift.

Die gewünschte tote Zone für kleine Sollwertabweichungen wird durch zwei negative Mamdani-Regeln realisiert. Um eine große Flexibilität in der Gestaltung der toten Zone zu erreichen, wird sowohl für jede Eingangsgröße als auch für die Ausgangsgröße jeweils ein zusätzlicher linguistischer Wert *Totzone* eingeführt². Durch die generalisierenden negativen Mamdani-Regeln

$$\text{WENN } \langle x_1 = \text{Totzone} \rangle \text{ DANN NICHT } \langle y = \text{Totzone} \rangle$$

$$\text{WENN } \langle x_2 = \text{Totzone} \rangle \text{ DANN NICHT } \langle y = \text{Totzone} \rangle$$

lässt sich die tote Zone in dem resultierenden Kennfeld einfach und nachvollziehbar realisieren. Deren Ausgestaltung lässt sich durch Änderungen an den Zugehörigkeitsfunktionen

²In diesem Beispiel wird die Strong-Veto-Hyperinferenz verwendet, daher sind nur die zwei äußeren der insgesamt vier Stützstellen einer Zugehörigkeitsfunktion für die Breite der toten Zone entscheidend.

der linguistischen Werte leicht anpassen. Wird beispielsweise für die Eingangsgröße X_1 die Zugehörigkeitsfunktion über den Wertebereich verbreitert, so ändert sich damit auch die Breite der toten Zone in x_1 -Richtung in dem resultierenden Kennfeld, sofern der linguistische Ausgangsgrößenwert einen entsprechend großen Wertebereich umschließt. In Abbildung 8.11 (links) ist die Breite der toten Zone in x_1 - und x_2 -Richtung gleich groß gewählt, wohingegen in Abbildung 8.11 (rechts) die Breite der toten Zone in x_2 -Richtung doppelt so groß gewählt ist wie in x_1 -Richtung. In den nicht durch die negativen Regeln berührten Bereichen bleibt die Approximationsgenauigkeit des ursprünglichen Kennfeldes durch das Fuzzy-Modell aufgrund des Einsatzes von TSK-Regeln unverändert erhalten.

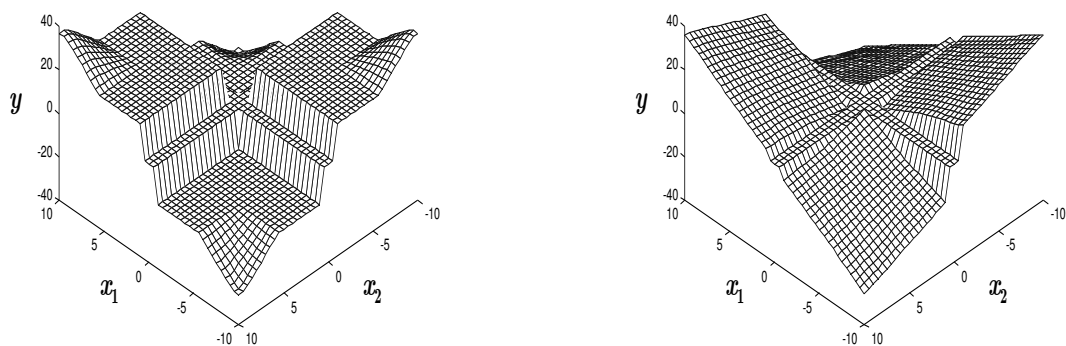


Abbildung 8.10: Kennfeld des datenbasiert generierten Mamdani-Modells (links) und des durch Transformation daraus hervorgehenden TSK-Modells (rechts).

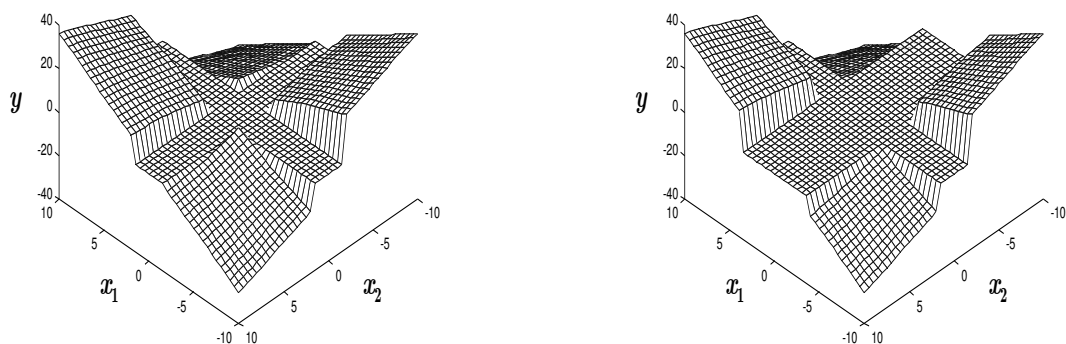


Abbildung 8.11: Kennfelder zweier hybrider Fuzzy-Module mit unterschiedlichen Ausgestaltungen der toten Zone.

Kapitel 9

Einsatz für komplexe Anwendungen und Ableitung von Einstellregeln

Dieses Kapitel beschäftigt sich mit der systematischen Exploration der Einstellmöglichkeiten des Fuzzy-ROSA-Verfahrens und der Ableitung systematischer Lösungsstrategien daraus [SKKS06, SKK06]. In Abschnitt 9.1 wird zunächst detailliert auf die verschiedenen Aspekte eingegangen, die diese Untersuchungen motivieren. Von übergeordnetem Interesse ist dabei, das Verfahren potenziellen Anwendern in der Weise zugänglich und nutzbar zu machen, dass sie es ohne größere Einarbeitung und ohne tiefer gehende Detailkenntnisse für ihre Zwecke erfolgreich einsetzen können. In Abschnitt 9.2 werden die verschiedenen Testszenerarien erläutert, die der Exploration der Einstellmöglichkeiten zu Grunde liegen. In Abschnitt 9.3 werden für die untersuchten Anwendungen und Benchmarkprobleme¹ erzielte Modellierungsergebnisse aufgeführt und diskutiert. Aufgrund des Umfangs der durchgeführten Untersuchungen können nicht alle Modellierungsergebnisse im Detail aufgeführt werden. Die Ergebnisse zeigen die Leistungsfähigkeit der in dieser Arbeit entwickelten Erweiterungen des Verfahrens und die damit erzielten Verbesserungen. In Abschnitt 9.4 werden die Modellierungsergebnisse zusammenfassend beschrieben. Aufsetzend auf den Ergebnissen der durchgeführten Untersuchungen wird die Entwicklung systematischer Lösungsstrategien angestrebt. In Abschnitt 9.5 werden Einstellempfehlungen für industrielle Anwender vorgestellt, die durch die Analyse der durch die Exploration gewonnenen Daten wissensbasiert formuliert werden können. Die so formulierten Einstellempfehlungen stellen ein Zwischenziel hinsichtlich der Entwicklung einer vollautomatisierten Vorgehensweise dar. Abschnitt 9.6 beschäftigt sich damit, wie auf der Basis der durchgeführten Untersuchungen die datenbasierte Generierung eines Moduls erfolgt, das problembezogene Prognosen der zu erwartenden Modellierungsfehler für unterschiedliche Einstellungen freier Modellparameter liefert.

¹Die Benchmarkprobleme werden in Anhang B vorgestellt.

9.1 Motivation und Zielsetzungen

Für die Akzeptanz des Fuzzy-ROSA-Verfahrens in der Industrie ist entscheidend, dass es auch von weniger versierten Anwendern ohne großen Aufwand zur Lösung komplexer Probleme eingesetzt werden kann. Häufig fordern Anwender, dass die Anwendung des Verfahrens ohne besondere Kenntnisse im Bereich der datenbasierten Fuzzy-Modellierung möglich sein muss. Dies ist auch verständlich, da sich andernfalls der Aufwand für die Modellerstellung insgesamt nicht verringert, sondern nur von einer Stelle – der intensiven Auseinandersetzung mit dem zu modellierenden System – an eine andere Stelle – der intensiven Auseinandersetzung mit dem eingesetzten Modellierungsverfahren – verlagert. Oft wird eine *Ein-Knopf-Lösung* für den Einsatz eines Verfahrens gefordert. Um dennoch eine hohe Flexibilität zu bieten, aber eine Überladung der Anwender mit Einstelloptionen zu vermeiden, ist es häufig sinnvoll, den Programmkern des Verfahrens automatisch aufzurufen, statt die Anwender mit dieser Aufgabe zu konfrontieren. Dennoch wird meist gefordert, dass die Wahl der Einstellwerte problemabhängig erfolgt. Deshalb kommt der Entwicklung einer Systematik zur Festlegung der Einstellwerte eine sehr große Bedeutung zu.

Mit der im Rahmen dieser Arbeit durchgeführten systematischen Exploration der Einstellmöglichkeiten des Fuzzy-ROSA-Verfahrens werden verschiedene, teilweise aufeinander aufbauende Zielsetzungen verfolgt. Ein Ziel besteht darin, die prinzipielle Leistungsfähigkeit des Verfahrens und der hierfür neu entwickelten Strategieelemente an unterschiedlichen Benchmarkproblemen zu erproben. Ein weiterführendes Ziel besteht darin, anhand der so gewonnenen Datenbasis qualitative Zusammenhänge zwischen Kenngrößen der Modellierungsaufgaben, den vorgenommenen Einstellungen und den damit erzielten Resultaten aufzudecken. Daraus werden problemspezifische Empfehlungen für die Einstellwerte freier Parameter abgeleitet, die industriellen Anwendern helfen, das Verfahren zur Lösung komplexer praktischer Probleme erfolgreich zu nutzen, ohne selbst über Detailwissen zum Verfahren verfügen zu müssen. Ein noch anspruchsvolleres Ziel ist es, die vermuteten Zusammenhänge durch das Verfahren selbst in Form von Regeln automatisch aufzufinden und daraus ein Modul zu generieren, das anhand von Kenngrößen eines gegebenen neuen Modellierungsproblems für jede Wahl der Einstellparameter den zu erwartenden Modellierungsfehler prognostiziert. Damit können vor der eigentlichen Modellgenerierung unterschiedliche Parametereinstellungen simulativ untersucht und eine möglichst gute Parametereinstellung ausgewählt werden, um so den Prozess der Modellbildung insgesamt zu beschleunigen oder vollständig zu automatisieren. Mit einem entsprechenden Modul können auch solche Parametereinstellungen simulativ bewertet werden, die im Rahmen der vorherigen Exploration noch gar nicht untersucht worden sind. Bei entsprechender Prognosegenauigkeit kann das Modul somit auch für eine Optimierung der Parametereinstellungen genutzt werden. Für die Optimierung können Verfahren aus dem Bereich der Computational Intelligence eingesetzt werden, beispielsweise Evolutionsstrategien [Rec94, Sch95] oder Partikel-Schwarm-Optimierungsverfahren [KE95, KSSH05, KSS06].

9.2 Experiment-Design

Ein Modell dient dazu, ein auf sein Eingangs-/Ausgangsverhalten abstrahiertes System nachzubilden. Der wahre Zusammenhang $y = f(\mathbf{x})$ zwischen den Eingangsgrößen und der Ausgangsgröße sei unbekannt. Bei einer datenbasierten Modellierung wird anhand einer begrenzten Stichprobe von protokollierten Daten ein wie auch immer geartetes Modell generiert, welches in der Regel über eine Reihe frei wählbarer Parameter verfügt, die eine möglichst gute Anpassung des Modells an das zu lösende Modellierungsproblem ermöglichen sollen. Es stellen sich zwei Fragen: Wie sollen die freien Parameter des Modells eingestellt und wie kann die Anpassung des Modells an das Problem bewertet werden? Mit diesen grundsätzlichen Fragestellungen befassen sich die zwei nachfolgenden Abschnitte. Im Anschluß daran wird auf die Einstellmöglichkeiten des Fuzzy-ROSA-Verfahrens eingegangen. Dieser Begriff ist in dieser Arbeit sehr weit gefasst. Grundsätzlich fallen hierunter alle Freiheiten, die sich einem Anwender bei dem Einsatz des Verfahrens für die Modellierung eines Problems bieten. In Abschnitt 4.2 werden diese unter dem Begriff der Projektdefinition vorgestellt. Unter gewissen Annahmen bleiben für industrielle Anwender zwei zentrale Entwurfsvollfreiheiten: die Festlegung der Partitionierung und die Wahl eines Test- und Bewertungsverfahrens sowie die Einstellung dessen freier Parameter. Diese beiden Aspekte werden jeweils genauer beleuchtet. Daran anschließend wird auf die Menge möglicher Kombinationen aus Partitionierung und Wahl bzw. Einstellung eines Test- und Bewertungsverfahrens eingegangen. Aufgrund der kombinatorischen Vielfalt können nicht alle Varianten untersucht werden, sondern es müssen gewisse Einschränkungen vorgenommen werden. Die in dieser Arbeit untersuchten Einstellmöglichkeiten lassen sich in zwei übergeordnete Gruppen gliedern, die genauer vorgestellt werden. Die erzielten Modellierungsergebnisse werden in Abschnitt 9.4 zusammenfassend beschrieben und diskutiert.

Grundsätzliche Überlegungen zur Bestimmung der Güte eines Modells

Die *wahre* Güte eines Modells ergibt sich, wenn an dem zu modellierenden System unendlich viele Daten erfasst und dann zum Testen an dem zuvor generierten Modell verwendet werden. In der Praxis lässt sich dies aus verschiedenen Gründen so aber nicht durchführen. Diese Überlegung zeigt jedoch, dass das letztlich entscheidende Kriterium zur Beurteilung der Güte das Verhalten des Modells für solche Daten ist, die *nicht* für dessen Generierung verwendet werden. Nur so kann festgestellt werden, ob das Modell tatsächlich die Wirkungszusammenhänge des zu modellierenden Systems erfasst oder lediglich die zu Grunde liegenden Lerndaten auswendig gelernt hat.

Da in der Praxis nur eine begrenzte Stichprobe an Daten zur Verfügung steht, behilft man sich häufig damit, dass man diese Stichprobe in eine *Lernstichprobe* und eine *Teststichprobe* aufteilt. Auf der Lernstichprobe wird ein Modell erstellt und dann anhand der Teststichprobe validiert. Um die *tatsächliche* Güte eines Modells zu ermitteln, müsste dieser Vorgang

bei gleichbleibender Größe der Lernstichprobe unendlich oft wiederholt werden. Auch hierbei handelt es sich um eine eher theoretische Überlegung, da dieser Vorgang in der Praxis viel zu zeitaufwendig wäre.

Um die tatsächliche Güte eines Modells möglichst gut zu schätzen, werden oftmals Methoden der Kreuzvalidierung und der mehrfachen Wiederholung eingesetzt. Unter einer N_R -fachen Kreuzvalidierung wird hier verstanden, dass die Gesamtstichprobe der Daten zufällig in $N_R \geq 2$ möglichst gleich große Teilstichproben geteilt wird. Es werden dann jeweils $N_R - 1$ dieser Teilstichproben zu einer Lernstichprobe zusammengefasst und damit ein Modell generiert, das auf der verbliebenen Teilstichprobe (Teststichprobe) getestet wird. Unter einer N_R -fachen Wiederholung wird hier verstanden, dass die Gesamtstichprobe der Daten N_R mal zufällig in Lern- und Teststichprobe geteilt und für jede Teilungsvariante entsprechend ein Modell gelernt und validiert wird.

Grundsätzliche Überlegungen zur Einstellung freier Modellparameter

In Anbetracht der vorhergehenden Überlegungen zur Beurteilung der tatsächlichen Güte eines Modells stellt sich die Frage, wie sich die Vorgehensweise zur Einstellung dessen freier Parameter gestalten sollte. Diese Frage führt zu dem von SALZBERG [Sal99] beschriebenen Problem des *Repeated Tuning*. Der zentrale Aspekt liegt dabei auf der Forderung, alle Einstellungen oder Modifikationen eines Modells vorzunehmen, *bevor* es zur abschließenden Validierung auf Testdaten kommt. Diese Vorgehensweise erscheint plausibel, wird jedoch oftmals nicht konsequent verfolgt. Als Beispiel werde folgendes Szenarium angenommen: Die vorhandenen Daten seien in eine Lern- und eine Teststichprobe geteilt worden, und mithilfe der Lernstichprobe werde ein Modell erstellt. Die freien Parameter des Modells seien dabei wie auch immer gewählt worden. Das so erstellte Modell wird nun auf der Teststichprobe validiert. Ausgehend von dem so erzielten Ergebnis auf der Teststichprobe werde nun die Einstellung der Modellparameter geändert und das Modell abermals auf Grundlage der Lernstichprobe erstellt. Dies geschieht mit dem Ziel, das Ergebnis auf der Teststichprobe zu verbessern. Diese Vorgehensweise verletzt die oben beschriebene Forderung, da die Teststichprobe (mittelbar oder unmittelbar) zur Anpassung der Parameter verwendet wird. Die Lösung dieses Problems liegt darin, Kreuzvalidierungen oder Wiederholungen in der Lernstichprobe selbst zu verwenden. Dieser Ansatz wird auch als *harte Validierung* bezeichnet. Als Beispiel hierfür werde folgendes Szenarium angenommen: Die vorhandenen Daten seien in eine Lern- und eine Teststichprobe geteilt worden. Die Lernstichprobe werde nun abermals geteilt in die Tuning-Lernstichprobe und die Tuning-Teststichprobe. Es seien mehrere Standard-Einstellungen für die freien Parameter eines Modells definiert. Diese Einstellungen werden auch als Standard-Parametrierungen bezeichnet. Für jede der Standard-Parametrierungen wird auf der Tuning-Lernstichprobe ein Modell erzeugt und dann auf der Tuning-Teststichprobe validiert. Das für eine Standard-Parametrierung erzielte Ergebnis wird in einer Liste gespeichert. Nachdem alle Standard-Parametrierungen untersucht sind, wird anhand der Liste diejenige ausgewählt, die das beste Ergebnis auf der Tuning-

Teststichprobe erzielt. Für das zugehörige Modell erfolgt dann abschließend die Validierung auf der bis dahin zurückgehaltenen Teststichprobe. Die Vorgehensweise der harten Validierung ist mit einem entsprechenden Aufwand verbunden. Eine manuelle Durchführung der beschriebenen Untersuchungen ist zu aufwendig und auch fehleranfällig. Daher muss die eingesetzte Software so konzipiert sein, dass eine automatisierte Durchführung und Auswertung der Experimente erfolgt. Ferner müssen Standard-Parametrierungen definiert und ein automatisiertes Parametertuning realisiert werden. Andere potenzielle Probleme, wie beispielsweise die Gefahr des Overfittings eines Modells, können auch durch diesen Ansatz nicht ausgeschlossen werden. Die ausgewählte Standard-Parametrierung ist zwar für die Tuning-Teststichprobe *optimal*, daraus folgt aber nicht zwangsläufig, dass dies auch für die Teststichprobe zutrifft.

Vorarbeiten

Als Orientierung für die in dieser Arbeit durchgeführten Untersuchungen dienen Vorarbeiten [Sla01], die im Rahmen des Projektes B1 des Sonderforschungsbereichs (SFB) 531 durchgeführt wurden. Dort wurde folgende systematische Vorgehensweise zur Einstellung freier Modellparameter vorgeschlagen und anhand von Benchmarkproblemen getestet: Für ein gegebenes Modellierungsproblem wird zunächst eine als geeignet angesehene Partitionierung der Größen vorgenommen (Suchraumstrukturierung). Hierfür wird die Anzahl der linguistischen Werte je Eingangs- bzw. Ausgangsgröße wissensbasiert festgelegt. Die Erzeugung der Zugehörigkeitsfunktionen zu den linguistischen Werten erfolgt daraufhin datenbasiert. Ausgehend von der so erzeugten Suchraumstrukturierung werden mit vier unterschiedlichen Test- und Bewertungsverfahren (J_{RI} , J_{RHR} , J_{CHR} und J_{MVB}) Modelle generiert. Jedes dieser Verfahren verfügt über einen Parameter κ , der die Schwere des jeweiligen Regeltests steuert. Ein weiterer Parameter τ steuert, welche der generierten Regeln aufgrund ihrer Relevanz tatsächlich in die Regelbasis übernommen werden. Der Parameter κ wird für jedes der Test- und Bewertungsverfahren auf einem hierfür typischen Standardwert belassen oder aber aufgrund von Zusatzwissen problemspezifisch angepasst. Für den Parameter τ werden für jedes Test- und Bewertungsverfahren die Werte $\tau_1 = 0.0$, $\tau_2 = 0.1$, \dots , $\tau_{10} = 0.9$, $\tau_{11} = 1.0$ vorgesehen. Die maximal zulässige Kombinationstiefe in einer Prämisse beträgt $c_{max} = 2$. Es werden ausschließlich Modelle vom Typ Mamdani mit nur positiven Regeln untersucht. Systematische Untersuchungen für TSK-Modelle, die durch Transformation von Mamdani-Modellen mit ausschließlich positiven Regeln entstehen, finden sich in [Kra02].

Diese Untersuchungen haben wichtige Zusammenhänge zwischen den angestrebten Zielen Relevanz und Treffsicherheit von Regeln sowie der Approximationsgüte des Modells einerseits und der Wahl und Parametrierung des Test- und Bewertungsverfahrens andererseits aufgedeckt. Es fehlen aber Erkenntnisse darüber, welches Test- und Bewertungsverfahren und welche Werte für κ und τ im konkreten Anwendungsfall zu wählen sind. Hierauf zielen die folgenden Untersuchungen ab.

Beschreibung der durchgeführten Experimente

Die datenbasierte Modellierung mithilfe des Fuzzy-ROSA-Verfahrens gliedert sich in vier Teilaufgaben: Auswahl der Eingangsgrößen, Strukturierung des Suchraumes, Regelgenerierung und Optimierung der Regelbasis. Im Rahmen der systematischen Exploration der Einstellmöglichkeiten wird auf eine vorgeschaltete Selektion von Eingangsgrößen zum einen verzichtet, um eine bessere Vergleichbarkeit mit vorherigen Modellierungsergebnissen zu ermöglichen, zum anderen auch deshalb, weil industrielle Anwender meist zunächst alle verfügbaren Eingangsgrößen berücksichtigen möchten. Auf eine nachgeschaltete Optimierung der Regelbasis wird ebenfalls verzichtet, weil diese so zeitintensiv ist, dass sie bei der großen Anzahl der hier erzeugten Modelle nicht in angemessener Zeit durchführbar ist. Die nachfolgende systematische Untersuchung erstreckt sich somit auf die Strukturierung des Suchraumes und die Regelgenerierung.

Die Strukturierung des Suchraumes besteht in der Festlegung der Anzahl der linguistischen Werte je Eingangs- bzw. Ausgangsgröße sowie der Festlegung der Zugehörigkeitsfunktionen für die Modellierung der linguistischen Werte. Im Prinzip kann die Festlegung von linguistischen Werten und Zugehörigkeitsfunktionen für jede Eingangs- und die Ausgangsgröße individuell erfolgen. Die daraus resultierende kombinatorische Vielfalt impliziert sehr viele Freiheitsgrade. In praktischen Anwendungen, in denen kein spezielles Problemwissen vorliegt, hat sich die Beschränkung auf gewisse Standard-Partitionierungen als ausreichend erwiesen. Die Anzahl linguistischer Werte wird dabei für alle Größen gleich – typischerweise zu drei, fünf oder sieben – gewählt. Ebenso werden die Zugehörigkeitsfunktionen für alle Größen in gleicher Weise – äquidistant, äquifrequent oder aufgrund eines unidimensionalen Clusterverfahrens – über den Wertebereich einer Größe platziert. Dabei werden rechteckförmige, trapezförmige oder dreieckförmige Zugehörigkeitsfunktionen verwendet¹. Für eine automatische Partitionierung wird das oben erwähnte unidimensionale Clusterverfahren eingesetzt, das auf einem bekannten K -Means Algorithmus [McQ67, HW79] basiert². Für jede Größe werden unterschiedliche Anzahlen an zu verwendenden Clustern vorgegeben. Schwerpunkt und Größe der Cluster werden durch den Algorithmus bestimmt und für jede resultierende Clusterkonfiguration wird ein Gütemaß³ berechnet. Anhand des Gütemaßes wird diejenige Konfiguration ausgewählt, die in diesem Sinne das beste Ergebnis erzielt. Anhand dieser Clusterkonfiguration werden entsprechend linguistische Werte und Zugehörigkeitsfunktionen hierzu erzeugt. Mit den zuvor beschriebenen Ansätzen ergeben sich so 22 Partitionierungsvarianten, die in Tabelle 9.1 zusammengefasst sind⁴.

Was das Test- und Bewertungsverfahren angeht, ist zunächst die diskrete Entscheidung zu treffen, welches Verfahren ausgewählt wird, und dann, welche Werte die zuvor erwähnten Parameter κ und τ haben sollen. Hier werden sieben verschiedene Test- und Bewertungs-

¹Das Clusterverfahren generiert ausschließlich trapezförmige Zugehörigkeitsfunktionen.

²In [KS98] wird ein anderer Ansatz für die datenbasierte Partitionierung vorgeschlagen.

³Das eingesetzte Gütemaß wurde von NuTech definiert.

⁴In Tabelle 9.1 wird Zugehörigkeitsfunktion durch ZGF abgekürzt.

verfahren (J_{RI} , J_{RHR} , J_{NHR} und J_{CHR} , sowie J_{MVB} , J_{VVB} und J_{CVB}) untersucht. Die vier erstgenannten Verfahren arbeiten auf der Grundlage der relativen Datenhäufigkeiten, die übrigen drei anhand der Datenverteilungen. Weitere Details zu den Test- und Bewertungsverfahren finden sich unter anderem in [Kro99, JS98b, Jes00a]. Zu jedem Test- und Bewertungsverfahren werden sieben verschiedene Werte für κ angesetzt⁵. Für τ werden die zehn Werte $\tau_1 = 0.0, \dots, \tau_{10} = 0.9$ verwendet. Es ergeben sich so $7 \cdot 10 = 70$ Parametrierungen für ein gewähltes Test- und Bewertungsverfahren und damit für die sieben Test- und Bewertungsverfahren insgesamt $7 \cdot 70 = 490$ Standard-Parametrierungen, die in Tabelle 9.2 zusammengefasst sind. Die maximal zulässige Kombinationstiefe in einer Prämisse wird wie in [Sla01] auf $c_{max} = 2$ festgelegt.

Anzahl ling. Werte	Platzierung der ZGF	Form der ZGF	Index
3	Äquidistant	Rechteck	1
		Trapez	2
		Dreieck	3
	Äquifrequent	Rechteck	4
		Trapez	5
		Dreieck	6
	Clusterbasiert	Trapez	7
5	Äquidistant	Rechteck	8
		Trapez	9
		Dreieck	10
	Äquifrequent	Rechteck	11
		Trapez	12
		Dreieck	13
	Clusterbasiert	Trapez	14
7	Äquidistant	Rechteck	15
		Trapez	16
		Dreieck	17
	Äquifrequent	Rechteck	18
		Trapez	19
		Dreieck	20
	Clusterbasiert	Trapez	21
Automatisch	Clusterbasiert	Trapez	22

Tabelle 9.1: Übersicht der vorgesehenen Partitionierungen.

⁵Je nach Test- und Bewertungsverfahren bewegen sich die Werte für κ in unterschiedlichen Wertebereichen, da die Verfahren unterschiedlich konzipiert sind (siehe Abschnitt 6.2 und Anhang A). Für die Test- und Bewertungsverfahren J_{RI} , J_{RHR} , J_{MVB} , J_{VVB} und J_{CVB} werden für κ folgende Werte angesetzt: $\kappa_1 = 0.0001$, $\kappa_2 = 0.001$, $\kappa_3 = 0.01$, $\kappa_4 = 0.03$, $\kappa_5 = 0.06$, $\kappa_6 = 0.1$ und $\kappa_7 = 0.2$. Für die Test- und Bewertungsverfahren J_{NHR} und J_{CHR} werden für κ folgende Werte angesetzt: $\kappa_1 = 0.3$, $\kappa_2 = 0.4$, $\kappa_3 = 0.5$, $\kappa_4 = 0.6$, $\kappa_5 = 0.7$, $\kappa_6 = 0.8$ und $\kappa_7 = 0.9$.

Verfahren	Parameter κ	Parameter τ	Index
J_{RI}	κ_1	τ_1	1
		τ_2	2
		τ_3	3
		τ_4	4
		τ_5	5
		τ_6	6
		τ_7	7
		τ_8	8
		τ_9	9
		τ_{10}	10
	κ_2	\vdots	11-20
	κ_3	\vdots	21-30
	κ_4	\vdots	31-40
	κ_5	\vdots	41-50
κ_6	\vdots	51-60	
κ_7	\vdots	61-70	
J_{RHR}	\vdots	\vdots	71-140
J_{NHR}	\vdots	\vdots	141-210
J_{CHR}	\vdots	\vdots	211-280
J_{MVB}	\vdots	\vdots	281-350
J_{VVB}	\vdots	\vdots	351-420
J_{CVB}	\vdots	\vdots	421-490

Tabelle 9.2: Übersicht der vorgesehenen Standard-Parametrierungen.

Aufgrund der hier vorgesehenen 22 Varianten für die Partitionierung und der 490 Standard-Parametrierungen ergeben sich insgesamt ca. 11.000 Kombinationsmöglichkeiten. Das Fuzzy-ROSA-Verfahren ist für die systematische Exploration der Einstellmöglichkeiten softwaretechnisch so erweitert worden, dass die Einstellungen im Batch-Betrieb verändert werden und die Generierung aller zugehörigen Modelle automatisch erfolgen kann. Auch für die nachgelagerte Analyse der Modellierungsergebnisse sind entsprechende Werkzeuge, die diese Datenmenge bewältigen können, geschaffen worden. Dennoch erscheint es nicht sinnvoll, alle möglichen Kombinationen aus Partitionierungen und Parametrierungen zu untersuchen. Hinsichtlich der Standard-Parametrierungen kann nämlich vorab gesagt werden, welchen qualitativen Einfluss Veränderungen der Parametrierung auf die Regelbasis haben. Wird beispielsweise bei Wahl des Testverfahrens J_{RI} der Parameter κ von κ_1 auf

κ_2 heraufgesetzt, so erleichtert dies den statistischen Test, wodurch sich die Anzahl der generierten Regeln im Allgemeinen erhöht. Wird bei Wahl des Testverfahrens J_{RI} der Parameter κ konstantgehalten und τ von τ_1 auf τ_2 erhöht, so verringert sich im Allgemeinen die Anzahl der akzeptierten Regeln. Die Auswirkungen einer derart veränderten Regelbasis auf die Qualität der Modellierung lassen sich zwar nicht exakt, aber zumindest qualitativ aufgrund von Erfahrungswerten vorhersagen. Wenn der Test zu leicht und die Relevanzschwelle zu niedrig gewählt wird, werden viele fragwürdige Regeln, die die Modellierung eher verschlechtern, generiert und in die Regelbasis übernommen. Ist der Test zu schwer und die Relevanzschwelle zu hoch gewählt, werden eher zu wenige Regeln generiert und in die Regelbasis übernommen. Dann werden viele Datenpunkte nicht durch Regeln abgedeckt und dies beeinträchtigt ebenfalls die Qualität der Modellierung. Die Auswirkungen einer veränderten Partitionierung erweisen sich dagegen als wesentlich problemspezifischer und lassen sich daher weitaus weniger gut abschätzen. Eine direkte Nutzung der zu allen Partitionierungen erhobenen Daten für die Aufdeckung der gesuchten Zusammenhänge ist problematisch, weil eine Parametrierung in Verbindung mit einer Partitionierung gut, in Verbindung mit einer anderen Partitionierung aber schlecht geeignet sein kann. Die Gesamtheit all dieser Daten kann daher verstärkt Mehrdeutigkeiten und Widersprüche beinhalten, die für die angestrebte datenbasierte Generierung eines Moduls zur Prognose von Modellierungsfehlern nachteilig sind. Um diese Probleme zu vermeiden und den Aufwand für die Untersuchungen in einem vertretbaren Rahmen zu halten, werden folgende Varianten für ein Experiment gewählt.

Experiment A) Es werden nicht alle 490 Parametrierungen, sondern eine deutlich geringere Anzahl verwendet. Dabei wird für das jeweils gewählte Test- und Bewertungsverfahren der Parameter κ bei dem jeweiligen Default-Wert belassen⁶. Der Parameter τ wird wie oben beschrieben variiert. Es ergeben sich so zehn Variationen für jedes der sieben Test- und Bewertungsverfahren und damit insgesamt 70 Parametrierungen. Diese werden für alle 22 beschriebenen Partitionierungen angesetzt. Die mit einer automatischen und mit verschiedenen Standard-Partitionierungen erzielten Modellierungsergebnisse werden miteinander verglichen. Dieses Experiment soll klären, ob die Verwendung einer automatisch erzeugten Partitionierung *in der Regel vertretbar* ist oder typischerweise zu einer signifikant schlechteren Modellierung als die der jeweils besten Standard-Partitionierung führt.

Experiment B) Es werden alle 490 Parametrierungen jeweils nur mit der automatisch erzeugten Partitionierung verwendet. Mit diesem Experiment soll der Einfluss der Parametrierung bei gegebener Partitionierung untersucht werden. Ferner soll aufgeklärt werden, welche Test- und Bewertungsverfahren mit welchen Werten der Einstellparameter für die unterschiedlichen Problemtypen am besten geeignet sind.

Zu jedem der beschriebenen Experimente können unterschiedliche Fälle (Untervarianten)

⁶Der Default-Wert für die Test- und Bewertungsverfahren J_{RI} , J_{RHR} , J_{MVB} , J_{VVB} und J_{CVB} ist $\kappa = 0.01$, für J_{NHR} und J_{CHR} ist der Default-Wert $\kappa = 0.5$.

betrachtet werden. Sie unterscheiden sich danach, ob Mamdani-Modelle oder TSK-Modelle, ob nur positive oder auch negative Regeln verwendet werden. Um den zeitlichen Aufwand für die Untersuchungen in einem vertretbaren Rahmen zu halten, werden für das Experiment A nur Mamdani-Modelle, und zwar nur mit positiven Regeln betrachtet. Für das Experiment B werden jedoch mehrere Untervarianten B1 bis B6 untersucht. Alle erzielten Modellierungsergebnisse sind für weitergehende Analysen in Datendateien niedergelegt sowie auch in tabellarischen und grafischen Übersichten dokumentiert. Im Folgenden werden wesentliche Ergebnisse werden getrennt nach Benchmarkproblemen detaillierter vorgestellt. Eine zusammenfassende Darstellung der Ergebnisse folgt in Abschnitt 9.4.

9.3 Einsatz der neuen Strategieelemente für komplexe Anwendungen

Die im Rahmen dieser Arbeit untersuchten Anwendungen werden in Anhang B vorgestellt. Die hierfür erzielten Modellierungsergebnisse werden in Abschnitt 9.4 zusammenfassend beschrieben. In den nachfolgenden Abschnitten werden Ergebnisse zu den einzelnen Experimenten (siehe Abschnitt 9.2) detaillierter vorgestellt und diskutiert. Die Ergebnisse werden getrennt nach Anwendungen vorgestellt. Zu jeder Anwendung ist dabei die Anzahl N der Eingangsgrößen und die Anzahl D der Datenpunkte des gesamten zur Verfügung stehenden Datensatzes angegeben.

Die Betrachtung der Ergebnisse konzentriert sich im Folgenden auf das Experiment A sowie die Experimente B1 und B2. Das Experiment A untersucht den Einfluss der Suchraumstrukturierung auf die Modellgenauigkeit bei fester Modellparametrierung, wohingegen im Experiment B1 der Einfluss der Modellparametrierung auf die Modellgenauigkeit bei fester Suchraumstrukturierung untersucht wird. In beiden Experimenten werden Mamdani-Modelle mit nur positiven Regeln eingesetzt. Im Experiment B2 werden die im Experiment B1 erzeugten Mamdani-Modelle in korrespondierende TSK-Modelle transformiert und die so erzielte Verbesserung der Modellgenauigkeit untersucht. Die genannten Experimente decken das typische Anwendungsspektrum des Fuzzy-ROSA-Verfahrens ab. Die Experimente B3 bis B6 beschäftigen sich mit den eher speziellen Anwendungsfeldern der negativen und der universellen Regeln und werden daher vergleichsweise knapp behandelt.

Vergleiche der mit dem Fuzzy-ROSA-Verfahren erzielten Modellierungsergebnisse mit denen anderer Ansätze finden sich unter anderem in [Sla01] und [Kra02]. Im Rahmen dieser Arbeit steht die systematische Anwendung des Verfahrens und die Formulierung von Einstellregeln im Vordergrund und nicht das Ausschöpfen aller verfügbaren Optionen zum Erreichen eines im Sinne des Modellierungsfehlers möglichst optimalen Ergebnisses. Die neuen Strategieelemente für das Verfahren zielen auf eine bessere Handhabung von Problemfällen ab, die in industriellen Anwendungen auftreten können. Sie verändern das Verfahren aber bewusst nicht in seinem grundsätzlichen Ansatz, sodass vergleichbare Modellierungsergeb-

nisse wie in vorangegangenen Arbeiten zu erwarten sind. Mit dem Fuzzy-ROSA-Verfahren werden die besten Ergebnisse für Mamdani-Modelle in der Regel nach einer optimierenden Konfliktreduktion der Regelbasis erzielt [Sla01]. Die besten Ergebnisse für TSK-Modelle werden meist bei Verwendung einer multilinearen Funktion als Konklusion einer TSK-Regel erzielt [Kra02]. Im Rahmen dieser Arbeit kommen diese speziellen Strategieelemente nicht zum Einsatz, da die systematische Durchführung einer optimierenden Konfliktreduktion aufgrund der Anzahl der erzeugten Modelle nicht möglich ist und eine multilineare Funktion mit einer hohen Anzahl an freien Parametern einhergeht und häufig zu einem unterbestimmten linearen Gleichungssystem führt. Ein Vergleich der hier erzielten Ergebnisse mit denjenigen aus [Sla01] und [Kra02], die ohne diese speziellen Strategieelemente erzielt werden, zeigt, dass sich tatsächlich vergleichbare, teilweise aber auch bessere Resultate ergeben. Es ist daher davon auszugehen, dass sich durch den Einsatz der oben genannten speziellen Strategieelemente bei der systematischen Anwendung weitere Verbesserungen erzielen lassen dürften. Damit dürfte die Performance des Verfahrens gegenüber anderen Ansätzen verbessert werden, sie wird sich aber in vielen Fällen nicht wesentlich ändern.

Ergebnisse Experiment A

Ziel des Experiments ist es, die Modellierungsergebnisse bei einer automatischen Partitionierung mit denen bei unterschiedlichen Standard-Partitionierungen zu vergleichen. Aufbauend auf den Ergebnissen wird eine Aussage darüber getroffen, ob die Verwendung einer automatisch erzeugten Partitionierung als Standard-Vorgehensweise geeignet ist.

In den nachfolgenden Abschnitten werden Modellierungsergebnisse bei unterschiedlichen Suchraumstrukturierungen für unterschiedliche Benchmarkprobleme verglichen. Dabei werden die wesentlichen Modellierungsergebnisse für ein Problem in einer Tabelle aufgeführt, die wie folgt aufgebaut ist: In der ersten Spalte ist die jeweilige Partitionierung der Eingangsgrößen und der Ausgangsgröße angegeben. Alle Größen werden in der gleichen Weise granularisiert. Es werden hier die Ergebnisse für neun der insgesamt 21 Standard-Partitionierungen (siehe Tabelle 9.1) und die automatische Partitionierung vorgestellt. Da durch das Clusterverfahren ausschließlich trapezförmige Zugehörigkeitsfunktionen generiert werden, werden hier zur besseren Vergleichbarkeit auch bei allen anderen Ansätzen ausschließlich diejenigen Varianten betrachtet, die trapezförmige Zugehörigkeitsfunktionen erzeugen. Zu jeder der Partitionierungen werden 70 Modelle erzeugt. Es werden hierfür sieben Test- und Bewertungsverfahren eingesetzt, wobei der Parameter κ konstant auf dem jeweiligen Default-Wert belassen und der Parameter τ in der in Abschnitt 9.2 beschriebenen Weise verändert wird. Zu jeder der zehn ausgewählten Partitionierungen ist in der Tabelle das jeweils beste der 70 Modellierungsergebnisse angegeben. Ferner ist die zugehörige Parametrierung des Modells (das eingesetzten Test- und Bewertungsverfahren J sowie die Einstellung der freien Parameter κ und τ des gewählten Test- und Bewertungsverfahrens) und die Anzahl R der verwendeten Regeln angegeben. Hinzu kommen der unnormierte Modellierungsfehler ϵ (der mittlere absolute Fehler für Approximationsprobleme bzw. der

relative Klassifikationsfehler für Klassifikationsprobleme) und der entsprechende normierte Modellierungsfehler ϵ_{norm} (siehe Anhang C). Ferner wird die Korrelation r zwischen den gemessenen und den modellierten Ausgangsgrößenwerten aufgeführt. Alle Angaben bezüglich der Anzahl der Regeln und der Modellierungsfehler sind gemittelte Werte nach der Durchführung einer Kreuzvalidierung. Die Angaben bezüglich der Modellierungsfehler und der Korrelation beziehen sich stets auf die gemittelten Ergebnisse auf Validierungsdaten.

Die im Experiment A erzielten Ergebnisse bei einer automatisch erzeugten Partitionierung finden sich in den Ergebnissen des Experiments B1 wieder. Bei Verwendung der gleichen Datensätze müssen die Ergebnisse eines Test- und Bewertungsverfahrens für die Standard-Einstellung der Parameter κ und τ in den beiden Experimenten identisch sein. Für die Benchmarkprobleme Wärmetauscher, Roboterkinematik und Box-Jenkins-Zeitreihe zeigen sich jedoch Differenzen zwischen den Ergebnissen der Experimente A und B1, die teilweise stark, teilweise aber auch nur gering ausfallen. Eine Untersuchung dieser Fälle zeigt, dass für die drei genannten Benchmarkprobleme nicht die gleichen Datensätze in den Experimenten A und B1 verwendet werden. Das Experiment A dient zur Beurteilung der Qualität automatisch erzeugter Partitionierungen und zur Entscheidungsfindung, ob die Experimente B1 bis B6, die auf diesen Partitionierungen aufsetzen, in dieser Weise durchgeführt werden sollen. Die Modelle des Experiments A wurden daher als erstes erstellt und mussten zu diesem Zeitpunkt manuell erzeugt werden. Hierbei wurden in den oben genannten Fällen andere Datensätze bzw. andere Teilungen der zur Verfügung stehenden Daten verwendet als in dem nachfolgenden Experiment B1 bzw. auch in den Experimenten B2 bis B6. In den Experimenten B1 bis B6 werden aber stets die gleichen Datensätze verwendet.

Wärmetauscher ($N = 3$, $D = 1126$)

Die automatische Partitionierung führt hier zu dem besten insgesamt erzielten Ergebnis. Für die Ausgangsgröße werden durch das Clusterverfahren in diesem Fall fünf linguistische Werte erzeugt. Vergleichbar gute Ergebnisse werden auch bei fünf bzw. sieben linguistischen Werten mit dem Clusterverfahren erzielt. In allen drei Fällen liefert das Mittelwertbasierte Test- und Bewertungsverfahren J_{MVB} die besten Resultate.

Partition	J	κ	τ	R	ϵ	ϵ_{norm}	r
3-Equal	J_{NHR}	0.5000	0.8000	3	5.9955	0.8104	0.6464
3-HistFit	J_{RI}	0.0100	0.4000	0	7.4527	1.0055	0.0000
3-Cluster	J_{RI}	0.0100	0.7000	2	5.6209	0.7607	0.6068
5-Equal	J_{MVB}	0.0100	0.2000	34	3.1735	0.4281	0.8371
5-HistFit	J_{VVB}	0.0100	0.0000	31	3.9287	0.5293	0.6445
5-Cluster	J_{MVB}	0.0100	0.2000	33	2.2627	0.3053	0.9021
7-Equal	J_{MVB}	0.0100	0.5000	32	4.4988	0.6073	0.8230
7-HistFit	J_{VVB}	0.0100	0.5000	53	3.4779	0.4680	0.7476
7-Cluster	J_{MVB}	0.0100	0.2000	50	2.1896	0.2970	0.9209
Auto-Cluster	J_{MVB}	0.0100	0.2000	27	2.1806	0.2935	0.9135

Tabelle 9.3: Jeweils beste Modellierungsergebnisse mit Mamdani-Modellen bei unterschiedlichen Partitionierungen für das Benchmarkproblem Wärmetauscher.

Mackey-Glass-Zeitreihe ($N = 4$, $D = 1000$)

Das Modellierungsergebnis bei einer automatischen Partitionierung ist hier vergleichsweise schlecht. Der Grund hierfür liegt in erster Linie darin, dass für die Ausgangsgröße nur vier linguistische Werte erzeugt werden (siehe Anhang B). Das Ergebnis ist vergleichbar mit den Fällen, in denen drei linguistische Werte verwendet werden. Es ist aber deutlich schlechter als in den Fällen, in denen fünf oder sieben linguistische Werte verwendet werden. Die besten Resultate werden für dieses Problem mit dem Varianzbasierten Test- und Bewertungsverfahren J_{VVB} erzielt.

Partition	J	κ	τ	R	ϵ	ϵ_{norm}	r
3-Equal	J_{CVB}	0.0100	0.5000	24	0.0976	0.4993	0.8867
3-HistFit	J_{CVB}	0.0100	0.5000	24	0.1108	0.5671	0.8807
3-Cluster	J_{CVB}	0.0100	0.4000	46	0.0869	0.4447	0.9001
5-Equal	J_{VVB}	0.0100	0.4000	83	0.0549	0.2811	0.9614
5-HistFit	J_{VVB}	0.0100	0.3000	102	0.0496	0.2540	0.9716
5-Cluster	J_{VVB}	0.0100	0.4000	105	0.0441	0.2258	0.9716
7-Equal	J_{NHR}	0.5000	0.0000	146	0.0455	0.2329	0.9663
7-HistFit	J_{VVB}	0.0100	0.5000	168	0.0373	0.1908	0.9794
7-Cluster	J_{VVB}	0.0100	0.5000	188	0.0374	0.1916	0.9798
Auto-Cluster	J_{VVB}	0.0100	0.4000	72	0.0665	0.3403	0.9440

Tabelle 9.4: Jeweils beste Modellierungsergebnisse mit Mamdani-Modellen bei unterschiedlichen Partitionierungen für das Benchmarkproblem Mackey-Glass-Zeitreihe.

Operator ($N = 5$, $D = 70$)

Die automatische Partitionierung schneidet hier eher schlecht ab. Für die Ausgangsgröße werden in diesem Fall zwar sieben linguistische Werte erzeugt (siehe Anhang B), diese sind aber ungünstig platziert. Eine ähnliche Tendenz ist auch bei der Partitionierung einiger Eingangsgrößen zu erkennen. Die besten Resultate für dieses Problem werden mit der normierten Trefferquote J_{NHR} erzielt.

Partition	J	κ	τ	R	ϵ	ϵ_{norm}	r
3-Equal	J_{NHR}	0.5000	0.5000	28	229.6864	0.1283	0.9897
3-HistFit	J_{NHR}	0.5000	0.5000	25	969.7871	0.5297	0.9458
3-Cluster	J_{VVB}	0.0100	0.0000	19	401.8604	0.2217	0.9752
5-Equal	J_{MVB}	0.0100	0.7000	23	366.6281	0.1953	0.9798
5-HistFit	J_{MVB}	0.0100	0.7000	28	495.1663	0.2685	0.9763
5-Cluster	J_{VVB}	0.0100	0.7000	16	392.3491	0.2174	0.9621
7-Equal	J_{MVB}	0.0100	0.8000	29	386.2610	0.1950	0.9480
7-HistFit	J_{MVB}	0.0100	0.9000	27	360.1667	0.1891	0.9452
7-Cluster	J_{NHR}	0.5000	0.5000	28	339.0174	0.1813	0.9816
Auto-Cluster	J_{VVB}	0.0100	0.8000	19	508.0831	0.2734	0.9275

Tabelle 9.5: Jeweils beste Modellierungsergebnisse mit Mamdani-Modellen bei unterschiedlichen Partitionierungen für das Benchmarkproblem Operator.

Roboterkinematik ($N = 8, D = 8192$)

Die erzielten Ergebnisse sind durchweg eher mäßig, wie die angegebenen normierten Fehlerwerte von $\epsilon_{norm} > 0.6$ zeigen. Bei nur drei linguistischen Werten schlägt die Modellierung völlig fehl, bei fünf bzw. sieben linguistischen Werten gelingt die Modellierung etwas besser. Die automatische Partitionierung liefert hier ein Ergebnis, das bezüglich des Modellierungsfehlers eher etwas schlechter ausfällt, allerdings ist die Anzahl der Regeln dafür deutlich geringer als in vielen anderen Fällen. Bezüglich der Eignung der Test- und Bewertungsverfahren zeigt sich für dieses Problem kein eindeutiger Trend.

Partition	J	κ	τ	R	ϵ	ϵ_{norm}	r
3-Equal	J_{RHR}	0.0100	0.0000	108	0.0632	0.9361	0.7824
3-HistFit	J_{CHR}	0.0100	0.8000	0	0.0675	1.0000	0.0000
3-Cluster	J_{CHR}	0.0100	0.5000	0	0.0675	1.0000	0.0000
5-Equal	J_{RHR}	0.0100	0.0000	369	0.0437	0.6473	0.7845
5-HistFit	J_{CVB}	0.0100	0.4000	279	0.0483	0.7148	0.6846
5-Cluster	J_{VVB}	0.0100	0.1000	290	0.0558	0.8270	0.6897
7-Equal	J_{RHR}	0.0100	0.0000	354	0.0426	0.6312	0.7714
7-HistFit	J_{CVB}	0.0100	0.0000	797	0.0463	0.6857	0.7687
7-Cluster	J_{MVB}	0.0100	0.4000	496	0.0523	0.7746	0.7240
Auto-Cluster	J_{MVB}	0.0100	0.4000	291	0.0528	0.7818	0.7270

Tabelle 9.6: Jeweils beste Modellierungsergebnisse mit Mamdani-Modellen bei unterschiedlichen Partitionierungen für das Benchmarkproblem Roboterkinematik (KinFH).

Box-Jenkins-Zeitreihe ($N = 10, D = 290$)

Für dieses Problem liegt die Performance der automatischen Partitionierung im Mittelfeld. Die besten Ergebnisse werden hier mit fünf bzw. sieben linguistischen Werten und äquidistant verteilten Zugehörigkeitsfunktionen in Verbindung mit der normierten Trefferquote J_{NHR} oder der relevanten Trefferquote J_{RHR} erzielt.

Partition	J	κ	τ	R	ϵ	ϵ_{norm}	r
3-Equal	J_{MVB}	0.0100	0.5000	170	0.3608	0.4122	0.9155
3-HistFit	J_{MVB}	0.0100	0.5000	172	0.3431	0.3920	0.9197
3-Cluster	J_{VVB}	0.0100	0.4000	184	0.5312	0.6069	0.8446
5-Equal	J_{NHR}	0.5000	0.4000	205	0.2596	0.2965	0.9533
5-HistFit	J_{NHR}	0.5000	0.2000	298	0.3423	0.3911	0.9394
5-Cluster	J_{MVB}	0.0100	0.6000	231	0.3672	0.4196	0.9069
7-Equal	J_{RHR}	0.0100	0.5000	86	0.2488	0.2842	0.9506
7-HistFit	J_{VVB}	0.0100	0.6000	299	0.3182	0.3634	0.9289
7-Cluster	J_{VVB}	0.0100	0.5000	91	0.3590	0.4101	0.8992
Auto-Cluster	J_{VVB}	0.0100	0.5000	179	0.3304	0.3775	0.9290

Tabelle 9.7: Jeweils beste Modellierungsergebnisse mit Mamdani-Modellen bei unterschiedlichen Partitionierungen für das Benchmarkproblem Box-Jenkins-Zeitreihe.

Stock Prices ($N = 10, D = 100$)

Auch für dieses Problem liegt die Performance der automatischen Partitionierung im Mittelfeld. Das beste Ergebnis wird mit fünf linguistischen Werten und äquidistant verteilten Zugehörigkeitsfunktionen erzielt. Als Test- und Bewertungsverfahren ist hier das Medianbasierte Verfahren J_{CVB} am besten geeignet. Es liefert das Modell mit der besten Genauigkeit, das zudem mit der kleinsten Regelbasis arbeitet. Da es sich hier um ein Problem mit zehn Eingangsgrößen und nur 100 Datenpunkten handelt (dies entspricht jeweils nur 50 Lerndatenpunkten bei einer Kreuzvalidierung), ist dieses Ergebnis auch eine Bestätigung dafür, dass das Medianbasierte Verfahren in einem solche Fall eine geeignete Wahl ist.

Partition	J	κ	τ	R	ϵ	ϵ_{norm}	r
3-Equal	J_{NHR}	0.5000	0.5000	108	5.8106	0.5811	0.7629
3-HistFit	J_{NHR}	0.5000	0.6000	94	5.5829	0.5559	0.8175
3-Cluster	J_{VVB}	0.0100	0.0000	53	8.1974	0.8234	0.7925
5-Equal	J_{CVB}	0.0100	0.6000	30	4.2645	0.4299	0.8570
5-HistFit	J_{NHR}	0.5000	0.0000	357	5.5593	0.5524	0.8197
5-Cluster	J_{NHR}	0.5000	0.1000	332	4.9881	0.4972	0.8434
7-Equal	J_{NHR}	0.5000	0.4000	176	4.8077	0.4844	0.8457
7-HistFit	J_{CVB}	0.0100	0.6000	30	6.7335	0.6724	0.7009
7-Cluster	J_{VVB}	0.0100	0.0000	41	5.8133	0.5811	0.7404
Auto-Cluster	J_{VVB}	0.0100	0.5000	65	5.6994	0.5696	0.8344

Tabelle 9.8: Jeweils beste Modellierungsergebnisse mit Mamdani-Modellen bei unterschiedlichen Partitionierungen für das Benchmarkproblem Stock Prices.

Boston Housing ($N = 13, D = 506$)

Die automatische Partitionierung liefert für dieses Problem das drittbeste Ergebnis. Auffällig ist, dass für die Ausgangsgröße dabei lediglich vier linguistische Werte verwendet werden (siehe Anhang B). Die besten Ergebnisse werden mit sieben linguistischen Werten erzielt. Für dieses Problem sind die Test- und Bewertungsverfahren, die anhand von Datenverteilungen arbeiten, besonders gut geeignet.

Partition	J	κ	τ	R	ϵ	ϵ_{norm}	r
3-Equal	J_{MVB}	0.0100	0.3000	149	4.9521	0.7447	0.7288
3-HistFit	J_{MVB}	0.0100	0.3000	102	5.3241	0.8031	0.7953
3-Cluster	J_{MVB}	0.0100	0.4000	146	4.7814	0.7198	0.7883
5-Equal	J_{CVB}	0.0100	0.0000	367	3.9343	0.5926	0.8350
5-HistFit	J_{MVB}	0.0100	0.5000	170	4.1525	0.6255	0.7676
5-Cluster	J_{CVB}	0.0100	0.0000	23	4.7502	0.7132	0.6171
7-Equal	J_{MVB}	0.0100	0.5000	164	3.4168	0.5143	0.8457
7-HistFit	J_{MVB}	0.0100	0.5000	131	4.7253	0.7141	0.6924
7-Cluster	J_{VVB}	0.0100	0.4000	244	3.3406	0.5028	0.8202
Auto-Cluster	J_{MVB}	0.0100	0.5000	179	3.7512	0.5650	0.8119

Tabelle 9.9: Jeweils beste Modellierungsergebnisse mit Mamdani-Modellen bei unterschiedlichen Partitionierungen für das Benchmarkproblem Boston Housing.

Iris ($N = 4$, $D = 150$)

Die automatische Partitionierung führt hier zu dem besten Ergebnis. Bei einer anderen Partitionierung (7-HistFit) wird zwar der gleiche relative Klassifikationsfehler erreicht, es werden aber deutlich mehr Regeln verwendet. Auffällig ist auch, dass in dem genannten Fall das Mittelwertbasierte Verfahren J_{MVB} das beste Ergebnis erzielt, obwohl dieses Verfahren eher für die Behandlung von Approximationsproblemen konzipiert ist. Im Gegensatz dazu führt bei der automatischen Partitionierung die normierte Trefferquote J_{NHR} zu dem kleinsten relativen Klassifikationsfehler.

Partition	J	κ	τ	R	ϵ	ϵ_{norm}	r
3-Equal	J_{NHR}	0.5000	0.4000	29	0.0400	0.0614	0.9702
3-HistFit	J_{NHR}	0.5000	0.5000	27	0.0667	0.1028	0.9536
3-Cluster	J_{CHR}	0.0100	0.0000	7	0.6667	1.0048	0.1986
5-Equal	J_{NHR}	0.5000	0.2000	46	0.0400	0.0602	0.9708
5-HistFit	J_{VVB}	0.0100	0.5000	35	0.0467	0.0697	0.9656
5-Cluster	J_{CHR}	0.0100	0.0000	4	0.6667	1.0048	0.5113
7-Equal	J_{MVB}	0.0100	0.7000	36	0.3667	0.5502	0.7776
7-HistFit	J_{MVB}	0.0100	0.7000	36	0.0333	0.0508	0.9627
7-Cluster	J_{MVB}	0.0100	0.0000	56	0.6667	1.0048	0.0000
Auto-Cluster	J_{NHR}	0.5000	0.7000	25	0.0333	0.0508	0.9762

Tabelle 9.10: Jeweils beste Modellierungsergebnisse mit Mamdani-Modellen bei unterschiedlichen Partitionierungen für das Benchmarkproblem Iris.

Wine ($N = 13$, $D = 178$)

Hier liegt das Ergebnis der automatischen Partitionierung im unteren Mittelfeld. Die mit Abstand besten Ergebnisse werden mit fünf linguistischen Werten erzielt, insbesondere bei dem Einsatz des Clusterverfahrens. Es stellt sich daher die Frage, warum das Resultat der automatischen Partitionierung, die auch clusterbasiert arbeitet, bei Verwendung des gleichen Test- und Bewertungsverfahrens so stark von diesem Ergebnis abweicht. Betrachtet man die Granularisierung der Eingangsgrößen (siehe Anhang B), so fällt auf, dass für neun der dreizehn Eingangsgrößen jeweils nur drei bzw. vier linguistische Werte erzeugt werden. Hier ist die Granularisierung unter Umständen zu grob gewählt. Auffällig ist für dieses Problem weiterhin, dass das Varianzbasierte Verfahren J_{VVB} oftmals zu den besten Ergebnissen führt, obwohl dieses eher für die Bearbeitung von Approximationsproblemen konzipiert ist.

Partition	J	κ	τ	R	ϵ	ϵ_{norm}	r
3-Equal	J_{NHR}	0.5000	0.0000	363	0.0674	0.0923	0.9475
3-HistFit	J_{NHR}	0.5000	0.8000	126	0.0281	0.0385	0.9640
3-Cluster	J_{NHR}	0.5000	0.7000	179	0.0337	0.0462	0.9724
5-Equal	J_{VVB}	0.0100	0.5000	174	0.0169	0.0231	0.9863
5-HistFit	J_{NHR}	0.5000	0.6000	508	0.0169	0.0231	0.9860
5-Cluster	J_{VVB}	0.0100	0.0000	319	0.0112	0.0154	0.9909
7-Equal	J_{VVB}	0.0100	0.0000	258	0.0225	0.0308	0.9687
7-HistFit	J_{VVB}	0.0100	0.8000	290	0.0674	0.0923	0.9196
7-Cluster	J_{VVB}	0.0100	0.0000	340	0.0225	0.0308	0.9683
Auto-Cluster	J_{VVB}	0.0100	0.5000	179	0.0281	0.0385	0.9777

Tabelle 9.11: Jeweils beste Modellierungsergebnisse mit Mamdani-Modellen bei unterschiedlichen Partitionierungen für das Benchmarkproblem Wine.

Fazit

Die Performance der automatischen Partitionierung liegt, verglichen mit den hier betrachteten Standard-Partitionierungen, meist im oberen Mittelfeld, in einigen Fällen auch in der Spitzengruppe. In zwei Fällen ist die Performance unterdurchschnittlich. Für das Benchmarkproblem Mackey-Glass-Zeitreihe werden offenbar zu wenige linguistische Werte für die Ausgangsgröße erzeugt. Für das Benchmarkproblem Operator werden die Zugehörigkeitsfunktionen offenbar ungünstig platziert. Wie die Ergebnisse insgesamt zeigen, zeichnet sich aber auch keine der Standard-Partitionierungen als stets beste Lösung ab. Daher ist der Einsatz der automatischen Partitionierung, sofern kein besonderes Vorwissen über das zu modellierende Problem vorliegt, eine durchaus angemessene Vorgehensweise.

Ergebnisse Experiment B1

Ziel dieses Experiments ist es, die Modellierungsergebnisse für verschiedene Standard-Parametrierungen bei einer fest vorgegebenen Partitionierung miteinander zu vergleichen. Aufbauend auf den Ergebnissen werden qualitative Empfehlungen formuliert, welche Standard-Parametrierungen für welche Arten von Problemen besonders geeignet erscheinen.

In den nachfolgenden Abschnitten werden die erzielten Modellierungsergebnisse für die untersuchten Benchmarkprobleme vorgestellt. Um einen möglichst umfassenden, gleichzeitig aber auch möglichst kompakten Überblick zu geben, ist ein Großteil der Ergebnisse in Form von Grafiken dargestellt, die wie folgt aufgebaut sind: Zum einen wird der normierte Modellierungsfehler als Kennfeld über die Parameter κ und τ dargestellt. In Abbildung 9.1 (links) ist dies exemplarisch dargestellt. Für jedes der sieben Test- und Bewertungsverfahren wird das entsprechende Kennfeld als separate Grafik erstellt. Zum anderen wird die Anzahl der verwendeten Regeln über dem Parameter τ dargestellt, und zwar getrennt für die unterschiedlichen Werte für κ . In Abbildung 9.1 (rechts) ist dies exemplarisch dargestellt. Da die Abbildungen für alle Benchmarkprobleme in gleicher Weise aufgebaut sind, wird aus Gründen der besseren Lesbarkeit soweitwie möglich auf Beschriftungen der Achsen verzichtet. Lediglich die Anzahl der verwendeten Regeln ist angegeben, da diese von Problem zu Problem variiert.

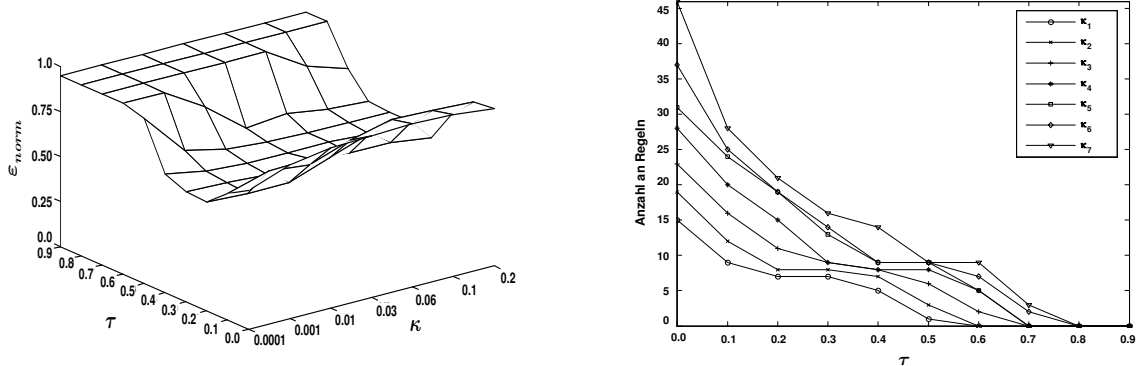


Abbildung 9.1: Exemplarische Darstellung des normierten Modellierungsfehlers ϵ_{norm} (links) und der Anzahl an Regeln (rechts) in Abhängigkeit von κ und τ .

Ausgewählte Ergebnisse werden detaillierter in einer Tabelle angegeben, die folgende Einträge enthält: In der ersten Spalte ist das jeweils eingesetzte Test- und Bewertungsverfahren angegeben, in der zweiten Spalte der Wert des Parameters κ . Für einen gewählten Wert von κ wird der Parameter τ in Schritten variiert und so die Anzahl der Regeln gemäß der Relevanz reduziert. Für die sich insgesamt ergebenden 70 Kombinationen von κ und τ sind die Anzahl der Regeln R und der jeweilige normierte gemittelte Modellierungsfehler auf Validierungsdaten ϵ_{norm} nach einer Kreuzvalidierung angegeben.

Wärmetauscher ($N = 3, D = 1126$)

Aus der Gruppe der Test- und Bewertungsverfahren, die anhand von Datenhäufigkeiten arbeiten, schneidet hier der Relevanzindex J_{RI} am besten ab. Aus der Gruppe der Test- und Bewertungsverfahren, die anhand von Datenverteilungen arbeiten, ist in diesem Fall das Mittelwertbasierte Verfahren J_{MVB} am besten geeignet, wobei die Unterschiede zu den anderen Verfahren (J_{VVB} und J_{CVB}) dieser Gruppe relativ gering ausfallen. Im Hinblick auf die erzielten Modellierungsfehler zeigt sich für das Verfahren J_{RI} , dass ein etwas leichter gewählter Test in Verbindung mit einer etwas höher gewählten Regelrelevanzschwelle günstig ist. Für das Verfahren J_{MVB} lässt sich sagen, dass auch hier ein etwas leichter gewählter Test von Vorteil ist. Das Anheben der Regelrelevanzschwelle führt in diesem Fall allerdings zu einer etwas schlechteren Modellierung. Die Ergebnisse sind für das Verfahren J_{MVB} insgesamt besser als für das Verfahren J_{RI} , allerdings werden bei diesem deutlich weniger Regeln erzeugt.

J	κ		$\tau=0.0$	$\tau=0.1$	$\tau=0.2$	$\tau=0.3$	$\tau=0.4$	$\tau=0.5$	$\tau=0.6$	$\tau=0.7$	$\tau=0.8$	$\tau=0.9$
J_{RI}	0.0001	R	15	9	7	7	5	1	0	0	0	0
		ϵ_{norm}	0.77	0.70	0.60	0.61	0.66	0.92	1.01	1.01	1.01	1.01
	0.0010	R	19	12	8	8	7	3	0	0	0	0
		ϵ_{norm}	0.85	0.77	0.59	0.60	0.61	0.68	1.01	1.01	1.01	1.01
	0.0100	R	23	16	11	9	8	6	2	0	0	0
		ϵ_{norm}	0.89	0.74	0.60	0.59	0.60	0.63	0.81	1.01	1.01	1.01
	0.0300	R	28	20	15	9	8	8	5	0	0	0
		ϵ_{norm}	0.90	0.82	0.67	0.62	0.60	0.61	0.64	1.01	1.01	1.01
	0.0600	R	31	24	19	13	9	9	5	0	0	0
		ϵ_{norm}	0.90	0.88	0.75	0.59	0.60	0.60	0.66	1.01	1.01	1.01
	0.1000	R	37	25	19	14	9	9	7	2	0	0
		ϵ_{norm}	0.90	0.87	0.76	0.59	0.59	0.60	0.62	0.81	1.01	1.01
	0.2000	R	46	28	21	16	14	9	9	3	0	0
		ϵ_{norm}	0.87	0.86	0.75	0.56	0.60	0.60	0.61	0.74	1.01	1.01

Tabelle 9.12: Ergebnisse für das Test- und Bewertungsverfahren J_{RI} mit nur positiven Regeln für das Benchmarkproblem Wärmetauscher.

J	κ		$\tau=0.0$	$\tau=0.1$	$\tau=0.2$	$\tau=0.3$	$\tau=0.4$	$\tau=0.5$	$\tau=0.6$	$\tau=0.7$	$\tau=0.8$	$\tau=0.9$
J_{MVB}	0.0001	R	23	23	21	20	17	17	11	8	6	4
		ϵ_{norm}	0.45	0.45	0.46	0.47	0.49	0.50	0.60	0.59	0.73	0.74
	0.0010	R	25	25	23	22	19	17	11	8	6	4
		ϵ_{norm}	0.45	0.45	0.45	0.47	0.49	0.50	0.60	0.59	0.73	0.74
	0.0100	R	30	30	27	26	23	22	13	8	6	4
		ϵ_{norm}	0.41	0.41	0.40	0.45	0.47	0.47	0.57	0.58	0.72	0.74
	0.0300	R	35	35	31	30	27	25	14	9	6	4
		ϵ_{norm}	0.41	0.41	0.40	0.45	0.46	0.47	0.57	0.57	0.72	0.74
	0.0600	R	39	38	33	32	29	26	15	9	6	4
		ϵ_{norm}	0.41	0.41	0.40	0.45	0.46	0.47	0.57	0.57	0.72	0.74
	0.1000	R	41	41	36	34	30	27	15	9	6	4
		ϵ_{norm}	0.41	0.41	0.39	0.45	0.46	0.47	0.57	0.57	0.72	0.74
	0.2000	R	44	43	37	35	31	29	15	9	6	4
		ϵ_{norm}	0.41	0.41	0.39	0.45	0.46	0.47	0.57	0.57	0.72	0.74

Tabelle 9.13: Ergebnisse für das Test- und Bewertungsverfahren J_{MVB} mit nur positiven Regeln für das Benchmarkproblem Wärmetauscher.

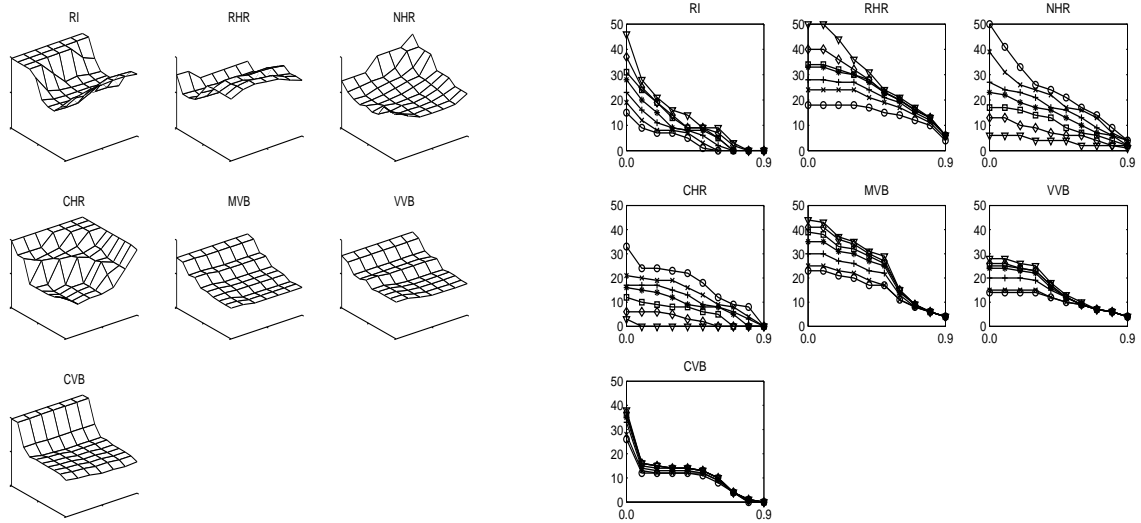


Abbildung 9.2: Normierte Modellierungsfehler (links) und Anzahl der erzeugten Regeln (rechts) für das Benchmarkproblem Wärmetauscher.

Mackey-Glass-Zeitreihe ($N = 4, D = 1000$)

Aus der Gruppe der Test- und Bewertungsverfahren, die anhand von Datenhäufigkeiten arbeiten, schneidet hier die relevante Trefferquote J_{RHR} am besten ab, aus der Gruppe der Test- und Bewertungsverfahren, die anhand von Datenverteilungen arbeiten, ist es das Varianzbasierte Verfahren J_{VVB} . Auch hier zeigen sich starke Ähnlichkeiten in den Ergebnissen der Verfahren dieser Gruppe. Im Hinblick auf die erzielten Modellierungsfehler zeigt sich für das Verfahren J_{RHR} , dass ein etwas leichter gewählter Test in Verbindung mit einer nicht zu hoch gewählten Regelrelevanzschwelle günstig ist. Für das Verfahren J_{VVB} zeigt sich dieser Trend in gleicher Form. Beide Verfahren generieren auch ähnlich viele Regeln.

J	κ		$\tau=0.0$	$\tau=0.1$	$\tau=0.2$	$\tau=0.3$	$\tau=0.4$	$\tau=0.5$	$\tau=0.6$	$\tau=0.7$	$\tau=0.8$	$\tau=0.9$
J_{RHR}	0.0001	R	35	35	35	35	35	31	23	20	13	9
		ϵ_{norm}	0.89	0.89	0.89	0.89	0.89	0.95	0.85	0.81	0.76	0.85
	0.0010	R	49	49	49	49	47	44	32	25	16	10
		ϵ_{norm}	0.60	0.60	0.60	0.60	0.70	0.74	0.85	0.81	0.79	0.85
	0.0100	R	66	66	66	66	63	58	42	30	20	11
		ϵ_{norm}	0.47	0.47	0.47	0.47	0.57	0.64	0.86	0.81	0.79	0.85
	0.0300	R	77	77	77	77	73	66	47	32	20	11
		ϵ_{norm}	0.44	0.44	0.44	0.44	0.55	0.63	0.81	0.82	0.79	0.85
	0.0600	R	84	84	84	84	79	70	48	33	20	11
		ϵ_{norm}	0.45	0.45	0.45	0.45	0.56	0.63	0.81	0.82	0.79	0.85
	0.1000	R	89	89	89	88	82	72	49	34	20	11
		ϵ_{norm}	0.43	0.43	0.43	0.44	0.56	0.63	0.81	0.81	0.79	0.85
	0.2000	R	97	97	97	95	88	74	51	36	22	12
		ϵ_{norm}	0.42	0.42	0.42	0.45	0.58	0.63	0.81	0.81	0.79	0.84

Tabelle 9.14: Ergebnisse für das Test- und Bewertungsverfahren J_{RHR} mit nur positiven Regeln für das Benchmarkproblem Mackey-Glass-Zeitreihe.

J	κ		$\tau=0.0$	$\tau=0.1$	$\tau=0.2$	$\tau=0.3$	$\tau=0.4$	$\tau=0.5$	$\tau=0.6$	$\tau=0.7$	$\tau=0.8$	$\tau=0.9$
J_{VVB}	0.0001	R	78	78	78	76	70	53	40	15	2	0
		ϵ_{norm}	0.36	0.36	0.36	0.35	0.34	0.40	0.53	0.69	0.99	1.00
	0.0010	R	82	82	82	79	71	54	41	16	3	0
		ϵ_{norm}	0.35	0.35	0.35	0.35	0.34	0.40	0.53	0.69	0.99	1.00
	0.0100	R	86	86	85	79	72	55	41	16	3	0
		ϵ_{norm}	0.35	0.35	0.35	0.35	0.34	0.40	0.53	0.69	0.99	1.00
	0.0300	R	88	88	87	80	72	55	41	16	3	0
		ϵ_{norm}	0.35	0.35	0.35	0.34	0.34	0.40	0.53	0.69	0.99	1.00
	0.0600	R	90	90	88	80	72	55	41	16	3	0
		ϵ_{norm}	0.36	0.36	0.35	0.34	0.34	0.40	0.53	0.69	0.99	1.00
	0.1000	R	92	91	89	81	73	55	42	16	3	0
		ϵ_{norm}	0.36	0.36	0.35	0.34	0.34	0.40	0.53	0.69	0.99	1.00
	0.2000	R	93	91	89	81	73	55	42	16	3	0
		ϵ_{norm}	0.36	0.36	0.35	0.34	0.34	0.40	0.53	0.69	0.99	1.00

Tabelle 9.15: Ergebnisse für das Test- und Bewertungsverfahren J_{VVB} mit nur positiven Regeln für das Benchmarkproblem Mackey-Glass-Zeitreihe.

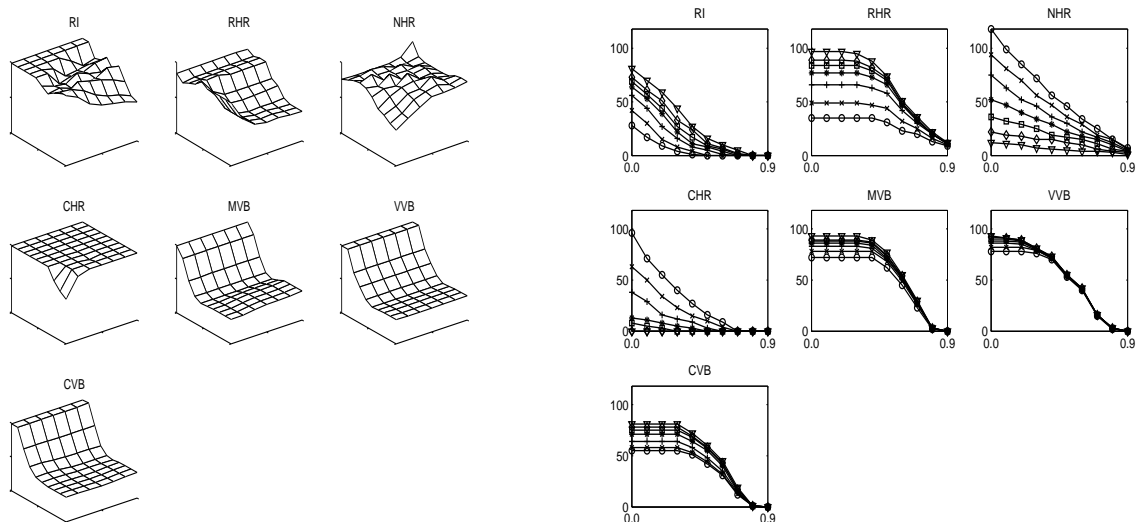


Abbildung 9.3: Normierte Modellierungsfehler (links) und Anzahl der erzeugten Regeln (rechts) für das Benchmarkproblem Mackey-Glass-Zeitreihe.

Operator ($N = 5, D = 70$)

Aus der Gruppe der Test- und Bewertungsverfahren, die anhand von Datenhäufigkeiten arbeiten, schneidet das Verfahren J_{RI} am besten ab, allerdings nur für ein sehr begrenztes Spektrum an Parametrierungen. Aus der anderen Gruppe der Test- und Bewertungsverfahren ist für dieses Problem das Medianbasierte Verfahren J_{CVB} gut geeignet. Allerdings ist auch bei diesem das Ergebnis sehr stark abhängig von der gewählten Parametrierung. Es zeigt sich für das Verfahren J_{RI} , dass ein deutlich leichter gewählter Test hier für ein akzeptables Ergebnis notwendig ist. Die Relevanzschwelle darf ebenfalls nur niedrig gewählt werden, da sonst zu viele Regeln entfernt werden und entsprechend häufig auf den

Default-Wert zurückgegriffen werden muss. Für das Verfahren J_{CVB} zeigt sich insbesondere, dass der Test hier nicht schwerer gewählt werden darf als die Standard-Einstellung $\kappa = 0.01$ vorsieht, da sonst zu wenige Regeln erzeugt werden. Ein zu leicht gewählter Test ist aber auch weniger gut geeignet, da dann offenbar widersprüchliche Regeln erzeugt werden. Die Relevanzschwelle ist für dieses Verfahren weniger sensibel, sofern die Schwelle nicht auf den Höchstwert $\tau = 0.9$ gesetzt wird. Es fällt für dieses Verfahren weiterhin auf, dass der Anteil von Regeln mit sehr hoher Relevanz, verglichen mit anderen Problemen, überdurchschnittlich hoch ist.

J	κ		$\tau=0.0$	$\tau=0.1$	$\tau=0.2$	$\tau=0.3$	$\tau=0.4$	$\tau=0.5$	$\tau=0.6$	$\tau=0.7$	$\tau=0.8$	$\tau=0.9$	
J_{RI}	0.0001	R	0	0	0	0	0	0	0	0	0	0	
		ϵ_{norm}	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02
	0.0010	R	0	0	0	0	0	0	0	0	0	0	0
		ϵ_{norm}	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02
	0.0100	R	0	0	0	0	0	0	0	0	0	0	0
		ϵ_{norm}	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02
	0.0300	R	2	1	0	0	0	0	0	0	0	0	0
		ϵ_{norm}	0.93	0.93	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02
	0.0600	R	4	3	1	0	0	0	0	0	0	0	0
		ϵ_{norm}	0.48	0.70	0.93	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02
	0.1000	R	6	4	3	0	0	0	0	0	0	0	0
		ϵ_{norm}	0.29	0.47	0.70	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02
	0.2000	R	10	7	4	3	1	0	0	0	0	0	0
		ϵ_{norm}	0.26	0.28	0.47	0.55	0.93	1.02	1.02	1.02	1.02	1.02	1.02

Tabelle 9.16: Ergebnisse für das Test- und Bewertungsverfahren J_{RI} mit nur positiven Regeln für das Benchmarkproblem Operator.

J	κ		$\tau=0.0$	$\tau=0.1$	$\tau=0.2$	$\tau=0.3$	$\tau=0.4$	$\tau=0.5$	$\tau=0.6$	$\tau=0.7$	$\tau=0.8$	$\tau=0.9$	
J_{CVB}	0.0001	R	0	0	0	0	0	0	0	0	0	0	
		ϵ_{norm}	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02
	0.0010	R	1	1	1	1	1	1	1	1	1	1	1
		ϵ_{norm}	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92
	0.0100	R	5	5	5	5	5	5	5	5	5	4	2
		ϵ_{norm}	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.35	0.75
	0.0300	R	12	12	12	11	10	10	10	10	10	9	3
		ϵ_{norm}	0.28	0.28	0.28	0.28	0.34	0.34	0.34	0.34	0.34	0.33	0.71
	0.0600	R	20	20	19	18	17	17	16	16	16	13	6
		ϵ_{norm}	0.30	0.31	0.31	0.31	0.39	0.39	0.39	0.39	0.39	0.36	0.61
	0.1000	R	25	24	23	22	21	20	19	18	18	15	7
		ϵ_{norm}	0.32	0.32	0.33	0.32	0.39	0.41	0.41	0.41	0.41	0.37	0.64
	0.2000	R	42	40	37	37	35	34	33	33	31	25	16
		ϵ_{norm}	0.45	0.45	0.45	0.45	0.46	0.46	0.47	0.47	0.44	0.38	0.51

Tabelle 9.17: Ergebnisse für das Test- und Bewertungsverfahren J_{CVB} mit nur positiven Regeln für das Benchmarkproblem Operator.

Roboterkinematik ($N = 8, D = 8192$)

Unabhängig von dem eingesetzten Test- und Bewertungsverfahren liefert die Modellierung für dieses Problem nur unbefriedigende Ergebnisse. Die Anzahl der generierten Regeln ist dabei in einigen Fällen aber durchaus hoch. Tabelle 9.18 zeigt beispielhaft die Ergebnisse

für das Mittelwertbasierte Test- und Bewertungsverfahren J_{MVB} . Das beste mit diesem Verfahren erzielte Ergebnis liefert einen normierten Fehlerwert von $\epsilon_{norm} = 0.78$ und ist damit nur unwesentlich besser als das zur Referenzbildung herangezogene Trivialmodell. Dieses Ergebnis ist durch die Verteilung der Ausgangsgrößenwerte dieses Problems (siehe Anhang B) zu erklären. Die Verteilung der Ausgangsgrößenwerte entspricht in Näherung einer Normalverteilung. Die überwiegende Mehrheit der Ausgangsgrößenwerte liegt eng um deren Mittelwert verteilt und wird daher durch das Trivialmodell auch entsprechend gut nachgebildet. Die vergleichsweise wenigen Ausgangsgrößenwerte, die deutlicher von dem Mittelwert entfernt liegen, werden durch das Fuzzy-Modell zwar besser angenähert als durch das Trivialmodell, aber in ihrer Tendenz trotzdem immer noch sehr ungenau nachgebildet. Der Großteil der Regeln des Fuzzy-Modells wird also für die Nachbildung des Mittelwertes verwendet, es fehlen aber die *richtigen* Regeln für die äußeren Randbereiche der Ausgangsgrößenwerte.

J	κ		$\tau=0.0$	$\tau=0.1$	$\tau=0.2$	$\tau=0.3$	$\tau=0.4$	$\tau=0.5$	$\tau=0.6$	$\tau=0.7$	$\tau=0.8$	$\tau=0.9$
J_{MVB}	0.0001	R	341	341	341	341	237	5	0	0	0	0
		ϵ_{norm}	0.78	0.78	0.78	0.78	0.78	0.94	1.00	1.00	1.00	1.00
	0.0010	R	377	377	377	377	260	5	0	0	0	0
		ϵ_{norm}	0.79	0.79	0.79	0.79	0.78	0.94	1.00	1.00	1.00	1.00
	0.0100	R	420	420	420	420	287	5	0	0	0	0
		ϵ_{norm}	0.79	0.79	0.79	0.79	0.79	0.94	1.00	1.00	1.00	1.00
	0.0300	R	447	447	447	447	302	5	0	0	0	0
		ϵ_{norm}	0.80	0.80	0.80	0.80	0.79	0.94	1.00	1.00	1.00	1.00
	0.0600	R	468	468	468	468	314	6	0	0	0	0
		ϵ_{norm}	0.80	0.80	0.80	0.80	0.79	0.93	1.00	1.00	1.00	1.00
	0.1000	R	491	491	491	491	325	6	0	0	0	0
		ϵ_{norm}	0.80	0.80	0.80	0.80	0.79	0.92	1.00	1.00	1.00	1.00
	0.2000	R	507	507	507	507	343	7	0	0	0	0
		ϵ_{norm}	0.81	0.81	0.81	0.81	0.80	0.93	1.00	1.00	1.00	1.00

Tabelle 9.18: Ergebnisse für das Test- und Bewertungsverfahren J_{MVB} mit nur positiven Regeln für das Benchmarkproblem Roboterkinematik (KinFH).

Box-Jenkins-Zeitreihe ($N = 10$, $D = 290$)

Anhand der grafischen Darstellung der normierten Modellierungsfehler lässt sich erkennen, dass für dieses Problem die Test- und Bewertungsverfahren, die anhand der Datenhäufigkeiten arbeiten, deutlich schlechter abschneiden als die Verfahren, die anhand der Datenverteilungen arbeiten. Tabelle 9.19 zeigt die Ergebnisse für das Varianzbasierte Verfahren J_{VVB} , das hier am besten abschneidet. Die besten Ergebnisse stellen sich ein, wenn die Relevanzschwelle etwas höher gewählt wird, beispielsweise zu $\tau = 0.4$. Das Absenken der Schwere des statistischen Tests führt hier zu keiner relevanten Verbesserung der Modellgenauigkeit, es erhöht sich nur die Anzahl der verwendeten Regeln.

J	κ		$\tau=0.0$	$\tau=0.1$	$\tau=0.2$	$\tau=0.3$	$\tau=0.4$	$\tau=0.5$	$\tau=0.6$	$\tau=0.7$	$\tau=0.8$	$\tau=0.9$
J_{VVB}	0.0001	R	42	42	42	42	33	16	4	0	0	0
		ϵ_{norm}	0.67	0.67	0.67	0.67	0.65	0.68	0.88	1.00	1.00	1.00
	0.0010	R	70	70	70	65	52	22	4	0	0	0
		ϵ_{norm}	0.53	0.53	0.53	0.50	0.46	0.58	0.88	1.00	1.00	1.00
	0.0100	R	144	144	143	130	95	29	5	0	0	0
		ϵ_{norm}	0.51	0.51	0.51	0.47	0.44	0.55	0.85	1.00	1.00	1.00
	0.0300	R	184	184	181	163	106	31	5	0	0	0
		ϵ_{norm}	0.53	0.53	0.52	0.49	0.45	0.55	0.85	1.00	1.00	1.00
	0.0600	R	205	205	201	177	110	32	5	0	0	0
		ϵ_{norm}	0.54	0.54	0.53	0.49	0.45	0.54	0.85	1.00	1.00	1.00
	0.1000	R	222	222	215	185	111	33	5	0	0	0
		ϵ_{norm}	0.55	0.55	0.54	0.50	0.45	0.54	0.85	1.00	1.00	1.00
	0.2000	R	241	240	230	191	113	34	6	0	0	0
		ϵ_{norm}	0.56	0.56	0.55	0.50	0.45	0.54	0.85	1.00	1.00	1.00

Tabelle 9.19: Ergebnisse für das Test- und Bewertungsverfahren J_{VVB} mit nur positiven Regeln für das Benchmarkproblem Box-Jenkins-Zeitreihe.

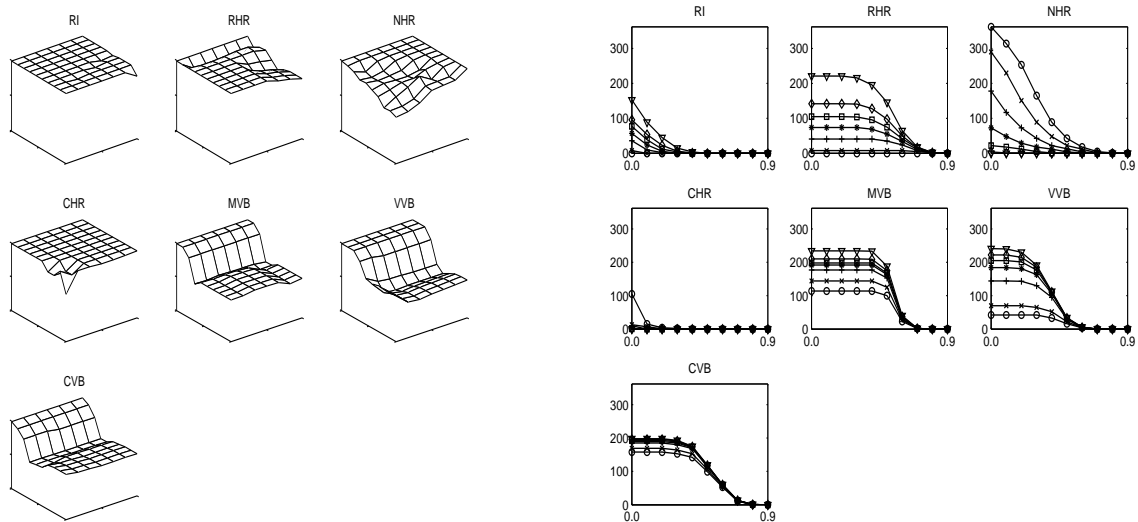


Abbildung 9.4: Normierte Modellierungsfehler (links) und Anzahl der erzeugten Regeln (rechts) für das Benchmarkproblem Box-Jenkins-Zeitreihe.

Stock Prices ($N = 10, D = 100$)

Die Test- und Bewertungsverfahren, die anhand der Datenhäufigkeiten arbeiten, sind in diesem Fall ungeeignet. Das Mittelwertbasierte Verfahren J_{MVB} und das Medianbasierte Verfahren J_{CVB} liefern vergleichbare Ergebnisse, auffällig ist das Ergebnis des Varianzbasierten Verfahrens J_{VVB} . Hier zeigt sich für einen bestimmten Wert von κ ein deutlich schlechteres Ergebnis als in anderen Fällen. Dieser besondere Trend zeigt sich in den übrigen Benchmarkproblemen nicht. Das Mittelwertbasierte Verfahren J_{MVB} liefert hier die besten Resultate. Der Test sollte nicht schwerer als üblicherweise vorgeschlagen gewählt

werden, die Relevanzschwelle kann durchaus auf einen etwas erhöhten Wert von zum Beispiel $\tau = 0.4$ gesetzt werden, der sich auch für andere Probleme als günstig erweist.

J	κ	$\tau=0.0$	$\tau=0.1$	$\tau=0.2$	$\tau=0.3$	$\tau=0.4$	$\tau=0.5$	$\tau=0.6$	$\tau=0.7$	$\tau=0.8$	$\tau=0.9$	
J_{MVB}	0.0001	R	15	15	15	15	15	15	14	13	4	0
		ϵ_{norm}	0.85	0.85	0.85	0.85	0.85	0.85	0.66	0.66	0.95	1.01
	0.0010	R	36	36	36	36	36	35	33	30	11	0
		ϵ_{norm}	0.62	0.62	0.62	0.62	0.62	0.71	0.66	0.63	0.69	1.01
	0.0100	R	75	75	75	75	74	69	60	53	26	2
		ϵ_{norm}	0.60	0.60	0.60	0.59	0.59	0.65	0.58	0.63	0.67	0.94
	0.0300	R	104	104	104	104	100	91	78	68	34	4
		ϵ_{norm}	0.57	0.57	0.57	0.57	0.58	0.63	0.57	0.61	0.67	0.92
	0.0600	R	134	134	134	134	125	114	97	81	41	6
		ϵ_{norm}	0.55	0.55	0.55	0.54	0.55	0.58	0.56	0.62	0.70	0.93
	0.1000	R	162	162	162	160	149	131	110	90	44	6
		ϵ_{norm}	0.55	0.55	0.55	0.56	0.57	0.60	0.58	0.64	0.70	0.93
	0.2000	R	233	230	226	219	204	179	149	117	51	9
		ϵ_{norm}	0.58	0.58	0.58	0.57	0.57	0.55	0.54	0.61	0.69	0.95

Tabelle 9.20: Ergebnisse für das Test- und Bewertungsverfahren J_{MVB} mit nur positiven Regeln für das Benchmarkproblem Stock Prices.

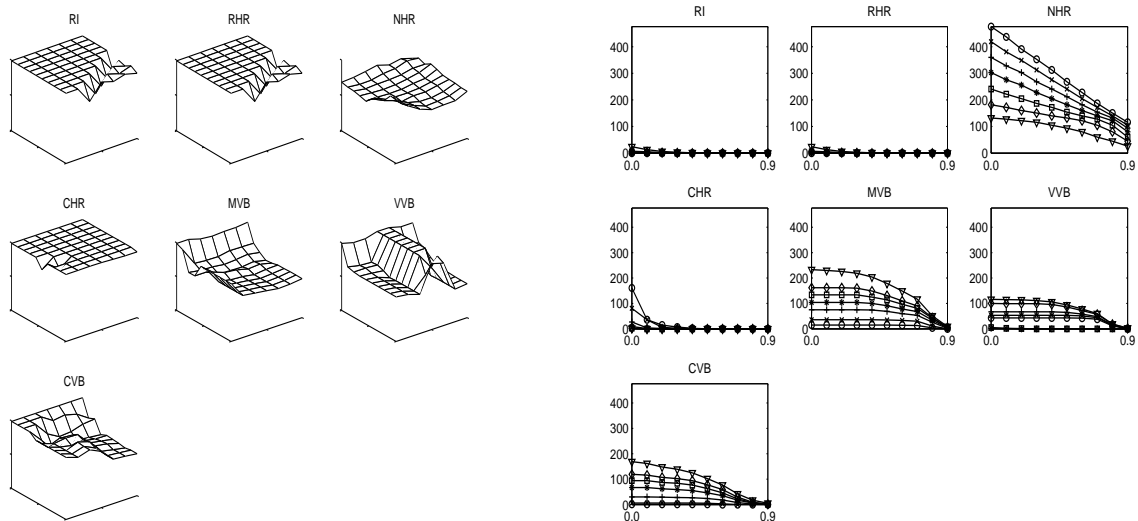


Abbildung 9.5: Normierte Modellierungsfehler (links) und Anzahl der erzeugten Regeln (rechts) für das Benchmarkproblem Stock Prices.

Boston Housing ($N = 13, D = 506$)

Akzeptable Ergebnisse erzielen hier nur die Test- und Bewertungsverfahren, die anhand der Datenverteilungen arbeiten. Das Mittelwertbasierte Verfahren J_{MVB} und das Varianzbasierte Verfahren J_{VVB} liefern vergleichbare Ergebnisse, wobei das erstgenannte etwas besser ist. Es fällt auf, dass mit abnehmender Schwere des Tests die Anzahl der generierten Regeln sehr stark ansteigt. Das Modellierungsergebnis wird dabei aber nicht besser. In

diesem Fall ist es vielmehr günstig, den Test eher etwas schwerer als üblich zu gestalten und die Relevanzschwelle auf einen erhöhten Wert zu setzen, beispielsweise auf $\tau = 0.5$.

J	κ		$\tau=0.0$	$\tau=0.1$	$\tau=0.2$	$\tau=0.3$	$\tau=0.4$	$\tau=0.5$	$\tau=0.6$	$\tau=0.7$	$\tau=0.8$	$\tau=0.9$
J_{MVB}	0.0001	R	112	112	112	112	99	76	56	14	2	1
		ϵ_{norm}	0.72	0.72	0.72	0.72	0.66	0.57	0.70	0.84	0.97	1.00
	0.0010	R	183	183	183	183	160	119	86	25	5	2
		ϵ_{norm}	0.73	0.73	0.73	0.73	0.69	0.57	0.66	0.87	0.97	1.00
	0.0100	R	288	288	288	287	248	179	127	42	9	2
		ϵ_{norm}	0.75	0.75	0.75	0.75	0.70	0.56	0.63	0.85	0.96	1.00
	0.0300	R	375	375	375	373	325	236	164	56	12	3
		ϵ_{norm}	0.75	0.75	0.75	0.75	0.70	0.58	0.61	0.84	0.95	1.00
	0.0600	R	449	449	449	446	389	287	199	73	20	4
		ϵ_{norm}	0.76	0.76	0.76	0.75	0.71	0.59	0.61	0.81	0.93	0.99
	0.1000	R	523	523	523	517	453	337	232	89	26	5
		ϵ_{norm}	0.76	0.76	0.76	0.76	0.71	0.60	0.61	0.79	0.93	0.99
0.2000	R	652	651	648	636	562	421	280	113	37	7	
	ϵ_{norm}	0.76	0.76	0.76	0.75	0.71	0.60	0.62	0.79	0.93	1.00	

Tabelle 9.21: Ergebnisse für das Test- und Bewertungsverfahren J_{MVB} mit nur positiven Regeln für das Benchmarkproblem Boston Housing.

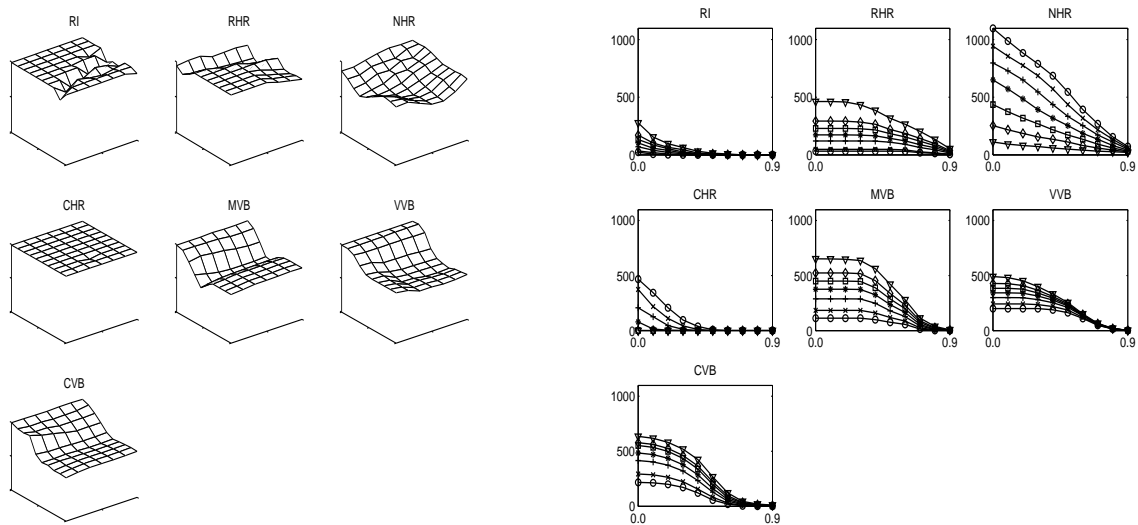


Abbildung 9.6: Normierte Modellierungsfehler (links) und Anzahl der erzeugten Regeln (rechts) für das Benchmarkproblem Boston Housing.

Iris ($N = 4, D = 150$)

Mit den Test- und Bewertungsverfahren, die anhand der Datenhäufigkeiten arbeiten, lassen sich – mit Ausnahme des Verfahrens J_{CHR} – für dieses Problem sehr genaue Modelle erzeugen, die sich jedoch stark in der Anzahl der verwendeten Regeln unterscheiden. Dabei zeichnet sich das Verfahren J_{RI} dahingehend aus, dass es hier vergleichsweise wenige

Regeln erzeugt. Auffällig ist hier insbesondere die steile Flanke, die sich in dem zugehörigen Kennfeld des Verfahrens abzeichnet. Ein akzeptables Ergebnis wird für dieses Problem bereits mit Modellen erzielt, die nur etwas mehr als fünf Regeln verwenden. Bei weniger Regeln werden offenbar relativ viele Datenpunkte nicht durch Regeln abgedeckt, was entsprechend hohe Fehlerwerte verursacht. Die insgesamt besten Modelle verwenden nicht wesentlich mehr als zehn Regeln.

J	κ		$\tau=0.0$	$\tau=0.1$	$\tau=0.2$	$\tau=0.3$	$\tau=0.4$	$\tau=0.5$	$\tau=0.6$	$\tau=0.7$	$\tau=0.8$	$\tau=0.9$
J_{RI}	0.0001	R	2	0	0	0	0	0	0	0	0	0
		ϵ_{norm}	1.01	1.51	1.51	1.51	1.51	1.51	1.51	1.51	1.51	1.51
	0.0010	R	5	4	2	0	0	0	0	0	0	0
		ϵ_{norm}	0.06	0.28	1.01	1.51	1.51	1.51	1.51	1.51	1.51	1.51
	0.0100	R	9	9	5	4	1	0	0	0	0	0
		ϵ_{norm}	0.05	0.05	0.06	0.28	1.01	1.51	1.51	1.51	1.51	1.51
	0.0300	R	11	10	10	6	5	2	0	0	0	0
		ϵ_{norm}	0.05	0.05	0.05	0.06	0.28	1.01	1.51	1.51	1.51	1.51
	0.0600	R	12	11	10	7	5	3	0	0	0	0
		ϵ_{norm}	0.05	0.04	0.04	0.05	0.06	0.71	1.51	1.51	1.51	1.51
	0.1000	R	13	12	11	10	7	5	2	0	0	0
		ϵ_{norm}	0.05	0.05	0.04	0.04	0.05	0.28	1.01	1.51	1.51	1.51
	0.2000	R	14	14	13	11	10	8	5	2	0	0
		ϵ_{norm}	0.05	0.05	0.05	0.04	0.04	0.05	0.28	1.01	1.51	1.51

Tabelle 9.22: Ergebnisse für das Test- und Bewertungsverfahren J_{RI} mit nur positiven Regeln für das Benchmarkproblem Iris.

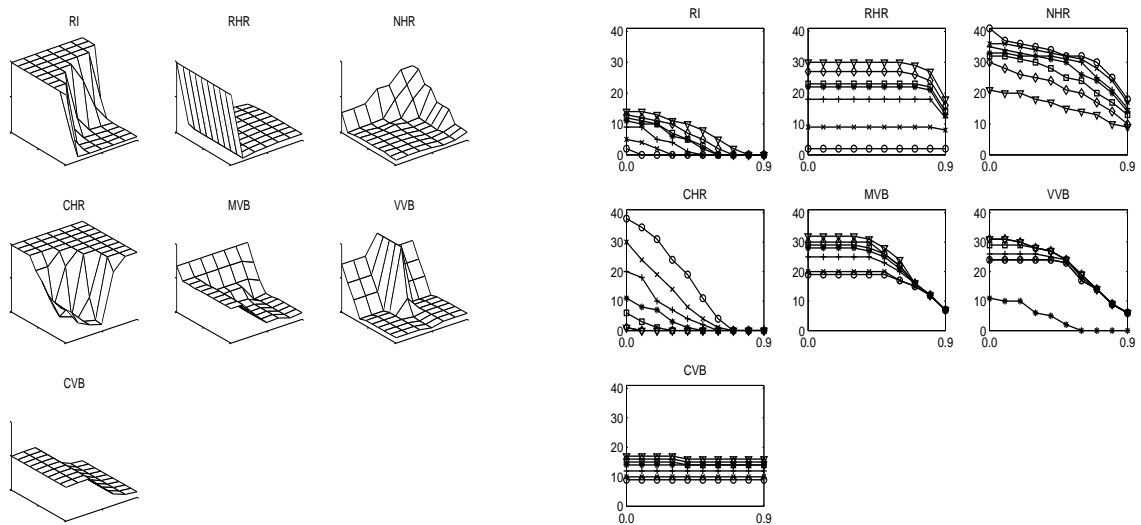


Abbildung 9.7: Normierte Modellierungsfehler (links) und Anzahl der erzeugten Regeln (rechts) für das Benchmarkproblem Iris.

Wine ($N = 13$, $D = 178$)

Die Kennfelder der Test- und Bewertungsverfahren, die anhand der Datenhäufigkeiten arbeiten, weisen starke Ähnlichkeiten mit denen auf, die sich für das Benchmarkproblem

Iris abzeichnen. Auch bezüglich der Anzahl erzeugter Regeln zeigen sich hier die zuvor beschriebenen Tendenzen. Die Ergebnisse des Verfahrens J_{RI} zeigen, dass es in diesem Fall günstig ist, die Schwere des Tests etwas leichter zu wählen und die Relevanzschwelle nicht zu hoch zu setzen.

J	κ		$\tau=0.0$	$\tau=0.1$	$\tau=0.2$	$\tau=0.3$	$\tau=0.4$	$\tau=0.5$	$\tau=0.6$	$\tau=0.7$	$\tau=0.8$	$\tau=0.9$
J_{RI}	0.0001	R	3	0	0	0	0	0	0	0	0	0
		ϵ_{norm}	1.01	1.37	1.37	1.37	1.37	1.37	1.37	1.37	1.37	1.37
	0.0010	R	10	5	1	0	0	0	0	0	0	0
		ϵ_{norm}	0.48	1.00	1.01	1.37	1.37	1.37	1.37	1.37	1.37	1.37
	0.0100	R	43	21	11	5	1	0	0	0	0	0
		ϵ_{norm}	0.15	0.15	0.38	1.00	1.01	1.37	1.37	1.37	1.37	1.37
	0.0300	R	82	57	27	12	5	1	0	0	0	0
		ϵ_{norm}	0.12	0.12	0.15	0.38	0.81	1.01	1.37	1.37	1.37	1.37
	0.0600	R	121	87	56	28	11	3	0	0	0	0
		ϵ_{norm}	0.12	0.10	0.09	0.15	0.39	1.00	1.37	1.37	1.37	1.37
	0.1000	R	152	118	86	50	22	8	2	0	0	0
		ϵ_{norm}	0.11	0.11	0.09	0.12	0.16	0.60	1.01	1.37	1.37	1.37
	0.2000	R	212	178	137	101	56	23	7	0	0	0
		ϵ_{norm}	0.09	0.09	0.10	0.09	0.12	0.11	0.60	1.37	1.37	1.37

Tabelle 9.23: Ergebnisse für das Test- und Bewertungsverfahren J_{RI} mit nur positiven Regeln für das Benchmarkproblem Wine.

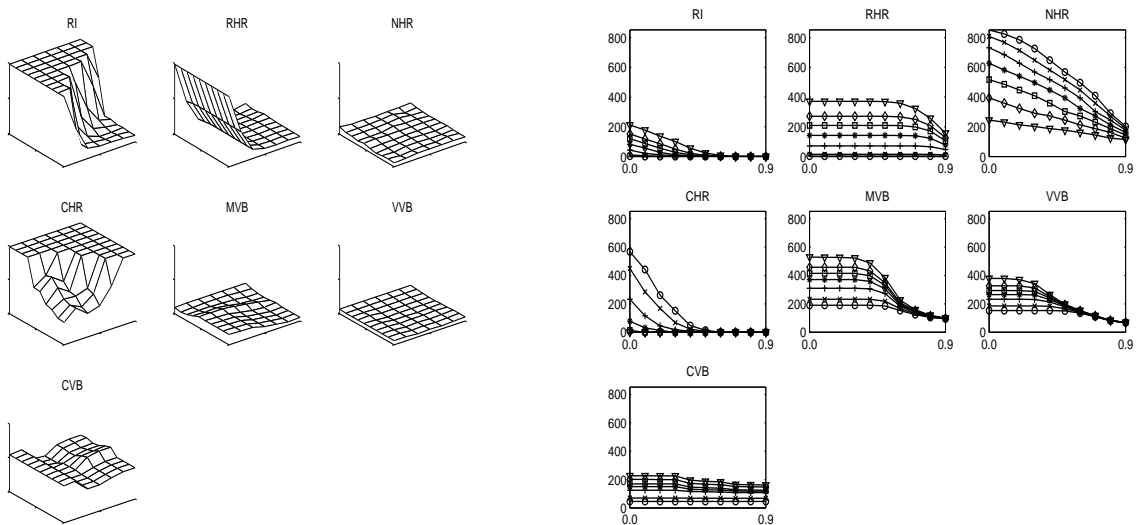


Abbildung 9.8: Normierte Modellierungsfehler (links) und Anzahl der erzeugten Regeln (rechts) für das Benchmarkproblem Wine.

Fazit

Für die meisten der hier untersuchten Benchmarkprobleme sind die Test- und Bewertungsverfahren, die anhand der Datenverteilungen arbeiten, im Hinblick auf die erzielte Modellgenauigkeit besser geeignet als die Test- und Bewertungsverfahren, die anhand der

Datenhäufigkeiten arbeiten. In Abschnitt 4.3 werden die unterschiedlichen Zielsetzungen (Relevanz, Treffsicherheit und Approximationsgüte) erläutert, die mit den einzelnen Test- und Bewertungsverfahren verfolgt werden. Mit den oben genannten Verfahren J_{MVB} , J_{VVB} und J_{CVB} wird gerade eine hohe Approximationsgüte angestrebt. Daher ist es auch nachvollziehbar, dass diese Verfahren für die Bearbeitung von Approximationsproblemen gut geeignet sind. Bei den hier untersuchten Klassifikationsproblemen ist hingegen der Relevanzindex J_{RI} , der anhand der Datenhäufigkeiten arbeitet, sowohl im Hinblick auf den Klassifikationsfehler als auch auf die Anzahl der verwendeten Regeln besser geeignet als die anderen Verfahren.

Der Schwerpunkt der Untersuchungen liegt auf Approximationsproblemen. Wie zuvor ausgeführt, sind für die Bearbeitung solcher Probleme die Test- und Bewertungsverfahren vorteilhaft, die anhand der Datenverteilungen arbeiten. Für die hier untersuchten Approximationsprobleme liefern die drei Verfahren J_{MVB} , J_{VVB} und J_{CVB} oftmals sehr ähnliche Ergebnisse im Hinblick auf die erzielte Modellgenauigkeit. Im Hinblick auf die Anzahl der generierten Regeln zeigt sich, dass das Medianbasierte Verfahren oftmals weniger Regeln erzeugt als die beiden anderen Verfahren. Dies zeigt sich insbesondere auch für die Benchmarkprobleme Operator und Stock Prices, bei denen nur vergleichsweise kleine Datensätze zur Verfügung stehen und die daher am ehesten den Szenarien entsprechen, für die dieses Verfahren konzipiert ist.

Von besonderem Interesse bei den durchgeführten Untersuchungen ist, wie sich die unterschiedlichen Parametrierungen der Test- und Bewertungsverfahren auf die erzielte Modellgenauigkeit auswirken. Die prinzipiellen Wirkungsweisen der Parameter κ und τ sind bekannt. Wie die Werte dieser Parameter aber im konkreten Anwendungsfall gewählt werden sollten, wurde in Vorarbeiten bisher nur ansatzweise untersucht. Die visuelle Inspektion der Ergebnisse und die Analyse der tabellarischen Aufstellungen zeigt, dass sich für die Parametrierung der Verfahren in vielen der untersuchten Probleme ein gemeinsamer Trend abzeichnet: Es ist oftmals günstig, ausgehend von den bisher vorgeschlagenen Standard-Einstellungen, die Schwere des statistischen Tests abzusenken und die Relevanzschwelle für die Regelübernahme anzuheben, um ein möglichst gutes Modellierungsergebnis zu erzielen.

Ergebnisse Experiment B2

Ziel dieses Experiments ist es, das Potenzial der Transformation von Mamdani-Modellen in TSK-Modelle im Hinblick auf die Verbesserung der Modellgenauigkeit einschätzen zu können. Es werden ausschließlich Approximationsprobleme untersucht, da TSK-Modelle für Klassifikationsaufgaben keine erkennbaren Vorteile gegenüber Mamdani-Modellen aufweisen. In Vorarbeiten [Kra02] wurde das Verbesserungspotenzial bereits untersucht. In dieser Arbeit wird aber eine deutlich größere Anzahl von Benchmarkproblemen verwendet, um die Ergebnisse besser abzusichern.

Bei TSK-Modellen sind im Fuzzy-ROSA-Verfahren vier Entwurfsvarianten zu unterscheiden, die in der nachfolgenden Tabelle 9.24 aufgeführt sind.

Bestimmung der Parameter	Berechnung des Ausgabewertes	Bezeichnung
Ordinary Least-Squares	Gewichteter Mittelwert	TSK-LS-MV
Ordinary Least-Squares	Gewichteter Median	TSK-LS-CV
Iteratively Reweighted Least-Squares	Gewichteter Mittelwert	TSK-RR-MV
Iteratively Reweighted Least-Squares	Gewichteter Median	TSK-RR-CV

Tabelle 9.24: Entwurfsvarianten für TSK-Modelle.

Die Abkürzungen erklären sich wie folgt: LS für Least-Squares, RR für Iteratively Reweighted Least-Squares (auch als *robust regression* bezeichnet), MV für Mittelwert (englisch: mean value) und CV für Median (englisch: central value). Die unterschiedlichen Ansätze für die Parameterbestimmung bzw. für die Berechnung des Ausgabewertes werden in den Abschnitten 7.2 und 7.4 bzw. 7.1 und 7.5 beschrieben.

Nachfolgend werden die Modellierungsergebnisse für die untersuchten Benchmarkprobleme vorgestellt. Es werden dabei fünf Modellvarianten verglichen: Mamdani-Modelle (Ergebnisse des Experiments B1) und die genannten Varianten für TSK-Modelle. Für jedes der Test- und Bewertungsverfahren ist das jeweils beste Modell, bezogen auf den mittleren betraglichen Modellierungsfehler ϵ , zu jeder Modellvariante angegeben. In der ersten Spalte einer Tabelle ist das eingesetzte Verfahren angegeben, es folgen die Werte der Parameter κ und τ und die Anzahl der Regeln R , dann folgen der unnormierte bzw. normierte gemittelte absolute Modellierungsfehler ϵ bzw. ϵ_{norm} . Weiterhin wird ein entsprechendes Maß η_{norm} auch für den quadratischen Fehler angegeben, da die Parameterbestimmung nach dem Least-Squares Ansatz auf die Minimierung dieses Fehlers abzielt. Da das in diesem Sinne optimale Ergebnis nicht zwangsläufig mit den gleichen Einstellungen für κ und τ erzielt wird wie das für den betraglichen Fehler ϵ_{norm} , ist dieser Fehlerwert η_{norm} hier in Klammern dargestellt. Die Korrelation r zwischen gemessenen und modellierten Ausgangswerten und die Bezeichnung der Modellvariante finden sich ebenfalls in einer Tabelle wieder: MAM für Mamdani-Modell, die übrigen Abkürzungen können nach Tabelle 9.24 zugeordnet werden.

Wärmetauscher ($N = 3, D = 1126$)

Das beste Ergebnis eines Mamdani-Modells wird für den Mittelwertbasierten Test J_{MVB} erzielt. Alle anderen Testverfahren, ausgenommen J_{RI} , erzielen vergleichbare Modellierungsfehler ϵ innerhalb [3.2, 3.7] bzw. Modellierungsfehler ϵ_{norm} innerhalb [0.39, 0.49]. Bei den Entwurfsvarianten für die TSK-Modelle zeigen sich durchaus Unterschiede. Als besonders gut geeignet – in Verbindung mit fast allen Test- und Bewertungsverfahren – stellt sich hier eine Kombination aus Parameterbestimmung nach dem IRLS-Ansatz und Berechnung des Modellausgabewertes über den gewichteten Median heraus (TSK-RR-CV). Der Zugewinn an Modellgenauigkeit beträgt im Mittel ca. 50%.

J	κ	τ	R	ϵ	ϵ_{norm}	η_{norm}	r	Modelltyp
J_{RI}	0.2000	0.3000	16	4.1376	0.5590	(0.5136)	0.7789	MAM
	0.2000	0.1000	28	1.7773	0.2401	(0.1022)	0.9514	TSK-LS-MV
	0.2000	0.0000	46	1.7785	0.2406	(0.1608)	0.9198	TSK-RR-MV
	0.2000	0.0000	46	1.6324	0.2204	(0.0752)	0.9624	TSK-LS-CV
	0.2000	0.0000	46	1.4615	0.1978	(0.0994)	0.9509	TSK-RR-CV
J_{RHR}	0.2000	0.7000	17	3.6139	0.4870	(0.3860)	0.8320	MAM
	0.2000	0.3000	36	1.7389	0.2346	(0.0902)	0.9577	TSK-LS-MV
	0.2000	0.2000	44	1.6222	0.2192	(0.1271)	0.9375	TSK-RR-MV
	0.2000	0.2000	44	1.5781	0.2128	(0.0729)	0.9637	TSK-LS-CV
	0.2000	0.1000	50	1.4529	0.1962	(0.1012)	0.9509	TSK-RR-CV
J_{NHR}	0.3000	0.4000	24	3.5951	0.4868	(0.3866)	0.8220	MAM
	0.3000	0.1000	41	1.6367	0.2211	(0.0840)	0.9604	TSK-LS-MV
	0.3000	0.1000	41	1.5482	0.2095	(0.1176)	0.9417	TSK-RR-MV
	0.3000	0.1000	41	1.5303	0.2066	(0.0701)	0.9649	TSK-LS-CV
	0.3000	0.1000	41	1.3606	0.1843	(0.0908)	0.9553	TSK-RR-CV
J_{CHR}	0.4000	0.4000	16	3.2792	0.4433	(0.5073)	0.7240	MAM
	0.4000	0.5000	13	2.0171	0.2711	(0.1409)	0.8994	TSK-LS-MV
	0.3000	0.0000	33	1.9476	0.2629	(0.1897)	0.9039	TSK-RR-MV
	0.3000	0.4000	22	1.9338	0.2607	(0.1209)	0.9424	TSK-LS-CV
	0.3000	0.4000	22	1.7921	0.2417	(0.1638)	0.9172	TSK-RR-CV
J_{MVB}	0.2000	0.2000	37	2.9198	0.3940	(0.2252)	0.9166	MAM
	0.2000	0.2000	37	1.6691	0.2253	(0.0874)	0.9596	TSK-LS-MV
	0.2000	0.2000	37	1.5311	0.2070	(0.1145)	0.9433	TSK-RR-MV
	0.2000	0.2000	37	1.4793	0.1996	(0.0675)	0.9665	TSK-LS-CV
	0.1000	0.2000	36	1.3551	0.1834	(0.0909)	0.9554	TSK-RR-CV
J_{VVB}	0.1000	0.0000	26	3.4332	0.4614	(0.4040)	0.8056	MAM
	0.2000	0.0000	28	1.9497	0.2627	(0.2271)	0.8824	TSK-LS-MV
	0.2000	0.0000	28	1.8397	0.2483	(0.2080)	0.8664	TSK-RR-MV
	0.2000	0.3000	25	1.9454	0.2619	(0.2230)	0.8852	TSK-LS-CV
	0.0300	0.0000	28	1.6629	0.2252	(0.1442)	0.9270	TSK-RR-CV
J_{CVB}	0.1000	0.0000	37	3.3299	0.4504	(0.3387)	0.8512	MAM
	0.2000	0.0000	38	1.7149	0.2314	(0.1013)	0.9504	TSK-LS-MV
	0.0001	0.0000	26	1.5984	0.2171	(0.1245)	0.9314	TSK-RR-MV
	0.0300	0.0000	35	1.8418	0.2481	(0.1455)	0.9257	TSK-LS-CV
	0.0300	0.0000	35	1.7139	0.2315	(0.1638)	0.9158	TSK-RR-CV

Tabelle 9.25: Jeweils beste Modellierungsergebnisse mit Mamdani- und TSK-Modellen mit nur positiven Regeln für das Benchmarkproblem Wärmetauscher.

Mackey-Glass-Zeitreihe ($N = 4, D = 1000$)

Die besten Ergebnisse mit Mamdani-Modellen (J_{VVB} , J_{MVB} und J_{CVB}) liefern Modellierungsfehler $\epsilon \approx 0.07$ bzw. Modellierungsfehler $\epsilon_{norm} \approx 0.34$. Bei den Entwurfsvarianten für die TSK-Modelle zeigen sich deutliche Unterschiede. Als besonders gut geeignet – unabhängig von dem gewählten Test- und Bewertungsverfahren – stellt sich hier ebenfalls die Entwurfsvariante TSK-RR-CV heraus. Der Zugewinn an Modellgenauigkeit beträgt im Mittel ca. 75%.

J	κ	τ	R	ϵ	ϵ_{norm}	η_{norm}	r	Modelltyp
J_{RI}	0.2000	0.0000	81	0.1075	0.5501	(0.2862)	0.9495	MAM
	0.2000	0.1000	70	0.0267	0.1366	(0.0203)	0.9907	TSK-LS-MV
	0.2000	0.1000	70	0.0227	0.1163	(0.0156)	0.9925	TSK-RR-MV
	0.2000	0.0000	81	0.0228	0.1166	(0.0169)	0.9926	TSK-LS-CV
	0.2000	0.1000	70	0.0191	0.0975	(0.0125)	0.9944	TSK-RR-CV
J_{RHR}	0.2000	0.0000	97	0.0814	0.4167	(0.1937)	0.9233	MAM
	0.0600	0.4000	79	0.0250	0.1281	(0.0175)	0.9922	TSK-LS-MV
	0.0600	0.4000	79	0.0207	0.1057	(0.0126)	0.9940	TSK-RR-MV
	0.0600	0.4000	79	0.0212	0.1085	(0.0146)	0.9940	TSK-LS-CV
	0.0600	0.4000	79	0.0174	0.0891	(0.0105)	0.9954	TSK-RR-CV
J_{NHR}	0.3000	0.0000	118	0.0984	0.5038	(0.2568)	0.9585	MAM
	0.3000	0.3000	72	0.0250	0.1279	(0.0178)	0.9919	TSK-LS-MV
	0.3000	0.2000	85	0.0208	0.1065	(0.0128)	0.9939	TSK-RR-MV
	0.4000	0.1000	81	0.0212	0.1087	(0.0150)	0.9935	TSK-LS-CV
	0.3000	0.2000	85	0.0175	0.0895	(0.0107)	0.9953	TSK-RR-CV
J_{CHR}	0.3000	0.0000	96	0.1288	0.6593	(0.4793)	0.9135	MAM
	0.4000	0.1000	50	0.0291	0.1488	(0.0229)	0.9896	TSK-LS-MV
	0.4000	0.1000	50	0.0229	0.1173	(0.0152)	0.9927	TSK-RR-MV
	0.3000	0.0000	96	0.0244	0.1251	(0.0189)	0.9922	TSK-LS-CV
	0.3000	0.1000	71	0.0172	0.0878	(0.0100)	0.9956	TSK-RR-CV
J_{MVB}	0.0010	0.5000	68	0.0672	0.3439	(0.1324)	0.9437	MAM
	0.2000	0.6000	55	0.0226	0.1158	(0.0159)	0.9921	TSK-LS-MV
	0.2000	0.5000	77	0.0192	0.0981	(0.0108)	0.9949	TSK-RR-MV
	0.2000	0.5000	77	0.0205	0.1048	(0.0141)	0.9944	TSK-LS-CV
	0.2000	0.5000	77	0.0156	0.0800	(0.0086)	0.9963	TSK-RR-CV
J_{VVB}	0.1000	0.4000	73	0.0660	0.3378	(0.1334)	0.9440	MAM
	0.1000	0.4000	73	0.0208	0.1062	(0.0126)	0.9946	TSK-LS-MV
	0.0001	0.4000	73	0.0164	0.0840	(0.0080)	0.9961	TSK-RR-MV
	0.1000	0.4000	73	0.0191	0.0979	(0.0129)	0.9949	TSK-LS-CV
	0.0001	0.4000	73	0.0145	0.0744	(0.0078)	0.9966	TSK-RR-CV
J_{CVB}	0.0300	0.0000	71	0.0694	0.3554	(0.1363)	0.9410	MAM
	0.0600	0.3000	75	0.0276	0.1410	(0.0215)	0.9901	TSK-LS-MV
	0.2000	0.4000	72	0.0234	0.1196	(0.0163)	0.9921	TSK-RR-MV
	0.0600	0.0000	75	0.0227	0.1160	(0.0163)	0.9930	TSK-LS-CV
	0.0600	0.3000	75	0.0185	0.0946	(0.0115)	0.9947	TSK-RR-CV

Tabelle 9.26: Jeweils beste Modellierungsergebnisse mit Mamdani- und TSK-Modellen mit nur positiven Regeln für das Benchmarkproblem Mackey-Glass-Zeitreihe.

Operator ($N = 5, D = 70$)

Die Modellierung mit Mamdani-Modellen liefert, mit Ausnahme der Testverfahren J_{NHR} und J_{CHR} , für dieses Beispiel relativ ähnliche Ergebnisse. Die Modellierungsfehler liegen im Mittel bei $\epsilon_{norm} \approx 0.28$. Bei den Entwurfsvarianten für die TSK-Modelle zeigen sich deutliche Unterschiede. Als besonders gut geeignet erweist sich in diesem Fall die Variante TSK-LS-CV. Der Zugewinn an Modellgenauigkeit ist für dieses Problem sehr groß und beträgt zwischen 80% und 85%.

J	κ	τ	R	ϵ	ϵ_{norm}	η_{norm}	r	Modelltyp
J_{RI}	0.2000	0.0000	10	463.7588	0.2560	(0.1302)	0.9536	MAM
	0.2000	0.1000	7	133.9977	0.0661	(0.0115)	0.9947	TSK-LS-MV
	0.2000	0.1000	7	143.7397	0.0721	(0.0147)	0.9931	TSK-RR-MV
	0.2000	0.0000	10	106.6303	0.0538	(0.0055)	0.9977	TSK-LS-CV
	0.2000	0.0000	10	148.7558	0.0761	(0.0206)	0.9905	TSK-RR-CV
J_{RHR}	0.1000	0.0000	8	500.9550	0.2772	(0.1511)	0.9348	MAM
	0.2000	0.8000	15	120.5428	0.0604	(0.0084)	0.9958	TSK-LS-MV
	0.1000	0.0000	8	159.7956	0.0790	(0.0180)	0.9918	TSK-RR-MV
	0.2000	0.0000	17	107.7343	0.0546	(0.0064)	0.9969	TSK-LS-CV
	0.1000	0.0000	8	167.0406	0.0848	(0.0248)	0.9875	TSK-RR-CV
J_{NHR}	0.6000	0.6000	44	705.1297	0.3842	(0.3055)	0.8617	MAM
	0.6000	0.6000	44	383.1412	0.2033	(0.0889)	0.9480	TSK-LS-MV
	0.6000	0.6000	44	477.2607	0.2494	(0.1331)	0.9344	TSK-RR-MV
	0.3000	0.2000	93	142.1595	0.0700	(0.0195)	0.9913	TSK-LS-CV
	0.6000	0.1000	69	250.0773	0.1209	(0.0666)	0.9585	TSK-RR-CV
J_{CHR}	0.3000	0.4000	2	1642.9390	0.8258	(0.9440)	0.3272	MAM
	0.5000	0.0000	29	336.3605	0.1716	(0.0618)	0.9708	TSK-LS-MV
	0.5000	0.0000	29	710.9053	0.3519	(0.1630)	0.9487	TSK-RR-MV
	0.5000	0.0000	29	141.4055	0.0758	(0.0148)	0.9934	TSK-LS-CV
	0.5000	0.0000	29	568.8226	0.2696	(0.1850)	0.9196	TSK-RR-CV
J_{MVB}	0.0100	0.8000	26	534.7713	0.2891	(0.1685)	0.9277	MAM
	0.0100	0.0000	34	145.0178	0.0760	(0.0117)	0.9946	TSK-LS-MV
	0.0100	0.7000	29	219.6272	0.1162	(0.0325)	0.9835	TSK-RR-MV
	0.0100	0.5000	33	90.1108	0.0462	(0.0045)	0.9978	TSK-LS-CV
	0.0100	0.5000	33	136.2251	0.0714	(0.0195)	0.9906	TSK-RR-CV
J_{VVB}	0.0100	0.8000	19	508.0831	0.2734	(0.1724)	0.9275	MAM
	0.0100	0.0000	26	183.9579	0.0917	(0.0332)	0.9849	TSK-LS-MV
	0.0100	0.7000	23	248.6669	0.1288	(0.0544)	0.9767	TSK-RR-MV
	0.0100	0.0000	26	131.2488	0.0651	(0.0106)	0.9953	TSK-LS-CV
	0.0100	0.7000	23	167.2189	0.0859	(0.0236)	0.9891	TSK-RR-CV
J_{CVB}	0.0300	0.3000	11	511.0691	0.2817	(0.1474)	0.9315	MAM
	0.0300	0.0000	12	185.9883	0.0992	(0.0255)	0.9861	TSK-LS-MV
	0.0300	0.3000	11	254.3075	0.1405	(0.0649)	0.9670	TSK-RR-MV
	0.2000	0.2000	37	91.7651	0.0471	(0.0039)	0.9981	TSK-LS-CV
	0.0300	0.2000	12	202.8466	0.1115	(0.0422)	0.9794	TSK-RR-CV

Tabelle 9.27: Jeweils beste Modellierungsergebnisse mit Mamdani- und TSK-Modellen mit nur positiven Regeln für das Benchmarkproblem Operator.

Roboterkinematik ($N = 8, D = 8192$)

Die Modellierung mit Mamdani-Modellen liefert für dieses Beispiel nur sehr unbefriedigende Resultate (siehe auch Ergebnisse der Experimente A und B1). Die Modellierungsfehler liegen bei $\epsilon_{norm} \approx 0.8$ und sind damit nur unwesentlich besser als eine einfache Mittelwert-Prognose. Bei den TSK-Modellen liegen die Modellierungsfehler bei $\epsilon_{norm} \approx 0.5$, der Zugewinn an Modellgenauigkeit beträgt also auch hier etwa 35%. Allerdings ist das Ergebnis insgesamt trotzdem unbefriedigend, verglichen mit anderen Benchmarkproblemen, bei denen TSK-Modelle Modellierungsfehler von $\epsilon_{norm} < 0.3$ erzielen.

J	κ	τ	R	ϵ	ϵ_{norm}	η_{norm}	r	Modelltyp
J_{RI}	0.0001	0.3000	0	0.0675	1.0000	(1.0000)	0.0000	MAM
	0.2000	0.0000	548	0.0343	0.5076	(0.2583)	0.8616	TSK-LS-MV
	0.2000	0.0000	548	0.0342	0.5062	(0.2580)	0.8617	TSK-RR-MV
	0.2000	0.0000	548	0.0343	0.5087	(0.2591)	0.8614	TSK-LS-CV
	0.2000	0.0000	548	0.0342	0.5071	(0.2588)	0.8613	TSK-RR-CV
J_{RHR}	0.2000	0.0000	860	0.0535	0.7926	(0.6144)	0.7947	MAM
	0.0300	0.1000	427	0.0343	0.5083	(0.2593)	0.8610	TSK-LS-MV
	0.0300	0.0000	454	0.0342	0.5071	(0.2592)	0.8610	TSK-RR-MV
	0.1000	0.0000	661	0.0344	0.5090	(0.2599)	0.8609	TSK-LS-CV
	0.1000	0.0000	661	0.0343	0.5080	(0.2599)	0.8607	TSK-RR-CV
J_{NHR}	0.3000	0.4000	0	0.0675	1.0000	(1.0000)	0.0000	MAM
	0.3000	0.0000	105	0.0353	0.5223	(0.2813)	0.8488	TSK-LS-MV
	0.3000	0.0000	105	0.0354	0.5242	(0.2858)	0.8472	TSK-RR-MV
	0.3000	0.0000	105	0.0354	0.5246	(0.2840)	0.8472	TSK-LS-CV
	0.3000	0.0000	105	0.0356	0.5267	(0.2886)	0.8454	TSK-RR-CV
J_{CHR}	0.3000	0.2000	0	0.0675	1.0000	(1.0000)	0.0000	MAM
	0.3000	0.0000	8	0.0575	0.8515	(0.7850)	0.4899	TSK-LS-MV
	0.3000	0.0000	8	0.0575	0.8518	(0.7857)	0.4899	TSK-RR-MV
	0.3000	0.0000	8	0.0575	0.8516	(0.7851)	0.4898	TSK-LS-CV
	0.3000	0.0000	8	0.0575	0.8518	(0.7858)	0.4898	TSK-RR-CV
J_{MVB}	0.0001	0.4000	237	0.0525	0.7770	(0.6209)	0.7114	MAM
	0.0001	0.0000	341	0.0343	0.5087	(0.2602)	0.8605	TSK-LS-MV
	0.0001	0.0000	341	0.0343	0.5082	(0.2609)	0.8603	TSK-RR-MV
	0.0001	0.0000	341	0.0344	0.5095	(0.2608)	0.8603	TSK-LS-CV
	0.0001	0.0000	341	0.0344	0.5090	(0.2617)	0.8598	TSK-RR-CV
J_{VVB}	0.0300	0.1000	314	0.0547	0.8106	(0.6818)	0.7104	MAM
	0.2000	0.0000	442	0.0347	0.5147	(0.2695)	0.8562	TSK-LS-MV
	0.2000	0.0000	442	0.0349	0.5164	(0.2734)	0.8548	TSK-RR-MV
	0.2000	0.0000	442	0.0348	0.5148	(0.2699)	0.8559	TSK-LS-CV
	0.2000	0.0000	442	0.0349	0.5168	(0.2742)	0.8543	TSK-RR-CV
J_{CVB}	0.0001	0.4000	253	0.0526	0.7796	(0.6436)	0.7195	MAM
	0.0001	0.0000	382	0.0344	0.5095	(0.2608)	0.8601	TSK-LS-MV
	0.0001	0.0000	382	0.0344	0.5090	(0.2614)	0.8600	TSK-RR-MV
	0.0001	0.0000	382	0.0344	0.5102	(0.2612)	0.8600	TSK-LS-CV
	0.0001	0.0000	382	0.0344	0.5095	(0.2617)	0.8598	TSK-RR-CV

Tabelle 9.28: Jeweils beste Modellierungsergebnisse mit Mamdani- und TSK-Modellen mit nur positiven Regeln für das Benchmarkproblem Roboterkinematik (KinFH).

Box-Jenkins-Zeitreihe ($N = 10$, $D = 290$)

Die besten Ergebnisse mit Mamdani-Modellen (J_{VVB} , J_{MVB} und J_{CVB}) liefern Modellierungsfehler ϵ im Bereich $[0.38, 0.48]$ bzw. Modellierungsfehler ϵ_{norm} im Bereich $[0.44, 0.54]$. Der Zugewinn an Modellgenauigkeit durch den Übergang zu TSK-Modellen ist hier extrem hoch. Die TSK-Modelle liefern Modellierungsfehler $\epsilon \approx 0.9$ bzw. $\epsilon_{norm} \approx 0.1$. Der Fehler reduziert sich somit um 75% (zum Beispiel für J_{VVB}), in den extremsten Fällen sogar um ca. 85% (zum Beispiel für J_{RI}). Zwischen den verschiedenen Entwurfsvarianten für TSK-Modelle zeigen sich hier eher geringe Unterschiede, insgesamt scheint die Parameterbestimmung nach dem IRLS-Ansatz hier am besten geeignet zu sein. Es zeigt sich, dass durch die Transformation auch bei den Testverfahren, die vergleichsweise schlechte Mamdani-Modelle erzeugen, ähnlich oder gleich gute TSK-Modelle entstehen können wie bei den Testverfahren, die gute Mamdani-Modelle generieren. Die Unterschiede liegen dann eher in der Anzahl der verwendeten Regeln.

J	κ	τ	R	ϵ	ϵ_{norm}	η_{norm}	r	Modelltyp
J_{RI}	0.2000	0.0000	153	0.7725	0.8825	(0.6942)	0.9104	MAM
	0.2000	0.0000	153	0.0929	0.1061	(0.0158)	0.9922	TSK-LS-MV
	0.2000	0.0000	153	0.0910	0.1040	(0.0153)	0.9924	TSK-RR-MV
	0.0600	0.0000	77	0.0948	0.1083	(0.0157)	0.9922	TSK-LS-CV
	0.2000	0.0000	153	0.0918	0.1049	(0.0154)	0.9924	TSK-RR-CV
J_{RHR}	0.2000	0.4000	196	0.6703	0.7657	(0.5472)	0.9002	MAM
	0.2000	0.4000	196	0.0912	0.1041	(0.0156)	0.9923	TSK-LS-MV
	0.2000	0.0000	221	0.0891	0.1017	(0.0149)	0.9926	TSK-RR-MV
	0.2000	0.4000	196	0.0933	0.1065	(0.0160)	0.9921	TSK-LS-CV
	0.2000	0.3000	216	0.0896	0.1023	(0.0149)	0.9926	TSK-RR-CV
J_{NHR}	0.3000	0.0000	362	0.5741	0.6559	(0.4216)	0.9043	MAM
	0.3000	0.0000	362	0.0910	0.1039	(0.0160)	0.9921	TSK-LS-MV
	0.3000	0.0000	362	0.0885	0.1011	(0.0149)	0.9926	TSK-RR-MV
	0.3000	0.1000	314	0.0930	0.1063	(0.0158)	0.9918	TSK-LS-CV
	0.3000	0.0000	362	0.0892	0.1018	(0.0149)	0.9926	TSK-RR-CV
J_{CHR}	0.3000	0.0000	105	0.6124	0.6994	(0.4850)	0.8749	MAM
	0.3000	0.0000	105	0.0925	0.1057	(0.0172)	0.9916	TSK-LS-MV
	0.3000	0.0000	105	0.0904	0.1033	(0.0158)	0.9922	TSK-RR-MV
	0.3000	0.0000	105	0.0944	0.1079	(0.0177)	0.9913	TSK-LS-CV
	0.3000	0.0000	105	0.0921	0.1053	(0.0161)	0.9921	TSK-RR-CV
J_{MVB}	0.1000	0.5000	175	0.4565	0.5215	(0.2939)	0.9019	MAM
	0.2000	0.0000	234	0.0907	0.1036	(0.0163)	0.9920	TSK-LS-MV
	0.2000	0.4000	233	0.0877	0.1002	(0.0149)	0.9926	TSK-RR-MV
	0.2000	0.0000	234	0.0924	0.1056	(0.0173)	0.9915	TSK-LS-CV
	0.2000	0.1000	234	0.0884	0.1010	(0.0149)	0.9926	TSK-RR-CV
J_{VVB}	0.0100	0.4000	95	0.3863	0.4412	(0.2263)	0.9127	MAM
	0.2000	0.0000	241	0.0906	0.1035	(0.0157)	0.9920	TSK-LS-MV
	0.1000	0.0000	222	0.0875	0.1000	(0.0148)	0.9926	TSK-RR-MV
	0.0010	0.0000	70	0.0909	0.1038	(0.0156)	0.9923	TSK-LS-CV
	0.2000	0.1000	240	0.0881	0.1007	(0.0148)	0.9926	TSK-RR-CV
J_{CVB}	0.0300	0.6000	59	0.4718	0.5389	(0.3241)	0.8243	MAM
	0.2000	0.3000	193	0.0964	0.1101	(0.0182)	0.9910	TSK-LS-MV
	0.2000	0.3000	193	0.0926	0.1057	(0.0163)	0.9920	TSK-RR-MV
	0.2000	0.3000	193	0.0979	0.1118	(0.0186)	0.9908	TSK-LS-CV
	0.0100	0.4000	168	0.0931	0.1063	(0.0163)	0.9920	TSK-RR-CV

Tabelle 9.29: Jeweils beste Modellierungsergebnisse mit Mamdani- und TSK-Modellen mit nur positiven Regeln für das Benchmarkproblem Box-Jenkins-Zeitreihe.

Stock Prices ($N = 10, D = 100$)

Die besten Ergebnisse mit Mamdani-Modellen (J_{VVB}, J_{MVB} und J_{CVB}) liefern Modellierungsfehler ϵ im Bereich [4.7, 6.0] bzw. Modellierungsfehler ϵ_{norm} im Bereich [0.47, 0.60]. Bei den besten TSK-Modellen liegen die Modellierungsfehler ϵ im Bereich [3.3, 3.9] bzw. ϵ_{norm} im Bereich [0.30, 0.39]. Der Modellierungsfehler reduziert sich in etwa um 30%. Zwischen den verschiedenen Entwurfsvarianten für TSK-Modelle zeigen sich hier teilweise deutliche Unterschiede, beispielsweise für das Medianbasierte Test- und Bewertungsverfahren J_{CVB} . Auffällig ist, dass durch die Transformation für die Test- und Bewertungsverfahren (J_{RI} und J_{RHR}), die vergleichsweise schlechte Mamdani-Modelle erzeugen, genauere TSK-Modelle entstehen als für die Test- und Bewertungsverfahren, die die besten Mamdani-Modelle generieren. Außerdem kommen die besten TSK-Modelle mit gerade einmal fünf Regeln aus, wohingegen die meisten anderen TSK-Modelle ca. 60 bis 120 Regeln verwenden.

J	κ	τ	R	ϵ	ϵ_{norm}	η_{norm}	r	Modelltyp
J_{RI}	0.0100	0.0000	1	7.5268	0.7567	(0.6151)	0.7700	MAM
	0.0600	0.0000	5	3.3415	0.3338	(0.1760)	0.9103	TSK-LS-MV
	0.0600	0.0000	5	3.5042	0.3508	(0.2017)	0.8980	TSK-RR-MV
	0.0600	0.0000	5	3.7031	0.3699	(0.1986)	0.8978	TSK-LS-CV
	0.0600	0.0000	5	3.8019	0.3802	(0.2181)	0.8884	TSK-RR-CV
J_{RHR}	0.0100	0.0000	1	7.5268	0.7567	(0.6151)	0.7700	MAM
	0.0600	0.0000	5	3.3415	0.3338	(0.1760)	0.9103	TSK-LS-MV
	0.0600	0.0000	5	3.5042	0.3508	(0.2017)	0.8980	TSK-RR-MV
	0.0600	0.0000	5	3.7031	0.3699	(0.1986)	0.8978	TSK-LS-CV
	0.0600	0.0000	5	3.8019	0.3802	(0.2181)	0.8884	TSK-RR-CV
J_{NHR}	0.6000	0.1000	276	5.8118	0.5809	(0.4350)	0.7839	MAM
	0.7000	0.3000	187	5.1376	0.5139	(0.3055)	0.8178	TSK-LS-MV
	0.8000	0.0000	182	5.1444	0.5148	(0.3140)	0.8095	TSK-RR-MV
	0.4000	0.4000	276	4.2213	0.4221	(0.2260)	0.8959	TSK-LS-CV
	0.4000	0.4000	276	4.1689	0.4168	(0.2262)	0.8952	TSK-RR-CV
J_{CHR}	0.3000	0.4000	2	8.6372	0.8571	(0.8867)	0.5655	MAM
	0.3000	0.0000	161	4.6454	0.4616	(0.2643)	0.8592	TSK-LS-MV
	0.3000	0.0000	161	4.2704	0.4234	(0.2590)	0.8611	TSK-RR-MV
	0.3000	0.1000	36	4.0509	0.4047	(0.2139)	0.8892	TSK-LS-CV
	0.3000	0.0000	161	3.5868	0.3569	(0.1838)	0.9112	TSK-RR-CV
J_{MVB}	0.2000	0.6000	149	5.4307	0.5430	(0.3583)	0.8037	MAM
	0.0100	0.6000	60	4.1894	0.4166	(0.2377)	0.8750	TSK-LS-MV
	0.0100	0.6000	60	4.0319	0.4001	(0.2288)	0.8802	TSK-RR-MV
	0.0100	0.6000	60	3.8823	0.3860	(0.2066)	0.8866	TSK-LS-CV
	0.0100	0.6000	60	3.8277	0.3804	(0.2053)	0.8882	TSK-RR-CV
J_{VVB}	0.1000	0.4000	96	4.7723	0.4759	(0.2970)	0.8559	MAM
	0.0600	0.0000	5	3.3415	0.3338	(0.1767)	0.9103	TSK-LS-MV
	0.1000	0.1000	99	4.5182	0.4508	(0.2838)	0.8352	TSK-RR-MV
	0.0600	0.0000	5	3.7031	0.3699	(0.1986)	0.8978	TSK-LS-CV
	0.1000	0.3000	99	3.8941	0.3870	(0.2162)	0.8794	TSK-RR-CV
J_{CVB}	0.2000	0.6000	77	5.9287	0.5963	(0.4507)	0.7925	MAM
	0.0100	0.2000	29	4.6007	0.4553	(0.3135)	0.8297	TSK-LS-MV
	0.1000	0.0000	119	4.4794	0.4424	(0.3152)	0.8217	TSK-RR-MV
	0.1000	0.1000	116	3.6256	0.3616	(0.1933)	0.9020	TSK-LS-CV
	0.1000	0.1000	116	3.6459	0.3635	(0.1967)	0.9004	TSK-RR-CV

Tabelle 9.30: Jeweils beste Modellierungsergebnisse mit Mamdani- und TSK-Modellen mit nur positiven Regeln für das Benchmarkproblem Stock Prices.

Boston Housing ($N = 13, D = 506$)

Die besten Ergebnisse mit Mamdani-Modellen (J_{MVB}, J_{VVB} und J_{CVB}) liefern Modellierungsfehler ϵ im Bereich [3.7, 4.3] bzw. Modellierungsfehler ϵ_{norm} im Bereich [0.56, 0.65]. Durch den Übergang zu TSK-Modellen reduziert sich der Modellierungsfehler signifikant. Der Modellierungsfehler ϵ liegt dann im Bereich [2.3, 2.9] bzw. ϵ_{norm} im Bereich [0.37, 0.44]. Der Fehler reduziert sich um etwa 35%. Zwischen den Entwurfsvarianten für TSK-Modelle gibt es nur geringfügige Unterschiede. Auffällig ist auch hier, dass durch die Transformation auch aus vergleichsweise schlechten Mamdani-Modellen gute TSK-Modelle entstehen können (siehe auch Benchmarkproblem Stock Prices).

J	κ	τ	R	ϵ	ϵ_{norm}	η_{norm}	r	Modelltyp
J_{RI}	0.0100	0.3000	9	5.4865	0.8251	(0.6576)	0.7729	MAM
	0.2000	0.0000	277	2.5466	0.3832	(0.1661)	0.9148	TSK-LS-MV
	0.2000	0.0000	277	2.8896	0.4351	(0.2638)	0.8675	TSK-RR-MV
	0.2000	0.0000	277	2.5271	0.3803	(0.1605)	0.9168	TSK-LS-CV
	0.2000	0.0000	277	2.8862	0.4344	(0.2825)	0.8488	TSK-RR-CV
J_{RHR}	0.2000	0.7000	204	4.5750	0.6887	(0.5440)	0.7784	MAM
	0.2000	0.7000	204	2.4239	0.3647	(0.1674)	0.9127	TSK-LS-MV
	0.2000	0.7000	204	2.6465	0.3985	(0.2374)	0.8793	TSK-RR-MV
	0.1000	0.6000	186	2.4587	0.3699	(0.1582)	0.9151	TSK-LS-CV
	0.2000	0.7000	204	2.6963	0.4061	(0.2589)	0.8653	TSK-RR-CV
J_{NHR}	0.6000	0.3000	393	4.3971	0.6618	(0.5186)	0.7927	MAM
	0.4000	0.4000	572	2.4555	0.3694	(0.1672)	0.9187	TSK-LS-MV
	0.5000	0.5000	334	2.7824	0.4189	(0.2559)	0.8650	TSK-RR-MV
	0.5000	0.0000	797	2.3893	0.3596	(0.1599)	0.9202	TSK-LS-CV
	0.6000	0.0000	650	2.6539	0.3997	(0.2419)	0.8826	TSK-RR-CV
J_{CHR}	0.6000	0.2000	4	6.3825	0.9601	(0.9565)	0.3189	MAM
	0.5000	0.1000	126	2.7822	0.4184	(0.2180)	0.8847	TSK-LS-MV
	0.5000	0.1000	126	2.7602	0.4156	(0.2427)	0.8862	TSK-RR-MV
	0.4000	0.1000	218	2.4990	0.3761	(0.1581)	0.9180	TSK-LS-CV
	0.5000	0.1000	126	2.6547	0.3996	(0.2363)	0.8798	TSK-RR-CV
J_{MVB}	0.0100	0.5000	179	3.7298	0.5617	(0.3555)	0.8134	MAM
	0.0100	0.4000	248	2.4173	0.3638	(0.1695)	0.9109	TSK-LS-MV
	0.0100	0.5000	179	2.7066	0.4075	(0.2432)	0.8744	TSK-RR-MV
	0.1000	0.4000	453	2.4335	0.3662	(0.1681)	0.9072	TSK-LS-CV
	0.0100	0.4000	248	2.6788	0.4035	(0.2320)	0.8809	TSK-RR-CV
J_{VVB}	0.1000	0.3000	366	3.9460	0.5944	(0.3871)	0.7883	MAM
	0.0300	0.0000	343	2.4752	0.3726	(0.1927)	0.8994	TSK-LS-MV
	0.0300	0.2000	339	2.9056	0.4378	(0.2831)	0.8637	TSK-RR-MV
	0.1000	0.0000	428	2.3951	0.3604	(0.1837)	0.9064	TSK-LS-CV
	0.2000	0.0000	490	2.7859	0.4196	(0.2596)	0.8762	TSK-RR-CV
J_{CVB}	0.0001	0.0000	217	4.2575	0.6415	(0.4458)	0.7543	MAM
	0.0010	0.4000	156	2.5300	0.3807	(0.1735)	0.9077	TSK-LS-MV
	0.0010	0.4000	156	2.8774	0.4336	(0.2644)	0.8663	TSK-RR-MV
	0.2000	0.4000	425	2.5086	0.3775	(0.1647)	0.9175	TSK-LS-CV
	0.0010	0.4000	156	2.8634	0.4315	(0.2573)	0.8678	TSK-RR-CV

Tabelle 9.31: Jeweils beste Modellierungsergebnisse mit Mamdani- und TSK-Modellen mit nur positiven Regeln für das Benchmarkproblem Boston Housing.

Fazit

Das Ziel der Transformation von Mamdani-Modellen in TSK-Modelle ist die Verbesserung der Modellgenauigkeit bei gleichzeitiger Erhaltung der Interpretierbarkeit. In den hier untersuchten Benchmarkproblemen werden Verbesserungen bezüglich des Modellierungsfehlers ϵ gegenüber den Mamdani-Modellen erzielt, die sich zwischen 30% und teilweise sogar über 80% bewegen. Vergleichbare Ergebnisse sind auch in früheren Arbeiten erzielt worden, wobei die Anzahl der untersuchten Benchmarkprobleme⁷ aber deutlich geringer war. Das verfolgte Ziel wird in allen Fällen klar erreicht. Aufgrund der Anzahl der untersuchten Benchmarkprobleme und der erzielten Ergebnisse erscheint es durchaus realistisch, auch für andere bisher nicht untersuchte Approximationsprobleme durch die Transformation von Mamdani-Modellen in TSK-Modelle eine Verbesserung des Modellierungsfehlers in der Größenordnung von 25% bis 50% erwarten zu können. Im Hinblick auf die Entwurfsvarianten für TSK-Modelle zeichnet sich keine eindeutig beste Lösung ab. Die Unterschiede zwischen den Entwurfsvarianten in den erzielten Modellierungsfehlern ϵ sind teilweise nur gering, in anderen Fällen teilweise aber auch deutlich und fallen insgesamt stark problemspezifisch aus. Im Hinblick auf die Bestimmung der freien Parameter in den Konklusionen der TSK-Regeln zeichnen sich insgesamt leichte Vorteile für den Iteratively Reweighted Least-Squares Ansatz ab. Im Hinblick auf die Berechnung des Ausgabewertes des TSK-Modells zeigen sich in einigen Fällen keine Unterschiede zwischen der Verwendung des gewichteten Mittelwertes und des gewichteten Median, in vier Fällen liefert jedoch der gewichtete Median bessere Resultate. Es erscheint daher durchaus vertretbar, die Berechnung des Modellausgabewertes über den gewichteten Median als Standard-Einstellung zu verwenden. Zusammenfassend erscheint eine Kombination aus IRLS-Ansatz für die Parameterbestimmung und Berechnung des Ausgabewertes über den gewichteten Median empfehlenswert. Die neu entwickelten Strategieelemente für TSK-Systeme haben damit ihre Leistungsfähigkeit erwiesen. Es zeigt sich zudem in einer Reihe von Fällen, dass die Parameterbestimmung über den IRLS-Ansatz nicht zwangsläufig einen größeren normierten mittleren quadratischen Fehler η_{norm} zur Folge hat als die Parameterbestimmung über den OLS-Ansatz, da für den resultierenden Fehler das Zusammenwirken aller Regeln maßgebend ist.

⁷In [Kra02] werden Ergebnisse für die Benchmarkprobleme Wärmetauscher und Mackey-Glass-Zeitreihe vorgestellt. Ferner werden dort auch Ergebnisse für andere Funktionsansätze (zum Beispiel eine multilineare Funktion) in der Konklusion einer TSK-Regel aufgeführt.

Ergebnisse Experiment B3

Ziel dieses Experiments ist es, das Potenzial negativer Regeln für die datenbasierte Modellierung zu untersuchen. Der prinzipielle Nutzen negativer Regeln ist bereits in Vorarbeiten gezeigt worden, eine flächendeckende Untersuchung an Benchmarkproblemen stand aber aus.

Im Rahmen dieser Arbeit werden die Weak-Veto Hyperinferenzstrategie und die Center of Largest Area Hyperdefuzzifizierungsstrategie verwendet. Aufgrund ihres einschränkenden Charakters können bereits wenige negative Regeln einen gravierenden Einfluss auf die erzielte Modellierungsqualität haben. Daher wird die Strong-Veto Hyperinferenz, die sehr rigoros wirkt, nicht eingesetzt. Bei der Fuzzy-Veto Hyperinferenz ergeben sich für die ausgangsseitige Zugehörigkeitsfunktion die gleichen Intervalle, in denen $\mu(y) = 0$ gilt, wie bei der Weak-Veto Hyperinferenz. Da die beiden Strategien somit identisch im Hinblick auf die Verwendung des Default-Wertes sind, wird die Fuzzy-Veto Hyperinferenz nicht untersucht. Wie in Abschnitt 8.1 ausgeführt, unterstützen nur bestimmte Test- und Bewertungsverfahren die Erzeugung negativer Mamdani-Regeln. Bei der Untersuchung negativer Mamdani-Regeln sind daher verschiedene Fragestellungen von Interesse: zum einen, ob bzw. welche Verbesserungen bezüglich derjenigen Test- und Bewertungsverfahren eintreten, die negative Regeln unterstützten, zum anderen, wie sich die Ergebnisse dieser Verfahren mit denen der Verfahren vergleichen, die keine negativen Regeln erzeugen. Weiterhin stellt sich die Frage, wie das Verhältnis von Mehraufwand – einer meist größeren Regelbasis und einer komplexeren Verarbeitung der Regeln – zu dem erzielten Zugewinn – der Verbesserung der Modellgenauigkeit – ausfällt.

Ausgewählte Ergebnisse werden in Form von Tabellen präsentiert. Diese sind vom Aufbau her identisch mit denen, die für die Vorstellung der Ergebnisse des Experiments B1 verwendet werden.

Wärmetauscher ($N = 3$, $D = 1126$)

Durch die Hinzunahme negativer Regeln verbessern sich die Ergebnisse für diejenigen Test- und Bewertungsverfahren, die Regeln dieser Art erzeugen können. Dargestellt sind hier beispielhaft die Modellierungsfehler für den Relevanzindex J_{RI} . Verglichen mit dem Fall nur positiver Regeln reduziert sich der Modellierungsfehler im Mittel um ca. 5% bis 10%. Für zum Beispiel $\kappa = 0.2000$ und $\tau = 0.3$ reduziert sich der Fehler von $\epsilon_{norm} = 0.56$ auf $\epsilon_{norm} = 0.50$. Dies entspricht einer Verbesserung von ca. 12%. Die Anzahl der Regeln ist in etwa doppelt so hoch wie zuvor und steigt von $R = 16$ auf $R = 37$. Dieses ist das beste Ergebnis des Relevanzindex J_{RI} , das, verglichen mit den besten Ergebnissen des Mittelwertbasierten Verfahrens J_{MVB} , deutlich schlechter abschneidet. Die besten Ergebnisse des Verfahrens J_{MVB} liegen in etwa bei $\epsilon_{norm} = 0.40$, sind also um ca. 20% besser.

J	κ		$\tau=0.0$	$\tau=0.1$	$\tau=0.2$	$\tau=0.3$	$\tau=0.4$	$\tau=0.5$	$\tau=0.6$	$\tau=0.7$	$\tau=0.8$	$\tau=0.9$
J_{RI}	0.0001	R	24	15	12	10	8	3	2	0	0	0
		ϵ_{norm}	0.67	0.62	0.59	0.62	0.66	0.92	1.01	1.01	1.01	1.01
	0.0010	R	30	22	14	12	11	7	3	1	0	0
		ϵ_{norm}	0.72	0.62	0.59	0.59	0.61	0.66	1.01	1.01	1.01	1.01
	0.0100	R	38	31	24	19	14	10	6	3	0	0
		ϵ_{norm}	0.70	0.58	0.56	0.56	0.58	0.62	0.80	1.01	1.01	1.01
	0.0300	R	47	37	30	21	18	13	9	4	1	0
		ϵ_{norm}	0.76	0.63	0.55	0.55	0.54	0.58	0.62	1.01	1.01	1.01
	0.0600	R	52	43	35	27	19	17	10	4	2	0
		ϵ_{norm}	0.69	0.67	0.54	0.53	0.54	0.55	0.64	1.01	1.01	1.01
	0.1000	R	60	47	38	30	22	19	12	6	3	0
		ϵ_{norm}	0.69	0.65	0.54	0.53	0.55	0.54	0.61	0.80	1.01	1.01
	0.2000	R	73	53	45	37	32	22	18	8	4	0
		ϵ_{norm}	0.67	0.62	0.54	0.50	0.52	0.53	0.55	0.72	1.01	1.01

Tabelle 9.32: Ergebnisse für das Test- und Bewertungsverfahren J_{RI} mit positiven und negativen Regeln für das Benchmarkproblem Wärmetauscher.

Box-Jenkins-Zeitreihe ($N = 10, D = 290$)

Die Modellierungsergebnisse für den Relevanzindex J_{RI} sind bei Verwendung nur positiver Regeln insgesamt sehr schlecht. Das beste Ergebnis $\epsilon_{norm} = 0.88$ wird mit $\kappa = 0.2000$ und $\tau = 0.0$ erzielt. Durch die Hinzunahme negativer Regeln kann das Ergebnis für den Relevanzindex J_{RI} hier in vielen Fällen deutlich verbessert werden. Das beste Ergebnis $\epsilon_{norm} = 0.66$ wird mit $\kappa = 0.2000$ und $\tau = 0.3$ erzielt. Für diesen speziellen Fall halbiert sich der Fehler gegenüber dem entsprechenden Modell mit nur positiven Regeln, allerdings vervierfacht sich die Anzahl der Regeln in etwa. Verglichen mit den Ergebnissen des Varianzbasierten Verfahrens J_{VVB} fällt auf, dass für dieses Verfahren häufig normierte Fehlerwerte von $\epsilon_{norm} \approx 0.45$ erzielt werden, die damit deutlich besser sind. Die Hinzunahme negativer Regeln verbessert zwar die Performance des Relevanzindex J_{RI} , diese bleibt aber insgesamt schlechter als die des Varianzbasierten Verfahrens J_{VVB} .

J	κ		$\tau=0.0$	$\tau=0.1$	$\tau=0.2$	$\tau=0.3$	$\tau=0.4$	$\tau=0.5$	$\tau=0.6$	$\tau=0.7$	$\tau=0.8$	$\tau=0.9$
J_{RI}	0.0001	R	0	0	0	0	0	0	0	0	0	0
		ϵ_{norm}	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.0010	R	8	0	0	0	0	0	0	0	0	0
		ϵ_{norm}	1.10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.0100	R	36	10	2	0	0	0	0	0	0	0
		ϵ_{norm}	1.59	1.14	0.97	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.0300	R	57	24	6	0	0	0	0	0	0	0
		ϵ_{norm}	1.49	1.46	1.10	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.0600	R	77	39	12	4	0	0	0	0	0	0
		ϵ_{norm}	1.20	1.60	1.20	0.97	1.00	1.00	1.00	1.00	1.00	1.00
	0.1000	R	96	54	23	6	1	0	0	0	0	0
		ϵ_{norm}	1.04	1.49	1.40	1.10	0.96	1.00	1.00	1.00	1.00	1.00
	0.2000	R	153	89	45	15	4	0	0	0	0	0
		ϵ_{norm}	0.88	1.04	1.60	1.33	0.97	1.00	1.00	1.00	1.00	1.00

Tabelle 9.33: Ergebnisse für das Test- und Bewertungsverfahren J_{RI} mit nur positiven Regeln für das Benchmarkproblem Box-Jenkins-Zeitreihe.

J	κ		$\tau=0.0$	$\tau=0.1$	$\tau=0.2$	$\tau=0.3$	$\tau=0.4$	$\tau=0.5$	$\tau=0.6$	$\tau=0.7$	$\tau=0.8$	$\tau=0.9$	
J_{RI}	0.0001	R	5	1	0	0	0	0	0	0	0	0	
		ϵ_{norm}	0.99	0.97	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.0010	R	23	5	2	0	0	0	0	0	0	0	0
		ϵ_{norm}	0.85	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.0100	R	60	31	6	4	1	0	0	0	0	0	0
		ϵ_{norm}	0.93	0.81	0.99	0.99	0.97	1.00	1.00	1.00	1.00	1.00	1.00
	0.0300	R	101	55	27	7	3	0	0	0	0	0	0
		ϵ_{norm}	0.97	0.81	0.89	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.00
	0.0600	R	135	81	50	22	6	2	0	0	0	0	0
		ϵ_{norm}	1.00	0.90	0.76	0.84	0.99	0.98	1.00	1.00	1.00	1.00	1.00
	0.1000	R	172	107	62	37	12	3	0	0	0	0	0
		ϵ_{norm}	0.99	0.90	0.73	0.75	0.93	0.99	1.00	1.00	1.00	1.00	1.00
	0.2000	R	234	184	102	62	35	12	3	0	0	0	0
		ϵ_{norm}	0.91	0.87	0.73	0.66	0.75	0.93	0.99	1.00	1.00	1.00	1.00

Tabelle 9.34: Ergebnisse für das Test- und Bewertungsverfahren J_{RI} mit positiven und negativen Regeln für das Benchmarkproblem Box-Jenkins-Zeitreihe.

Fazit

Die Ergebnisse zeigen, dass die Hinzunahme negativer Regeln durchaus zu einer Verbesserung der Modellgenauigkeit führen kann, wobei diese meist mit einer größeren Regelbasis einhergeht. Dies legt die Vermutung nahe, dass die Anforderungen an die Erzeugung negativer Regeln unter Umständen *zu leicht* gewählt sind. Andererseits fällt es aber auch Menschen oftmals leichter, zu formulieren, was man lieber nicht tun sollte, als präzise auszudrücken, was man tun sollte. Beispielsweise sollte ein Autofahrer bei vereister Straße starkes Bremsen oder abruptes Lenken vermeiden. Es lässt sich aber wesentlich schlechter beschreiben, wie man in dieser Situation konkret bremsen oder lenken soll. Die Anzahl negativer Regeln ist daher ein eher unkritischer Faktor, zumal in den negativen Regeln häufig noch Redundanzen enthalten sein dürften. Durch das Ausdünnen der Regelbasis nach der Regelrelevanz lässt sich dies nicht sicher feststellen. Hierfür müsste beispielsweise die strukturbasierte Regelreduktion (siehe Abschnitt 4.3) eingesetzt werden. Aus Zeitgründen konnten im Rahmen dieser Arbeit keine flächendeckenden Untersuchungen hierzu durchgeführt werden. Es zeigt sich ferner, dass oftmals Test- und Bewertungsverfahren, die keine negativen Regeln verwenden, die insgesamt besten Ergebnisse erzielen. Wenn also, wie es hier der Fall ist, primär eine hohe Modellgenauigkeit von Interesse ist, so bedeutet der Verzicht auf negative Regeln nicht zwangsläufig eine Einschränkung der erreichbaren Modellgenauigkeit. Ist aber die kompakte und transparente Darstellung entsprechender Zusammenhänge – beispielsweise sicherheitsrelevanter Aspekte – oder die Einbringung menschlicher Erfahrungen von Bedeutung, so sind negative Regeln vorteilhaft.

Ergebnisse Experiment B4

Der Einsatz negativer Regeln in TSK-Systemen war bisher nicht möglich. Durch den in Abschnitt 8.3 vorgestellten Ansatz für ein hybrides Fuzzy-System wird diese Lücke ge-

geschlossen. Der prinzipielle Nutzen dieses Ansatzes wird in Abschnitt 8.4 an einem Beispiel gezeigt. Ziel dieses Experiments ist es, die Leistungsfähigkeit dieses Ansatzes an einer Reihe von Benchmarkproblemen zu evaluieren.

Für das Experiment B4 werden nicht alle Entwurfsvarianten für TSK-Systeme (siehe Tabelle 9.24) untersucht. Die Parameterbestimmung erfolgt sowohl über den OLS-Ansatz als auch über den IRLS-Ansatz, die Berechnung des Ausgabewertes erfolgt aber ausschließlich über den gewichteten Mittelwert. In den nachfolgenden Tabellen sind zur besseren Einordnung der erzielten Ergebnisse auch die Resultate hybrider Fuzzy-Systeme angegeben, die mit Test- und Bewertungsverfahren erstellt werden, die keine negativen Regeln unterstützen. Vergleicht man die entsprechenden Einträge in den Tabellen mit denen des Experiments B2, so zeigen sich Unterschiede, die wie folgt erklärt werden können: Wie in Abschnitt 8.1 erläutert, kann die resultierende ausgangsseitige Zugehörigkeitsfunktion aus qualitativ unterschiedlichen Gründen aus mehreren Teilempfehlungsgebirgen bestehen. Dies kann daher auch für die genannten Test- und Bewertungsverfahren der Fall sein, ist dann aber nicht die Folge negativer Regeln, sondern resultiert lediglich daher, dass bestimmte Ausgangsgrößenwerte einfach nicht von positiven Regeln empfohlen werden. Auch in einem solchen Fall wird in diesem Experiment das Teilempfehlungsgebirge mit dem größten Gewicht ausgewählt und für die in Abschnitt 8.3 beschriebene Selektion der TSK-Regeln genutzt. Falls die resultierende ausgangsseitige Zugehörigkeitsfunktion des zu Grunde liegenden Mamdani-Systems nicht zusammenhängend ist, erfolgt daher auch in dem Fall, dass keine negativen Regeln wirken, eine Regelselektion. Hierauf sind die genannten Differenzen zurückzuführen.

Die Ergebnisse werden in Form von Tabellen präsentiert. Diese sind vom Aufbau her im Prinzip identisch mit denen, die für die Vorstellung der Ergebnisse des Experiments B2 verwendet werden. Auf die Angabe des normierten quadratischen Approximationsfehlers η_{norm} wird jedoch verzichtet.

Wärmetauscher ($N = 3$, $D = 1126$)

Für die Test- und Bewertungsverfahren, die negative Regeln unterstützen, ergeben sich für dieses Problem gegenüber den zu Grunde liegenden Mamdani-Systemen des Experiments B3 Verbesserungen der Modellgenauigkeit, wie sie sich auch zwischen den Experimenten B1 und B2 zeigen. Ein Vergleich der Modellierungsfehler der hybriden Systeme mit denen der gewöhnlichen TSK-Systeme (Experiment B2) zeigt nur sehr geringe Unterschiede. Die bereits nur geringen Unterschiede, die sich für dieses Problem zwischen den Mamdani-Systemen ohne und mit negativen Regeln zeigen (Experimente B1 und B3), fallen hier für die TSK-Systeme (Experimente B2 und B4) noch geringer aus.

J	κ	τ	R	ϵ	ϵ_{norm}	r	Modelltyp
J_{RI}	0.2000	0.3000	37	3.7539	0.5049	0.8002	MAM
	0.2000	0.1000	53	1.8502	0.2497	0.9457	TSK-LS-MV
	0.2000	0.1000	53	1.8322	0.2477	0.9256	TSK-RR-MV
J_{RHR}	0.1000	0.7000	47	3.4436	0.4641	0.8495	MAM
	0.2000	0.2000	84	1.8145	0.2447	0.9518	TSK-LS-MV
	0.2000	0.2000	84	1.8195	0.2451	0.9351	TSK-RR-MV
J_{NHR}	0.7000	0.0000	147	3.3325	0.4491	0.8312	MAM
	0.6000	0.2000	150	1.9914	0.2686	0.9444	TSK-LS-MV
	0.6000	0.3000	147	2.0919	0.2819	0.9239	TSK-RR-MV
J_{CHR}	0.6000	0.0000	139	3.9155	0.5273	0.7330	MAM
	0.3000	0.0000	161	1.9842	0.2670	0.8983	TSK-LS-MV
	0.3000	0.0000	161	1.7693	0.2374	0.9094	TSK-RR-MV
J_{MVB}	0.2000	0.2000	37	3.0335	0.4086	0.9047	MAM
	0.0600	0.2000	33	1.6966	0.2292	0.9413	TSK-LS-MV
	0.2000	0.2000	37	1.7226	0.2323	0.9264	TSK-RR-MV
J_{VVB}	0.2000	0.4000	18	3.4895	0.4693	0.8123	MAM
	0.2000	0.4000	18	2.0833	0.2797	0.8557	TSK-LS-MV
	0.2000	0.4000	18	1.9345	0.2600	0.8445	TSK-RR-MV
J_{CVB}	0.2000	0.0000	38	3.4076	0.4604	0.8408	MAM
	0.1000	0.0000	37	1.7553	0.2368	0.9423	TSK-LS-MV
	0.0600	0.0000	36	1.6531	0.2233	0.9269	TSK-RR-MV

Tabelle 9.35: Jeweils beste Modellierungsergebnisse mit Mamdani- und TSK-Modellen mit positiven und negativen Regeln für das Benchmarkproblem Wärmetauscher.

Box-Jenkins-Zeitreihe ($N = 10, D = 290$)

Die Bewertung fällt im Prinzip identisch mit dem Benchmarkproblem Wärmetauscher aus. Der Zugewinn an Modellgenauigkeit gegenüber den Mamdani-Modellen ist wie erwartet deutlich. Aber auch für dieses Problem ergibt sich kein signifikanter Unterschied zu den herkömmlichen TSK-Modellen.

J	κ	τ	R	ϵ	ϵ_{norm}	r	Modelltyp
J_{RI}	0.2000	0.3000	62	0.5737	0.6556	0.9185	MAM
	0.2000	0.1000	184	0.0908	0.1037	0.9926	TSK-LS-MV
	0.2000	0.1000	184	0.0889	0.1015	0.9928	TSK-RR-MV
J_{RHR}	0.2000	0.6000	318	0.4319	0.4934	0.9388	MAM
	0.2000	0.3000	482	0.0889	0.1016	0.9928	TSK-LS-MV
	0.2000	0.3000	482	0.0870	0.0994	0.9930	TSK-RR-MV
J_{NHR}	0.3000	0.0000	1312	0.4050	0.4628	0.8845	MAM
	0.4000	0.2000	1111	0.0886	0.1012	0.9929	TSK-LS-MV
	0.4000	0.1000	1181	0.0860	0.0982	0.9932	TSK-RR-MV
J_{CHR}	0.6000	0.0000	455	0.4342	0.4959	0.8844	MAM
	0.3000	0.3000	892	0.0878	0.1003	0.9928	TSK-LS-MV
	0.3000	0.1000	940	0.0857	0.0979	0.9931	TSK-RR-MV
J_{MVB}	0.1000	0.6000	194	0.3282	0.3749	0.9375	MAM
	0.2000	0.0000	363	0.0883	0.1009	0.9930	TSK-LS-MV
	0.2000	0.0000	363	0.0870	0.0994	0.9932	TSK-RR-MV
J_{VVB}	0.0300	0.4000	255	0.3125	0.3570	0.9398	MAM
	0.2000	0.0000	346	0.0898	0.1025	0.9928	TSK-LS-MV
	0.2000	0.0000	346	0.0878	0.1003	0.9931	TSK-RR-MV
J_{CVB}	0.2000	0.6000	193	0.3628	0.4144	0.9111	MAM
	0.2000	0.0000	335	0.1005	0.1148	0.9871	TSK-LS-MV
	0.2000	0.0000	335	0.1088	0.1243	0.9673	TSK-RR-MV

Tabelle 9.36: Jeweils beste Modellierungsergebnisse mit Mamdani- und TSK-Modellen mit positiven und negativen Regeln für das Benchmarkproblem Box-Jenkins-Zeitreihe.

Fazit

Zusammenfassend ist festzustellen, dass durch die Transformation von Mamdani-Modellen mit negativen Regeln in hybride Fuzzy-Modelle im Prinzip identische Verbesserungen erzielt werden wie bei der Transformation von Mamdani-Modellen mit nur positiven Regeln in herkömmliche TSK-Modelle. Der Vergleich der Mamdani-Systeme ohne und mit negativen Regeln (Experimente B1 und B3) zeigt, dass sich durch die Hinzunahme negativer Regeln oftmals durchaus Verbesserungen erzielen lassen. Aufgrund des einschränkenden Charakters negativer Regeln fallen diese Verbesserungen aber erwartungsgemäß nicht so groß aus wie beispielsweise bei der Umwandlung positiver Mamdani-Regeln in entsprechende TSK-Regeln. Daher sind auch zwischen den Experimenten B2 und B4 keine großen Veränderungen der Modellierungsfehler zu erwarten, was die gezeigten Ergebnisse auch bestätigen. Der prinzipielle Vorteil des hybriden Fuzzy-Systems – eine hohe Approximationsgenauigkeit bei gleichzeitiger Berücksichtigung von Warnungen oder Verboten – wird hierdurch jedoch nicht berührt.

Ergebnisse Experiment B5

Das von KIENDL vorgeschlagene Konzept des Inkrementellen Relevanzindex zur Regelgenerierung ist in Vorarbeiten prototypisch erprobt worden. Abweichungen in der bisherigen Implementierung zu dem in der Literatur beschriebenen Konzept machten eine Neuimplementierung notwendig.

Das Konzept des Inkrementellen Relevanzindex zielt auf eine möglichst gleichberechtigte Behandlung positiver und negativer Regeln ab. Die Basis dieses Konzepts bildet ein Satz universeller Regeln, die stets aktiviert werden und so einen Basis-Ausgabewert des Modells liefern. Die universellen Regeln werden so festgelegt, dass der Basis-Ausgabewert dem mittleren Ausgangsgrößenwert der Lerndatenpunkte entspricht. Alle hinzukommenden Regeln wirken inkrementell und erzeugen so Abweichungen von dem Basis-Ausgabewert. Speziell für dieses Konzept wurden zwei Test- und Bewertungsverfahren – die inkrementelle Trefferquote J_{ITQ} und die konfidente inkrementelle Trefferquote J_{KITQ} – konzipiert, die hier zusätzlich zu den anderen sieben Verfahren eingesetzt werden.

Wärmetauscher ($N = 3$, $D = 1126$)

Im Experiment B3 wird für den Relevanzindex J_{RI} das beste Ergebnis $\epsilon_{norm} = 0.50$ für $\kappa = 0.2000$ und $\tau = 0.3$ mit $R = 37$ Regeln erzielt. Bei dieser Parametrierung ergibt sich in diesem Experiment ein Wert von $\epsilon_{norm} = 0.45$ mit $R = 62$ Regeln. Die Modellgenauigkeit verbessert sich um ca. 10%, dafür steigt die Anzahl der Regeln um ca. ein Drittel. Das beste Ergebnis wird hier mit $\kappa = 0.0300$ und $\tau = 0.7$ erzielt und liegt bei $\epsilon_{norm} = 0.40$. In

diesem Fall werden nur $R = 10$ Regeln verwendet. Die Ergebnisse mit den Verfahren J_{ITQ} und J_{KITQ} fallen für dieses Problem schlechter aus als die Ergebnisse des Relevanzindex J_{RI} .

J	κ		$\tau=0.0$	$\tau=0.1$	$\tau=0.2$	$\tau=0.3$	$\tau=0.4$	$\tau=0.5$	$\tau=0.6$	$\tau=0.7$	$\tau=0.8$	$\tau=0.9$
J_{RI}	0.0001	R	150	45	25	19	14	9	8	6	6	6
		ϵ_{norm}	0.52	0.49	0.51	0.50	0.48	0.62	0.85	1.09	1.09	1.09
	0.0010	R	186	57	31	23	19	13	9	7	6	6
		ϵ_{norm}	0.52	0.49	0.52	0.51	0.51	0.47	0.58	1.07	1.09	1.09
	0.0100	R	234	79	48	32	25	17	12	9	6	6
		ϵ_{norm}	0.52	0.47	0.42	0.48	0.52	0.50	0.43	0.59	1.09	1.09
	0.0300	R	288	92	60	36	30	22	15	10	7	6
		ϵ_{norm}	0.53	0.46	0.45	0.49	0.50	0.51	0.49	0.40	1.01	1.09
	0.0600	R	315	107	69	45	33	26	17	10	8	6
		ϵ_{norm}	0.54	0.47	0.44	0.48	0.50	0.47	0.50	0.41	0.91	1.09
	0.1000	R	361	119	75	48	37	29	20	12	9	6
		ϵ_{norm}	0.55	0.48	0.44	0.47	0.50	0.50	0.51	0.45	0.60	1.09
	0.2000	R	444	139	94	62	49	35	26	14	10	6
		ϵ_{norm}	0.56	0.48	0.43	0.45	0.48	0.51	0.48	0.47	0.41	1.09

Tabelle 9.37: Ergebnisse für das Test- und Bewertungsverfahren J_{RI} mit universellen und positiven und negativen Regeln für das Benchmarkproblem Wärmetauscher.

Box-Jenkins-Zeitreihe ($N = 10$, $D = 290$)

Im Experiment B3 wird für den Relevanzindex J_{RI} das beste Ergebnis $\epsilon_{norm} = 0.66$ für $\kappa = 0.2000$ und $\tau = 0.3$ mit $R = 62$ Regeln erzielt. Bei dieser Parametrierung ergibt sich in diesem Experiment ein Wert von $\epsilon_{norm} = 0.58$ mit $R = 67$ Regeln. Die Modellgenauigkeit verbessert sich um ca. 12%, wobei sich die Anzahl der Regeln kaum erhöht. Das beste Ergebnis $\epsilon_{norm} = 0.45$ wird hier für $\kappa = 0.0300$ und $\tau = 0.2$ erzielt. In diesem Fall beträgt der Zugewinn an Modellgenauigkeit gegenüber dem besten Ergebnis aus dem Experiment B3 ca. 32% und es werden nur $R = 33$ statt $R = 62$ Regeln verwendet. Mit den Verfahren J_{ITQ} und J_{KITQ} lassen sich für dieses Problem vergleichbare Ergebnisse erzielen wie mit dem Relevanzindex J_{RI} , allerdings ist die Anzahl der verwendeten Regeln dabei größer.

J	κ		$\tau=0.0$	$\tau=0.1$	$\tau=0.2$	$\tau=0.3$	$\tau=0.4$	$\tau=0.5$	$\tau=0.6$	$\tau=0.7$	$\tau=0.8$	$\tau=0.9$
J_{RI}	0.0001	R	28	6	5	5	5	5	5	5	5	5
		ϵ_{norm}	0.91	0.95	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.0010	R	120	11	7	5	5	5	5	5	5	5
		ϵ_{norm}	0.76	0.88	0.91	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.0100	R	305	43	11	9	6	5	5	5	5	5
		ϵ_{norm}	0.52	0.54	0.87	0.87	0.94	1.00	1.00	1.00	1.00	1.00
	0.0300	R	510	96	33	12	8	5	5	5	5	5
		ϵ_{norm}	0.74	0.54	0.45	0.77	0.87	1.00	1.00	1.00	1.00	1.00
	0.0600	R	678	149	58	27	11	7	5	5	5	5
		ϵ_{norm}	0.80	0.63	0.54	0.47	0.77	0.89	1.00	1.00	1.00	1.00
	0.1000	R	861	212	75	42	17	8	5	5	5	5
		ϵ_{norm}	0.81	0.67	0.58	0.46	0.63	0.87	1.00	1.00	1.00	1.00
	0.2000	R	1174	358	140	67	40	17	8	5	5	5
		ϵ_{norm}	0.80	0.72	0.59	0.58	0.51	0.64	0.88	1.00	1.00	1.00

Tabelle 9.38: Ergebnisse für das Test- und Bewertungsverfahren J_{RI} mit universellen und positiven und negativen Regeln für das Benchmarkproblem Box-Jenkins-Zeitreihe.

Boston Housing ($N = 13$, $D = 506$)

Für dieses Problem werden hier die besten Ergebnisse mit der inkrementellen Trefferquote J_{ITQ} erzielt. Wie die Ergebnisse zeigen, ist es dabei eindeutig von Vorteil, die Relevanzschwelle auf einen Wert $\tau \geq 0.5$ zu setzen. Es ist zu vermuten, dass andernfalls zu viele negative Regeln mit eher geringer Relevanz wirken (worauf die vergleichsweise große Anzahl an Regeln in diesen Fällen schließen lässt) und daher *zu restriktiv* gearbeitet wird. Im Experiment B1 werden die insgesamt besten Ergebnisse mit dem Mittelwertbasierten Verfahren J_{MVB} erzielt, wobei das beste Ergebnis mit $\kappa = 0.0100$ und $\tau = 0.5$ erzielt wird und bei $\epsilon_{norm} = 0.56$ mit $R = 179$ Regeln liegt. In diesem Experiment wird für die inkrementelle Trefferquote J_{ITQ} das beste Ergebnis mit $\kappa = 0.2000$ und $\tau = 0.6$ erzielt und liegt bei $\epsilon_{norm} = 0.54$ mit $R = 93$ Regeln. Die Fehlerwerte sind nahezu identisch, aber die Anzahl der Regeln halbiert sich fast.

J	κ		$\tau=0.0$	$\tau=0.1$	$\tau=0.2$	$\tau=0.3$	$\tau=0.4$	$\tau=0.5$	$\tau=0.6$	$\tau=0.7$	$\tau=0.8$	$\tau=0.9$
J_{ITQ}	0.0001	R	90	65	46	43	38	25	15	11	6	4
		ϵ_{norm}	0.86	0.80	0.81	0.76	0.67	0.62	0.64	0.71	0.92	0.99
	0.0010	R	144	94	71	64	48	28	16	9	5	4
		ϵ_{norm}	0.95	0.89	0.90	0.77	0.63	0.59	0.61	0.76	0.90	0.99
	0.0100	R	410	262	191	155	105	54	32	12	7	4
		ϵ_{norm}	1.21	1.12	1.14	0.99	0.73	0.59	0.58	0.74	0.92	0.99
	0.0300	R	688	427	290	210	146	74	45	15	8	5
		ϵ_{norm}	1.26	1.18	1.22	0.97	0.74	0.59	0.60	0.80	0.95	0.97
	0.0600	R	1048	649	410	285	201	122	62	31	8	5
		ϵ_{norm}	1.19	1.12	1.15	0.92	0.79	0.65	0.58	0.70	0.91	0.98
	0.1000	R	1440	854	525	345	238	135	65	33	11	6
		ϵ_{norm}	1.34	1.25	1.27	1.00	0.82	0.62	0.56	0.67	0.88	0.97
	0.2000	R	2388	1367	788	481	318	177	93	43	19	9
		ϵ_{norm}	1.44	1.39	1.29	1.03	0.85	0.67	0.54	0.66	0.87	0.96

Tabelle 9.39: Ergebnisse für das Test- und Bewertungsverfahren J_{ITQ} mit universellen und positiven und negativen Regeln für das Benchmarkproblem Boston Housing.

Fazit

Bei den Untersuchungen dieses Experiments steht die Frage im Vordergrund, ob sich durch die Regelgenerierung nach dem speziellen Konzept des Inkrementellen Relevanzindex, der auf eine möglichst gleichberechtigte Behandlung positiver und negativer Regeln abzielt, bessere Ergebnisse erzielen lassen als bei der *herkömmlichen* Generierung negativer Regeln (Experiment B3). Die Ergebnisse zeigen, dass die Regelgenerierung nach dem genannten Konzept in fast allen untersuchten Problemen tatsächlich bessere Resultate liefert und daher von Vorteil ist. Die Ergebnisse erhärten deshalb die Vermutung, dass die bisherigen Ansätze zur Regelgenerierung Schwachstellen im Hinblick auf die Erzeugung negativer Regeln aufweisen. Auffällig ist aber auch, dass oftmals solche Test- und Bewertungsverfahren die insgesamt besten Ergebnisse erzielen, die im Gegensatz zu der inkrementellen Trefferquote J_{ITQ} und der konfidenten inkrementellen Trefferquote J_{KITQ} nicht speziell im Hinblick auf diese Art der Regelgenerierung konzipiert wurden.

Ergebnisse Experiment B6

Mit Hilfe des Ansatzes für ein hybrides Fuzzy-System kann das Konzept des Inkrementellen Relevanzindex zur Regelgenerierung auch für TSK-Systeme angewandt werden.

Boston Housing ($N = 13$, $D = 506$)

Die Test- und Bewertungsverfahren J_{NHR} und J_{CHR} sind in diesem Beispiel (bzw. in den gesamten Experimenten B5 und B6) nicht untersucht worden, da sie in Verbindung mit dem Konzept des Inkrementellen Relevanzindex sehr häufig zu einer unverhältnismäßig großen Regelbasis führen und diese einen entsprechenden Zeitaufwand für die Auswertung benötigt. Die beiden Verfahren wurden stichprobenartig eingesetzt, lieferten aber – wie in den anderen Experimenten auch – nur vergleichsweise schlechte Resultate.

Die bei den Mamdani-Systemen (Experiment B5) recht deutlich ausgeprägten Unterschiede finden sich in den entsprechenden TSK-Systemen nicht mehr wieder. Diese Auffälligkeit zeigte sich bereits bei den Experimenten B1 und B2 bzw. B3 und B4.

J	κ	τ	R	ϵ	ϵ_{norm}	r	Modelltyp
J_{RI}	0.0100	0.2000	23	4.4564	0.6704	0.7991	MAM
	0.2000	0.3000	117	2.4536	0.3691	0.9173	TSK-LS-MV
	0.0600	0.1000	155	2.5209	0.3795	0.9002	TSK-RR-MV
J_{RHR}	0.0600	0.7000	210	4.2135	0.6345	0.7955	MAM
	0.1000	0.7000	331	2.4935	0.3754	0.9167	TSK-LS-MV
	0.2000	0.8000	302	2.4241	0.3648	0.9036	TSK-RR-MV
J_{MVB}	0.0001	0.3000	124	3.8872	0.5851	0.8206	MAM
	0.0600	0.4000	404	2.4688	0.3715	0.9142	TSK-LS-MV
	0.0100	0.2000	543	2.4126	0.3632	0.8967	TSK-RR-MV
J_{VVB}	0.2000	0.1000	1173	4.0269	0.6059	0.7889	MAM
	0.0600	0.1000	949	2.4958	0.3758	0.9076	TSK-LS-MV
	0.1000	0.1000	1022	2.6172	0.3941	0.8965	TSK-RR-MV
J_{CVB}	0.1000	0.1000	1378	4.1958	0.6313	0.7629	MAM
	0.0001	0.1000	447	2.6211	0.3943	0.9011	TSK-LS-MV
	0.0100	0.2000	623	2.4421	0.3677	0.9035	TSK-RR-MV
J_{ITQ}	0.2000	0.6000	93	3.5540	0.5353	0.8355	MAM
	0.1000	0.1000	835	2.5733	0.3872	0.9174	TSK-LS-MV
	0.1000	0.1000	862	2.4673	0.3713	0.9062	TSK-RR-MV
J_{KITQ}	0.2000	0.4000	20	4.1782	0.6288	0.7948	MAM
	0.2000	0.1000	296	2.7638	0.4158	0.9020	TSK-LS-MV
	0.0600	0.1000	190	2.5694	0.3869	0.8891	TSK-RR-MV

Tabelle 9.40: Jeweils beste Modellierungsergebnisse mit Mamdani- und TSK-Modellen mit universellen und positiven und negativen Regeln für das Benchmarkproblem Boston Housing.

Fazit

Wie auch bei den Experimenten B2 und B4 beobachtet, sind die Unterschiede der Test- und Bewertungsverfahren im Hinblick auf die erzielten Modellierungsfehler bei TSK-Systemen geringer als bei Mamdani-Systemen. Bezogen auf den Übergang von Mamdani-Regeln zu TSK-Regeln (hier vom Experiment B5 zu B6) zeigen sich die zu erwartenden Verbesserungen. Die teilweise erheblichen Verbesserungen, die sich für Mamdani-Systeme zwischen den Experimenten B3 und B5 zeigen, finden sich bei entsprechenden TSK-Systemen (Experimente B4 und B6) jedoch nicht in dieser Form wieder. Gleiches gilt für den Vergleich der TSK-Systeme der Experimente B2 und B6.

9.4 Zusammenfassende Darstellung und Bewertung der Modellierungsergebnisse

Es werden insgesamt zwölf bekannte Benchmarkprobleme (zum Beispiel Mackey-Glass Zeitreihe, Boston Housing Problem, etc.) untersucht, die in Anhang B genauer vorgestellt werden. Sie überdecken ein breites Spektrum, beispielsweise im Hinblick auf die Anzahl der zu beachtenden Eingangsgrößen. Zehn der untersuchten Benchmarkprobleme sind Approximationsprobleme, zwei sind Klassifikationsprobleme. Allen erzielten Ergebnissen liegt eine Kreuzvalidierung zu Grunde, die auf einer Aufteilung der verfügbaren Daten in zwei gleich große Datensätze basiert.

Ergebnisse Experiment A) Es werden Mamdani-Modelle mit nur positiven Regeln eingesetzt. Bei zwei der untersuchten Benchmarkprobleme liefert die automatische Partitionierung im Vergleich zu den hierfür jeweils besten Standard-Partitionierungen deutlich schlechtere Modellierungsergebnisse. In einem der Fälle werden offensichtlich nicht genügend linguistische Ausgangsgrößenwerte erzeugt, daher ist die resultierende Granularisierung zu grob. In dem zweiten Fall werden zwar ausreichend viele linguistische Ausgangsgrößenwerte erzeugt, es ergeben sich aber ungünstige Zugehörigkeitsfunktionen: Wertebereiche, in denen nur wenige Datenpunkte liegen, werden sehr fein, Wertebereiche mit sehr vielen Datenpunkten aber zu grob aufgelöst. Bei allen anderen untersuchten Benchmarkproblemen führt die automatische Partitionierung aber zu Modellierungsergebnissen, die, bezogen auf die Gesamtheit aller untersuchten Standard-Partitionierungen, immer mindestens im Mittelfeld, teilweise sogar in der Spitzengruppe liegen. Weiterhin hat sich gezeigt, dass keine der Standard-Partitionierungen immer zum besten Ergebnis führt. Zusammenfassend erweist sich die automatische Partitionierung basierend auf einem unidimensionalen Clusterverfahren als eine akzeptable Standard-Vorgehensweise.

Ergebnisse Experiment B1) Es werden Mamdani-Modelle mit nur positiven Regeln

eingesetzt. Die datenbasierte Generierung solcher Modellen ist die typische Domäne des Fuzzy-ROSA-Verfahrens. Dieses Experiment dient deshalb auch als Grundlage für die wissensbasierte bzw. datenbasierte Suche nach Einstellregeln für die Parametrierung. Ein Vergleich der hier erzielten Ergebnisse mit denen aus [Sla01] ist aufgrund der unterschiedlichen Partitionierungen und der unterschiedlichen Anzahl der untersuchten Einstelloptionen nur bedingt möglich, es zeigen sich aber häufig recht ähnliche Ergebnisse. Bei dem Vergleich mit den Ergebnissen aus [Sla01] ist ferner zu beachten, dass dort die wesentlich kleinere Anzahl untersuchter Fälle eine nachträgliche Verbesserung der Modellgenauigkeit durch eine nachgeschaltete Optimierung der Regelbasis erlaubte. Die insgesamt besten Ergebnisse in [Sla01] werden typischerweise mit einer in der Zusammensetzung optimierten Regelbasis erzielt.

Es zeigt sich in den Ergebnissen, dass kein Test- und Bewertungsverfahren stets das beste Resultat liefert. Von den sieben untersuchten Test- und Bewertungsverfahren schneiden der Relevanzindex J_{RI} oder der Mittelwertbasierte Regeltest J_{MVB} meist am besten ab. In einigen Fällen sind aber auch der Varianzbasierte Regeltest J_{VVB} oder der Medianbasierte Regeltest J_{CVB} vorzuziehen. Insgesamt zeichnen sich Vorteile für diejenigen Test- und Bewertungsverfahren ab, die auf Datenverteilungen arbeiten. Dies war auch zu erwarten, da hier überwiegend Approximationsprobleme untersucht werden, für die diese Test- und Bewertungsverfahren konzipiert wurden. Zur Erzielung einer möglichst hohen Approximationsgüte erweist es sich als günstig, für die Parameter κ und τ ein breiteres Spektrum an Einstellvarianten zu untersuchen. Die jeweils besten Ergebnisse eines Test- und Bewertungsverfahrens für ein Benchmarkproblem werden in der Regel mit einer solchen Einstellung von κ und τ erreicht, die von der jeweils vorgeschlagenen Default-Einstellung abweicht. Für viele Benchmarkprobleme zeigt sich dabei ein gemeinsamer Trend: Die besten Ergebnisse werden oftmals dann erzielt, wenn die Schwelle zum Bestehen des Regeltests *eher etwas leichter* und die Relevanzschwelle für die Regelübernahme *eher etwas höher* gewählt wird. Dieser Trend zeigt sich insbesondere bei den Test- und Bewertungsverfahren, die auf Datenverteilungen basieren.

Ergebnisse Experiment B2) Es werden TSK-Modelle mit nur positiven Regeln eingesetzt. Diese Modelle gehen durch Transformation aus den Modellen des Experiments B1 hervor. Hier ist von Interesse, wie groß der Zugewinn an Modellgenauigkeit durch den Übergang von Mamdani-Modellen zu TSK-Modellen ausfällt. Hierfür werden ausschließlich Approximationsprobleme untersucht, denn bei Klassifikationsproblemen bieten TSK-Modelle keine erkennbaren Vorteile gegenüber den bereits hierfür untersuchten Mamdani-Modellen. Die hier erzielten Modellierungsergebnisse ähneln stark denen in [Kra02], werden nun aber anhand einer viel größeren Anzahl von Benchmarkproblemen abgesichert.

Für die hier untersuchten Benchmarkprobleme lässt sich zusammenfassend feststellen, dass die mit der Transformation in TSK-Modelle verbundene Erhöhung der Mo-

dellgenauigkeit stets signifikant ausfällt. Die erzielten Verbesserungen⁸ gegenüber den korrespondierenden Mamdani-Modellen bewegen sich dabei in einem weiten Spektrum und liegen meist zwischen 30% und 50%, in Einzelfällen teilweise bei sogar über 80%. Für die Transformation von Mamdani-Modellen in TSK-Modelle wird als neues Strategieelement die Bestimmung der freien Parameter in der Konklusion einer TSK-Regel, basierend auf einem Iteratively Reweighted Least-Squares Ansatz (IRLS), eingesetzt. Eine weitere neue Option ist die Berechnung des Ausgabewertes eines TSK-Modells über den gewichteten Median aller Einzelempfehlungen. Damit stehen neue Alternativen zu den bereits vorhandenen Ansätzen (herkömmlicher Least-Squares Ansatz zur Bestimmung der Parameter in der Konklusion einer Regel und Verwendung des gewichteten Mittelwertes aller Einzelempfehlungen als Ausgabewert) zur Verfügung. Durch Kombination der verschiedenen Optionen für die Parameterbestimmung und die Ausgabeberechnung ergeben sich insgesamt vier TSK-Modellvarianten, die hier sämtlich untersucht werden. Dabei hat sich keine der Varianten als stets beste Lösung erwiesen. Es zeichnet sich aber ab, dass eine Kombination aus IRLS-Ansatz für die Parameterbestimmung und die Verwendung des gewichteten Median aller Einzelempfehlungen als Ausgabewert eine empfehlenswerte Standard-Einstellung ist. Die neuen Strategieelemente, die auf eine Verbesserung der Robustheit der TSK-Modelle gegenüber fragwürdigen oder fehlerhaften Daten abzielen, haben sich somit als erfolgreich erwiesen.

Ergebnisse Experiment B3) Es werden Mamdani-Modelle mit positiven und negativen Regeln eingesetzt. In früheren Arbeiten wurde nachgewiesen, dass negative Regeln in Mamdani-Systemen eine transparente Berücksichtigung von Expertenwissen erlauben, das in Form von Warnungen oder Verboten vorliegt. Ebenso wurde am Beispiel der datenbasierten Modellierung des Verhaltens eines menschlichen Prozessbedieners gezeigt, dass die Berücksichtigung von negativen Regeln verglichen mit dem Fall nur positiver Regeln einen deutlichen Zugewinn an Modellgenauigkeit bringen kann [Kie97a].

Im Rahmen dieser Arbeit wird eine flächendeckende Untersuchung des Potenzials negativer Regeln anhand von Benchmarkproblemen durchgeführt, und zwar bezüglich aller Test- und Bewertungsverfahren (J_{RI} , J_{RHR} , J_{NHR} und J_{CHR}), die die Erzeugung negativer Regeln unterstützen. Verglichen mit der Verwendung nur positiver Regeln zeigt sich, dass sich die resultierende Modellgenauigkeit für alle genannten Test- und Bewertungsverfahren oftmals tatsächlich verbessert, aber nur selten deutlich. Zudem ergeben sich so oftmals (teilweise deutlich) größere Regelbasen mit überwiegend negativen Regeln. Ferner ist anzumerken, dass das mit den genannten Test- und Bewertungsverfahren beste erzielte Modellierungsergebnis nicht selten noch durch das beste Modellierungsergebnis übertroffen wird, das mit den Test- und Bewertungsverfahren (J_{MVB} , J_{VVB} und J_{CVB}), die keine negativen Regeln generieren, erzielt wird.

Zusammenfassend ist festzustellen, dass die Einbeziehung negativer Regeln die Mo-

⁸Diese Aussage bezieht sich auf die jeweilige Verringerung des Modellierungsfehlers.

dellgenauigkeit verbessern kann, dies aber nicht zwingend muss. Liegt die hier behandelte Aufgabe vor, ein möglichst genaues Modell allein aus vorhandenen Daten zu generieren, so bedeutet der Verzicht auf negative Regeln nicht zwangsläufig einen Verlust an erreichbarer Modellgenauigkeit. Besteht aber ein Interesse daran, auch qualitatives Anwenderwissen für die Erstellung eines möglichst transparenten Modells zu nutzen, beispielsweise zur Gewährleistung sicherheitsrelevanter Aspekte, oder menschliche Bedienstrategien verstehen und darstellen zu können, dann ist die Verwendung negativer Regeln vorteilhaft.

Ergebnisse Experiment B4) Aufsetzend auf den in Experiment B3 generierten Regeln werden die daraus hervorgehenden hybriden Fuzzy-Modelle untersucht.

Im Vergleich zu den TSK-Modellen mit nur positiven Regeln (Experiment B2) werden durch die Einbeziehung negativer Regeln ähnliche Verbesserungen erzielt wie bei der Einbeziehung negativer Regeln in Mamdani-Systemen. Daher gelten hinsichtlich des Stellenwertes negativer Regeln die Schlussfolgerungen des Experiments B3.

Ergebnisse Experiment B5) Es werden Mamdani-Modelle mit universellen und positiven und negativen Regeln eingesetzt. Die Regeln werden nach dem Konzept des Inkrementellen Relevanzindex von KIENDL generiert. Für die Generierung der inkrementell wirkenden Regeln sind zwei spezielle Test- und Bewertungsverfahren, die inkrementelle Trefferquote J_{ITQ} und die konfidente inkrementelle Trefferquote J_{KITQ} , konzipiert worden, die hier zusätzlich zu den sieben anderen Test- und Bewertungsverfahren eingesetzt werden. Die Anzahl der zu untersuchenden Parametrierungen erhöht sich dadurch von 490 auf 630.

Im Vergleich zu den Ergebnissen des Experiments B1 wird auf diese Weise für eine Reihe von Benchmarkproblemen eine nennenswerte Verbesserung der Modellgenauigkeit, meist bei einer vergrößerten Regelbasis, erzielt. Diese Verbesserung fällt deutlicher aus als bei dem Experiment B3. Daher bestätigt sich, dass das Konzept der Verwendung universeller Regeln in Verbindung mit inkrementell wirkenden Regeln im Hinblick auf die angestrebte Gleichbehandlung positiver und negativer Regeln günstig ist. Allerdings liefern im Hinblick auf die Modellgenauigkeit nicht in allen Fällen die für die inkrementell wirkenden Regeln eigens konzipierten Test- und Bewertungsverfahren J_{ITQ} und J_{KITQ} die besten Resultate, sondern in Einzelfällen andere Verfahren, beispielsweise der Relevanzindex J_{RI} .

Ergebnisse Experiment B6) Aufsetzend auf den in Experiment B5 generierten Regeln werden die daraus hervorgehenden hybriden Fuzzy-Systeme untersucht.

Im Gegensatz zu den deutlichen Unterschieden, die sich zwischen den Mamdani-Systemen (Experimente B3 und B5) zeigen, fallen die Unterschiede in den Ergebnissen der Experimente B4 und B6 nur gering aus. Daher gelten hinsichtlich des Stellenwertes negativer Regeln die Schlussfolgerungen des Experiments B4.

9.5 Qualitative Empfehlungen für die Einstellung freier Modellparameter

Aufbauend auf den durch die systematische Exploration gewonnenen Daten wird nach Zusammenhängen zwischen Kenngrößen eines Modellierungsproblems und den für unterschiedliche Parametrierungen resultierenden Modellierungsfehlern gesucht. Hierfür wird der zu einem Modellierungsproblem erhobene Datensatz durch Kenngrößen charakterisiert, die sich für unterschiedliche Modellierungsprobleme stets in gleicher Weise berechnen und auswerten lassen.

Die einer datenbasierten Modellierung zu Grunde liegenden Daten unterscheiden sich typischerweise stark von Problem zu Problem, beispielsweise in der Anzahl der Eingangsgrößen. Außerdem sind die Wertebereiche der einzelnen Größen einerseits von Problem zu Problem verschieden, meist aber auch innerhalb des Datensatzes eines Problems selbst. Um eine überschaubare Anzahl aussagekräftiger Kenngrößen zu erhalten, die zudem eine Vergleichbarkeit verschiedener Probleme ermöglichen, wird bei der Berechnung der Kenngrößen wie folgt vorgegangen: Die Eingangsgrößen und die Ausgangsgröße eines Datensatzes werden zunächst jeweils individuell auf den Wertebereich $[0, 1]$ normiert. Ein normierter Wert von 0 entspricht dem kleinsten unnormierten Wert und ein normierter Wert von 1 entsprechend dem größten unnormierten Wert einer Größe. Nach der Normierung werden für jede Eingangsgröße verschiedene Kenngrößen, wie beispielsweise die Standardabweichung oder die Korrelation mit der Ausgangsgröße, berechnet. Da die Anzahl der Eingangsgrößen von Problem zu Problem variieren kann, gilt dies folglich auch für die Anzahl der hierzu insgesamt gebildeten Kenngrößen. Die Abhängigkeit der Anzahl der Kenngrößen von der Anzahl der Eingangsgrößen wird hier für den Entwurf des angestrebten Prognosemoduls als ungünstig angesehen. Deshalb werden die Kenngrößen geeignet aggregiert, beispielsweise werden die Standardabweichungen der normierten Eingangsgrößen zu einer mittleren Standardabweichung zusammengefasst. Damit einzelne Eingangsgrößen das Gesamtbild eines Datensatzes nicht allzu stark verzerren können, wird für die Aggregation nicht der Mittelwert, sondern der Median verwendet. In dieser Weise werden auch für die Ausgangsgröße entsprechende Kenngrößen auf Grundlage der normierten Daten berechnet. Diese Vorgehensweise liefert für jedes Modellierungsproblem stets die gleiche Anzahl von Kenngrößen, unabhängig von der Anzahl der Eingangsgrößen. Zudem bewegen sich die so berechneten Kenngrößen stets innerhalb bekannter Wertebereiche, da sie anhand normierter Daten berechnet werden.

Zur Bewertung der Modellierungsergebnisse werden hier in erster Linie normierte Fehlerwerte ϵ_{norm} herangezogen, die in Anhang C erläutert sind. Dabei wird beispielsweise der mittlere absolute Fehler eines Fuzzy-Modells in das Verhältnis zu dem Fehler eines so genannten Trivialmodells gesetzt, das unabhängig von dem vorliegenden Eingangsgrößenvektor \mathbf{x} stets den gleichen Ausgabewert \hat{y}_{def} liefert. Der Ausgabewert eines solchen Trivialmodells wird oft auch als Baseline-Prognose bezeichnet. Gilt $\epsilon_{norm} = 1$, so arbeitet das Fuzzy-Modell nicht besser als dieses Trivialmodell und bringt daher offensichtlich kei-

nen Zugewinn an Modellgenauigkeit. Der betriebene Mehraufwand im Vergleich zu dem Trivialmodell erscheint daher nicht gerechtfertigt. Erst bei $\epsilon_{norm} \ll 1$ lohnt der mit einem Fuzzy-Modell verbundene Mehraufwand.

Die berechneten Kenngrößen und erzielten Modellierungsergebnisse werden für jedes untersuchte Problem in eine Matrix eingetragen. Die einzelnen Matrizen zu den untersuchten Problemen werden in einer Übermatrix zusammengefasst, die alle Daten zu einem bestimmten Experiment enthält. Die Analyse einer solchen Übermatrix zielt nicht auf die Untersuchung eines speziellen Problems ab, sondern vielmehr darauf, allgemeine Zusammenhänge aufzudecken, die sich für unterschiedliche Probleme in ähnlicher Form abzeichnen. Eine solche Übermatrix ist daher für die Suche nach den fraglichen Zusammenhängen geeignet. Die Suche nach Zusammenhängen kann wissensbasiert erfolgen, beispielsweise durch eine geeignete Visualisierung der Daten und darauf aufbauenden Schlussfolgerungen, und führt zu qualitativen Regeln, die die Zusammenhänge beschreiben. Damit wird das einleitend beschriebene Zwischenziel eingelöst.

Die wissensbasierte Formulierung von Einstellempfehlungen gestaltet sich aufgrund der großen Anzahl potenziell relevanter Kenngrößen schwierig. Um überhaupt eine Chance hierfür zu haben, erscheint es angebracht, diese auf ihre Relevanz hin zu untersuchen und die Anzahl der zu beachtenden Kenngrößen gegebenenfalls zu reduzieren. Hierfür wird das in Abschnitt 5.3 beschriebene evolutionäre Verfahren für die Eingangsgrößenselektion eingesetzt. Das Verfahren stuft dabei zwei Kenngrößen eines Problems als besonders relevant ein, zum einen die gemittelte Korrelation zwischen einer Eingangsgröße und der Ausgangsgröße, zum anderen die gemittelte Standardabweichung der Eingangsgrößen. Das erste Merkmal kann als ein Indikator dafür aufgefasst werden, wie eindeutig sich Zusammenhänge zwischen Eingangsgröße(n) und Ausgangsgröße in den Daten abzeichnen und wie schwierig sich eine Modellierung wohl grundsätzlich gestalten wird. Das zweite Merkmal ist ein Maß für die Streuung. Bei einer starken Streuung sind die Daten ungleichmäßig verteilt und es ist zu vermuten, dass es sich in solchen Fällen schwieriger gestaltet, datenbasiert Zusammenhänge herzustellen. Die von dem Verfahren auf Grundlage von Daten als relevant eingestuft Kenngrößen eines Problems erscheinen somit auch bei einer wissensbasierten Betrachtung durchaus plausibel. Die untersuchten Benchmarkprobleme wurden unter anderem anhand der beiden genannten Kenngrößen in einer Art Landkarte visualisiert. In weitergehenden Analysen zeigt sich, dass für unterschiedliche Benchmarkprobleme, die sich im Hinblick auf diese Kenngrößen auf dieser Landkarte zusammen gruppieren, meist ein und dasselbe Test- und Bewertungsverfahren besonders gut geeignet ist. Offenbar sind diese Kenngrößen für eine problemabhängige Wahl eines Test- und Bewertungsverfahrens relevant. In Verbindung mit den im Experiment B1 aufgedeckten Tendenzen und den daraus abgeleiteten Einstellempfehlungen für die Parameter κ und τ werden auf diese Weise qualitative Regeln für die Wahl und die Parametrierung eines günstigen Test- und Bewertungsverfahrens formuliert.

9.6 Modul zur Prognose des Modellierungsfehlers

Die Suche nach Einstellregeln für freie Modellparameter kann auch datenbasiert erfolgen, beispielsweise mit dem Fuzzy-ROSA-Verfahren selbst. Das Erstellen eines entsprechenden Moduls, das die Zusammenhänge zwischen den Kenngrößen eines Modellierungsproblems, den Einstellungen freier Modellparameter und den erzielten Ergebnissen widerspiegelt, ist das hier verfolgte weiterführende Ziel.

Für ein gegebenes neues Modellierungsproblem soll eine möglichst günstige Einstellung freier Modellparameter vorgenommen werden, um so ein möglichst genaues und korrektes Modell zu erstellen. Als freie Modellparameter werden hier folgende Einstelloptionen festgelegt: die Wahl eines Test- und Bewertungsverfahrens und die Einstellung der zugehörigen Parameter κ und τ . Grundsätzlich kommen zwei unterschiedliche Ansätze für Typen von Modulen für diese Aufgabe in Frage, deren Generierung auf der Grundlage der im Zuge der systematischen Exploration der Einstellmöglichkeiten gewonnenen Daten möglich erscheint. Der erste Ansatz besteht darin, charakteristische Kenngrößen des Modellierungsproblems zu bestimmen und einem Modul zuzuführen, das Empfehlungen für die Einstellung der oben genannten freien Modellparameter als Ausgabe liefert. Ein solches Modul hat somit drei Ausgabewerte. Der andere Ansatz besteht darin, ein Modul zu entwerfen, das die charakteristischen Kenngrößen des Modellierungsproblems *und* die Einstellung der freien Modellparameter als Eingangsgrößen verwendet und als Ausgabe eine Prognose des in diesem Fall zu erwartenden Modellierungsfehlers liefert. Ein solches Modul hat somit nur einen Ausgabewert.

Der erste Ansatz ist aus unterschiedlichen Gründen eher ungeeignet. Zum einen stellt sich die Frage, wie die Regeln in diesem Fall auf Grundlage der zuvor gewonnenen Daten erlernt werden können. Man müsste hierbei wohl so vorgehen, dass man zunächst für jedes untersuchte Modellierungsproblem diejenigen Modelle bzw. Einstelloptionen bestimmt, die die besten Modellierungsergebnisse erzielen. In den so ausgedünnten Daten würde man dann die Kenngrößen der Modellierungsprobleme als Eingangsgrößen und die Einstelloptionen der Modelle als Ausgangsgrößen ansehen und nach entsprechenden Zusammenhängen zwischen den Eingangs- und den Ausgangsgrößen suchen. Würde man die Daten nicht reduzieren, könnten keine Zusammenhänge aufgedeckt werden, da in den vollständigen Daten alle Einstelloptionen gleich häufig auftreten und sich daher keine Regeln abzeichnen können. Ein Nachteil dieses Ansatzes liegt zudem darin, dass eine modellgestützte Optimierung der Einstellparameter so nicht möglich ist. Für einen gegebenen Satz von Kenngrößen eines Modellierungsproblems liefert ein solches Modul genau einen bestimmten Satz von Empfehlungen für die Einstelloptionen. Aufgrund der beschriebenen Schwierigkeiten wird dieser Ansatz im Rahmen dieser Arbeit nicht verfolgt.

Als weiterführendes Ziel wird hier die datenbasierte Generierung eines Moduls zur Prognose des Modellierungsfehlers angestrebt. Die Kenngrößen eines gegebenen Problems sind die Eingangsgrößen des Moduls. Im praktischen Einsatz werden die Kenngrößen des Modellie-

rungsproblems vorab berechnet und nachfolgend als Eingangsgrößen des Moduls konstantgehalten. Die Einstelloptionen – das angewandte Test- und Bewertungsverfahren sowie die Werte für dessen Parameter κ und τ – sind ebenfalls Eingangsgrößen des Moduls. Sie werden variiert und simulieren so die unterschiedlichen Parametrierungen und die zugehörigen Modelle. Jedem simulierten Modell wird entsprechend der jeweils prognostizierte Modellierungsfehler zugewiesen. Aus der Menge aller simulierten Modelle wird dann das Modell mit dem kleinsten zu erwartenden Modellierungsfehler ausgewählt und tatsächlich generiert. Anhand des prognostizierten und des tatsächlichen Modellierungsfehlers kann man beurteilen, ob das Prognosemodul hinreichend exakt arbeitet und ob diesem vertraut werden kann. Als ein solches Modul kann grundsätzlich bereits das zellenbasierte Modell eingesetzt werden, das im Zuge der Relevanzbewertung von Kenngrößen erstellt wird (siehe Abschnitt 9.4). Aufgrund seines Prognosemechanismus (siehe Abschnitt 5.3) liefert es jedoch vergleichsweise ungenaue Prognosen. Es zeigt sich aber, dass ein akzeptabel genaues Prognosemodul mit dem Fuzzy-ROSA-Verfahren erzeugt werden kann. Zunächst wird versucht, *alle* in der Übermatrix⁹ gesammelten Daten für das Erlernen der Regeln zu verwenden. Dass dies nicht den gewünschten Erfolg bringt, lässt sich unter anderem dadurch erklären, dass innerhalb der Übermatrix Mehrdeutigkeiten und Widersprüche auftreten, die die Regelgenerierung erschweren. Zur Überwindung dieses Problems wird der für das Erlernen verwendete Datensatz ausgedünnt. Zu jedem der im Rahmen der systematischen Exploration der Einstelloptionen untersuchten Benchmarkproblem werden 490 Modelle generiert, davon werden aber jeweils nur die 25 besten betrachtet. Dies stellt sicher, dass alle Benchmarkprobleme in den Lerndaten berücksichtigt werden und damit der Raum der Problemkenngrößen gut abgedeckt wird. Ferner gewährleistet dies, dass ein breites Spektrum von Modellgüten abgedeckt wird, denn das beste für ein Benchmarkproblem erhaltene Modell weist nicht notwendigerweise auch eine hohe Modellierungsgüte auf. Auf Grundlage der so zusammengestellten Daten werden ein Mamdani-Modell und ein TSK-Modell generiert. Die Genauigkeit des Mamdani-Modells ist noch nicht ausreichend, die des TSK-Modells ist aber akzeptabel. Dieses TSK-Modell ist ein Prototyp des angestrebten Moduls zur Prognose des zu erwartenden Modellierungsfehlers. Seine Validierung an einem neuen Benchmarkproblem¹⁰, dessen Kenngrößen nicht Bestandteil des Lerndatensatzes sind, ergibt als Differenz zwischen dem tatsächlichen und dem prognostizierten Modellierungsfehler in einem Großteil der 490 Fälle (Standard-Parametrierungen) einen Wert kleiner 10%, in einigen Fällen zeigen sich jedoch größere Abweichungen. Dies sind überwiegend die Fälle, in denen ein realer normierter Modellierungsfehler $\epsilon_{norm} > 1$ erzielt wird. Die Prognosen sind für diese Fälle sehr ungenau, daher werden zunächst nur die Standard-Parametrierungen bestimmt, die für dieses Modellierungsproblem zu einem realen normierten Modellierungsfehler $\epsilon_{norm} < 1$ führen. Dies sind ca. 380 der 490 Standard-Parametrierungen. In Abbildung 9.9 sind hierfür die prozentualen Abweichungen in Form eines Histogramms dargestellt. In Abbildung

⁹Die Ausführungen beziehen sich hier auf die Ergebnisse des Experiments B1.

¹⁰Es handelt sich hierbei um das Benchmarkproblem *Cars* mit sieben Eingangsgrößen, bei dem der durchschnittliche Treibstoffverbrauch (miles per gallon) unterschiedlicher Autotypen aus den Jahren 1970 bis 1982 modelliert werden soll (siehe auch <http://www.ics.uci.edu/~mlearn/MLSummary.html> und <http://lib.stat.cmu.edu/datasets/>).

9.10 sind die Abweichungen noch einmal in anderer Form dargestellt. Die realen normierten Modellierungsfehler werden in aufsteigender Reihenfolge sortiert und es werden hierzu der $\pm 10\%$ - und der $\pm 20\%$ -Toleranzbereich dargestellt. Die Darstellung der zugehörigen prognostizierten normierten Modellierungsfehler zeigt, wie auch in Abbildung 9.9 zu erkennen, dass sich der Großteil der prognostizierten Fehlerwerte innerhalb dieser Bereiche bewegt. Auffällig ist, dass für kleinere reale Modellierungsfehler die Prognosen tendenziell eher zu hoch und für größere reale Modellierungsfehler eher zu niedrig ausfallen. Weiterhin zeigt sich, dass in diesem Beispiel die Parametrierung, die von dem Prognosemodul als am geeignetsten eingestuft wird, also den kleinsten prognostizierten Modellierungsfehler erzielt, nicht den tatsächlich kleinsten realen Modellierungsfehler liefert.

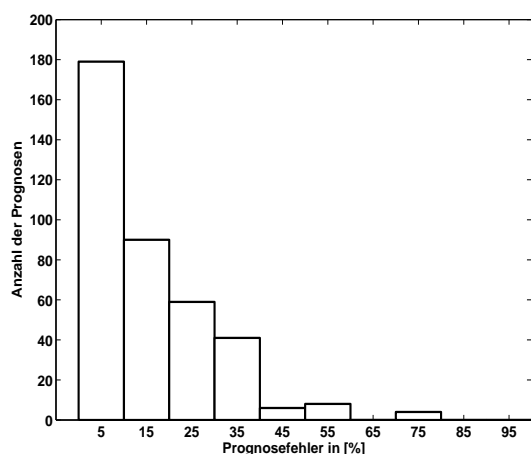


Abbildung 9.9: Histogramm der Prognosefehler für das Modellierungsproblem *Cars*.

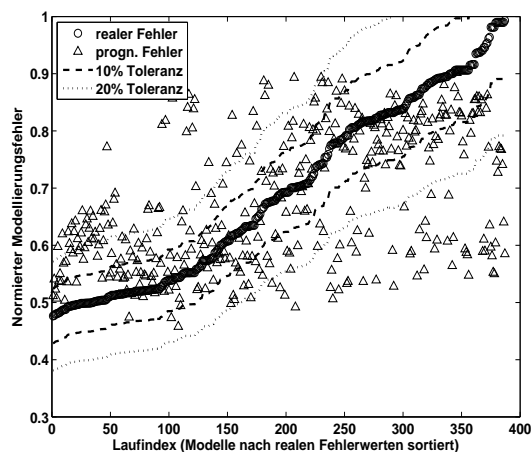


Abbildung 9.10: Reale und prognostizierte Modellierungsfehler für das Modellierungsproblem *Cars*.

Es stellt sich die Frage, welche realistischen Erwartungen an ein solches Prognosemodul grundsätzlich gestellt werden können. Es erscheint eher unrealistisch, exakte Prognosen erzielen zu können. Jedes neue Modellierungsproblem unterscheidet sich von den bisher untersuchten Modellierungsproblemen in irgendeiner Art und Weise soweit, dass das Zusammenspiel aus automatischer Suchraumstrukturierung und Regelgenerierung zu einem Ergebnis führt, dass sich nicht genau vorhersagen lässt. Für das hier betrachtete Modellierungsproblem ist eine Verbesserung des Prognosemoduls aber dahingehend erstrebenswert, dass die Bandbreite der Abweichungen zwischen realen und prognostizierten Modellierungsfehlern geringer wird. Es erscheint anhand der bereits erzielten Ergebnisse durchaus möglich, das Prognosemodul soweit zu verbessern, dass keine prozentualen Abweichungen auftreten, die größer als $\pm 20\%$ sind. Es wird nie gänzlich auszuschließen sein, dass ein solches Prognosemodul nicht die tatsächliche optimale, sondern eine suboptimale Parametrierung vorschlägt. Bei einer Prognosegenauigkeit, die sich in der skizzierten Bandbreite bewegt, dürfte dies aber für industrielle Anwender oftmals durchaus ausreichend sein. Ins-

besondere auch dann, wenn man den unter Umständen nicht unerheblichen Zeitgewinn gegenüber einer manuellen oder batchartigen Modellgenerierung bedenkt.

Zusammenfassend lässt sich feststellen, dass es tatsächlich möglich erscheint, sowohl wissensbasiert als auch datenbasiert Regeln für die Einstellung freier Modellparameter des Fuzzy-ROSA-Verfahrens aufzustellen. Der hier verfolgte Ansatz zielt auch darauf ab, die Ergebnisse, die bei der Erprobung des datenbasiert generierten Moduls gewonnen werden, für dessen Weiterentwicklung zu nutzen. Die Kenngrößen eines untersuchten Problems und die Modellierungsfehler der hierzu real erzeugten Modelle werden dabei der bereits vorhandenen Datenbasis hinzugefügt und das Prognosemodul auf der so erweiterten Grundlage neu erstellt.

Kapitel 10

Zusammenfassung und Ausblick

In den Natur- und Ingenieurwissenschaften ist die Verfügbarkeit von Modellen von großer Bedeutung. Modelle dienen zum Entwurf, zur Simulation, Analyse und gezielten Beeinflussung komplexer technischer Systeme und Prozesse. Modelle besitzen aber auch in anderen Bereichen, beispielsweise den Wirtschafts- und Sozialwissenschaften, einen hohen Stellenwert, da sie zum Verständnis von Wirkungszusammenhängen beitragen, die sich nicht mathematisch exakt formulieren lassen. Aufgrund der unterschiedlichen Zielsetzungen steht im Bereich der Natur- und Ingenieurwissenschaften eher die Modellgenauigkeit und im Bereich der Wirtschafts- und Sozialwissenschaften eher die Interpretierbarkeit der Modelle im Vordergrund. Grundsätzlich sind aber beide Eigenschaften wünschenswert und sollten daher möglichst auch parallel angestrebt werden.

Aufgrund der Komplexität moderner Systeme und Prozesse ist eine wissensbasierte Modellierung mit klassischen Methoden häufig sehr zeit- und kostenaufwendig, teilweise aber auch gänzlich unmöglich. Dieses Problem stellt sich gerade auch bei vielen nicht-technischen Systemen, da dort die zu Grunde liegenden Wirkungsmechanismen häufig nur sehr vage bekannt sind oder nur vermutet werden können. Daher nimmt die Bedeutung von Verfahren zu, die Modelle auf der Basis vorhandener Daten erzeugen, die zum Beispiel durch Experimente oder durch Befragungen erhoben worden sind. Dabei geht man davon aus, dass sich die vorhandenen Wirkungsmechanismen des zu modellierenden Systems, die dem Menschen jedoch unbekannt sind, in irgendeiner Form in den protokollierten Daten widerspiegeln. Die Aufdeckung solcher Zusammenhänge in Daten wird auch oft allgemein unter dem Begriff des Data Mining zusammengefasst. Für die Lösung dieser Aufgabe kommen häufig auf Methoden der Computational Intelligence (CI) basierende Verfahren zum Einsatz. Unter dieses gemeinsame Dach fallen viele unterschiedliche Ansätze, insbesondere auch Fuzzy-Logik bzw. Fuzzy-Control, künstliche neuronale Netze und evolutionäre Algorithmen¹. Diese Ansätze werden häufig auch als Problemlösung durch Imitation der

¹Unter diesem Begriff werden hier sowohl Evolutionsstrategien als auch genetische Algorithmen und andere thematisch verwandte Ansätze verstanden.

Natur bezeichnet, da sie Strategien, mit deren Hilfe die Natur hervorragende und hochangepasste Lösungen hervorgebracht hat, für technische Problemstellungen imitieren und adaptieren. In der Kombination der genannten Wissenschaftszweige liegt somit ein großes Synergiepotenzial zur Lösung komplexer Probleme unterschiedlichster Art.

Unter den etablierten Ansätzen aus dem Bereich der CI zeichnet sich die datenbasierte Fuzzy-Modellierung dadurch aus, dass sie zum einen transparente und genaue Modelle generieren kann, auch wenn nur sehr wenig Wissen über das zu modellierende System vorliegt. Zum anderen ermöglicht sie es aber auch, bereits vorhandenes Wissen in das Modell zu integrieren. Damit ergeben sich Vorteile gegenüber anderen Ansätzen, wie beispielsweise künstlichen neuronalen Netzen, die diese Möglichkeit nicht bzw. nicht in diesem Umfang bieten. Diese Vorzüge motivierten die Entwicklung des Fuzzy-ROSA-Verfahrens, das Ausgangspunkt dieser Arbeit ist. Grundidee des Verfahrens ist es, die Suche nach einer geeigneten Regelbasis auf die Suche nach einzelnen geeigneten Regeln zu reduzieren. Diese Vorgehensweise hat zwei wesentliche Vorteile: Zum einen ist der Suchraum meist wesentlich kleiner als bei der Suche nach einer kompletten Regelbasis. Zum anderen ist gewährleistet, dass nur lokal vernünftige Regeln erzeugt werden, die zuvor einen entsprechenden Test bestanden haben. In Vorarbeiten wurden unterschiedlichste Strategieelemente für das Verfahren entwickelt und dessen Leistungsfähigkeit an vielen praktischen Anwendungen und Benchmarkproblemen gezeigt.

Übergeordnetes Ziel dieser Arbeit ist es, die datenbasierte Modellierung mit dem Fuzzy-ROSA-Verfahren im Hinblick auf die Anforderungen in industriellen Problemstellungen zu erweitern und Anwendern systematische Lösungsstrategien zur Verfügung zu stellen. Aus Anwendersicht ist entscheidend, dass die eingesetzten Modellierungsverfahren einerseits hohe Anforderungen an Güte und Interpretierbarkeit der erzeugten Modelle erfüllen, andererseits aber auch ohne großen Einarbeitungsaufwand eingesetzt werden können. Weiterhin sollen die Verfahren für hochdimensionale Problemstellungen und auch bei nur kleinen Datenbasen anwendbar und robust gegenüber potenziell fehlerhaften Messdaten sein.

In dieser Arbeit wird ein Medianbasiertes Test- und Bewertungsverfahren vorgestellt, das auf die Generierung relevanter Mamdani-Regeln im Fall von nur wenigen und/oder gestörten Daten abzielt. Verglichen mit Verfahren, die auf Grundlage von Datenhäufigkeiten arbeiten, sind Verfahren, die auf Grundlage von Datenverteilungen arbeiten, besser in der Lage, schwach ausgeprägte Zusammenhänge in den Daten aufzudecken. Entsprechende bereits vorhandene Test- und Bewertungsverfahren (Mittelwertbasiertes Verfahren und Varianzbasiertes Verfahren) sind jedoch vergleichsweise anfällig gegenüber Störungen in den Daten (Ausreißer). Ferner gehen die dort angewandten statistischen Testverfahren von bestimmte Annahmen (Normalverteilung und Varianzhomogenität) über die Art der Datenverteilung aus, die insbesondere im Fall von nur wenigen Datenpunkten kritisch zu sehen sind. Im Gegensatz dazu ist der Median deutlich robuster gegenüber potenziell fehlerhaften Daten, und der eingesetzte statistische Test macht keine speziellen Annahmen bezüglich der Datenverteilung.

Das Fuzzy-ROSA-Verfahren erzeugt Mamdani-Modelle, die sich durch ihre gute Interpretierbarkeit auszeichnen. Im Hinblick auf die erreichbare Modellgenauigkeit sind jedoch Takagi-Sugeno-Kang-Modelle meist deutlich besser geeignet. Um die Vorzüge beider Modellarten miteinander zu verbinden, ist in Vorarbeiten die Möglichkeit geschaffen worden, ein Mamdani-Modell unter Erhaltung der Interpretierbarkeit in ein korrespondierendes TSK-Modell zu transformieren. Kernaufgabe bei der Transformation ist es, ein lineares Gleichungssystem zu lösen, wobei in den allermeisten Fällen jedoch keine eindeutige bzw. keine exakte Lösung existiert. Im Fall eines unterbestimmten linearen Gleichungssystems wurde bisher auf eine einfach zu berechnende Ersatzlösung zurückgegriffen, bei der der mittlere Ausgangsgrößenwert der regelstützenden Datenpunkte als konstanter Term in der Konklusion der TSK-Regel gesetzt wird und alle übrigen Parameter zu Null gewählt werden. Diese Vorgehensweise setzt allerdings einen bestimmten Ansatz für die Konklusion voraus (siehe Abschnitt 7.2.1) und liefert eine vergleichsweise ungenaue Approximation der zu Grunde liegenden Datenpunkte. Das Verfahren ist so erweitert worden, dass mit Hilfe einer QR-Zerlegung mit Spaltenpivotierung eine Lösung \mathbf{a} mit $k = \text{rang}(\mathbf{G})$ von Null verschiedenen Komponenten berechnet wird. Dadurch wird ein in den Ausgangsgrößenwerten vorhandener Trend wesentlich besser erfasst und die Modellgenauigkeit erhöht. Diese Vorgehensweise ist unabhängig von dem gewählten Ansatz für die Konklusion anwendbar. Der Fall eines überbestimmten Gleichungssystems tritt in der Praxis sehr häufig auf und wurde bisher durch eine herkömmliche Least-Squares Regression (OLS) gelöst. Dieser Ansatz ist jedoch anfällig gegenüber Ausreißern in den Daten, die einen ungewollt starken Einfluss auf die so gefundene Lösung haben können. Um die Robustheit gegenüber solchen Ausreißern zu erhöhen, kann nun alternativ eine Lösung über einen Iteratively Reweighted Least-Squares (IRLS) Ansatz bestimmt werden. Dabei werden die einzelnen Datenpunkte bzw. die zugehörigen Approximationsfehler geeignet gewichtet, wodurch fragwürdige oder fehlerhafte Datenpunkte einen nur geringen bzw. gar keinen Beitrag zu der zu minimierenden Gütefunktion liefern und so bei der Bestimmung der Lösung mehr oder weniger stark ausgeblendet werden.

Durch die Entwicklung des neuen hybriden Fuzzy-Systems ist es gelungen, den Anwendungsbereich der von KIENDL eingeführten negativen Regeln auf TSK-Systeme zu erweitern. Negative Regeln erlauben eine kompakte Formulierung von Zusammenhängen, die in Form von Warnungen oder Verboten vorliegen. Sie eignen sich daher insbesondere für die Gewährleistung sicherheitsrelevanter Aspekte eines Fuzzy-Systems, beispielsweise des Verbots von Prozesseingriffen bei kritischen Zuständen. Der prinzipielle Nutzen des hybriden Fuzzy-Systems ist an Beispielen gezeigt worden.

Wesentliche Fortschritte wurden bei der Formulierung systematischer Lösungsstrategien erzielt. Dieser Aspekt ist für die Akzeptanz des Verfahrens bei potenziellen Anwendern von großer Bedeutung, da eine zeit- und kostenintensive Einarbeitung in das eingesetzte Modellierungsverfahren oftmals für industrielle Anwender nicht in Frage kommt. Durch die systematische Exploration der Einstellmöglichkeiten des Fuzzy-ROSA-Verfahrens und die Erprobung an einer großen Anzahl von Benchmarkproblemen konnte einerseits die

Leistungsfähigkeit der neu entwickelten Strategieelemente gezeigt, andererseits konnten wissensbasiert problemspezifische Einstellempfehlungen für freie Modellparameter formuliert werden. Ferner ist datenbasiert ein prototypisches Modul erzeugt worden, das auf Grundlage von charakteristischen Merkmalen eines zu modellierenden Datensatzes entsprechende Prognosen der zu erwartenden Modellierungsfehler für gewählte Einstellungen freier Modellparameter liefert. Ein solches Modul kann prinzipiell die vollautomatische problemspezifische Einstellung freier Modellparameter übernehmen, falls sich der Anwender hiermit nicht befassen will oder kann.

Für die Weiterentwicklung des Fuzzy-ROSA-Verfahrens und dessen Anwendbarkeit für industrielle Problemstellungen sind verschiedene Aspekte von Interesse. Für die Generierung von Mamdani-Regeln stehen eine Reihe verschiedener Test- und Bewertungsverfahren zur Verfügung, die bei identischer Projektdefinition (siehe Abschnitt 4.2) jeweils ein individuelles Modell liefern. Die so insgesamt erzeugten Modelle können im Hinblick auf die erzielten Modellierungsfehler, die Größe der Regelbasis und andere charakteristische Merkmale miteinander verglichen werden. Ein interessanter weiterführender Ansatz besteht darin, aus den verschiedenen individuellen Modellen ein gemeinsames Modell bzw. eine gemeinsame Regelbasis zu erzeugen. Dieser Ansatz kann bereits auf den Prozess der Regelgenerierung übertragen werden: Eine Hypothese wird nicht nur anhand eines, sondern anhand einer Vielzahl von Test- und Bewertungsverfahren und entsprechender Einstellungen der Parameter κ und τ geprüft. Die endgültige Entscheidung, ob eine Hypothese zur Regel erklärt wird, fällt dann auf Grundlage aller Einzelentscheidungen. Es stellt sich dann die Aufgabe, die verschiedenen Einzelentscheidungen in einer angemessenen Weise zu einer Gesamtentscheidung zu verrechnen. Beispielsweise können Hypothesen, die von unterschiedlichen Test- und Bewertungsverfahren bestätigt werden, als besonders relevant angesehen werden. Weiterhin kann geprüft werden, ob ein bestimmtes Test- und Bewertungsverfahren eine Hypothese bei unterschiedlichen Werten von κ und τ , also bei unterschiedlich strengen Anforderungen, als Regel empfiehlt. Auf diese Weise lassen sich unter Umständen Richtlinien formulieren, wann eine Hypothese tatsächlich zu einer Regel erklärt werden sollte und wann nicht. Ein weiterer Schwerpunkt zukünftiger Arbeiten sollte in der Fortführung der Entwicklung von systematischen Lösungsstrategien liegen. Diesem Aspekt kommt insbesondere im Hinblick auf die Akzeptanz des Verfahrens bei potenziellen Anwendern eine sehr große Bedeutung zu. Die in dieser Arbeit erzielten Ergebnisse erhärten die Vermutung, dass solche Strategien auf der Grundlage von Experimenten formuliert werden können. Die bisherigen Untersuchungen zur Entwicklung einer Systematik haben sich bezüglich der generierten Modelle auf die Merkmale Modellierungsfehler und Größe der Regelbasis konzentriert. Eine Ausdehnung auf weitere Merkmale, wie beispielsweise den prozentualen Anteil nicht durch Regeln abgedeckter Daten oder die Untersuchung der Vollständigkeit der Regelbasis, dürfte lohnenswert sein. Auch zwischen diesen Merkmalen eines Modells und den Merkmalen des zugehörigen Lerndatensatzes wird es Zusammenhänge geben, deren Beachtung bei einer systematischen Vorgehensweise Vorteile bringt.

Anhang A

Test- und Bewertungsverfahren für die Regelgenerierung

Wie in Abschnitt 4.3 ausgeführt, lassen sich die im Fuzzy-ROSA-Verfahren zur Verfügung stehenden Test- und Bewertungsverfahren in zwei Klassen aufteilen. Eine Klasse bilden diejenigen Verfahren, die auf Datenverteilungen arbeiten. Diese werden in Abschnitt 6.2 genauer vorgestellt. Die andere Klasse bilden diejenigen Verfahren, die auf Datenhäufigkeiten arbeiten. Diese Verfahren werden hier detailliert beschrieben.

Relevanzindex (RI)

Der Regeltest prüft die Frage, ob zwischen der Prämisse P und der Konklusion C einer Regel ein kausaler Zusammenhang besteht. Der Bewertungsindex für positive bzw. negative Regeln ist wie folgt:

$$J_{RI} = \begin{cases} \frac{\hat{p}_u^\alpha(C|P) - \hat{p}_l^\alpha(C)}{1 - \hat{p}(C)} & \text{falls } \hat{p}_u^\alpha(C|P) > \hat{p}_l^\alpha(C) \quad \text{für positive Regeln} \\ \frac{\hat{p}_u^\alpha(C) - \hat{p}_l^\alpha(C|P)}{1 - \hat{p}(C)} & \text{falls } \hat{p}_l^\alpha(C|P) < \hat{p}_u^\alpha(C) \quad \text{für negative Regeln} \end{cases} \quad (\text{A.1})$$

Dabei kennzeichnet α das Alpha-Niveau für die Konfidenzintervalle. Je größer α gewählt wird, desto größer ist die zulässige Unsicherheit, dass der tatsächliche Wert nicht innerhalb des Konfidenzintervalls liegt. Je kleiner α gewählt wird, desto schwerer ist folglich der Regeltest. Für den Relevanzindex steuert α die Schwere des Regeltests und ist daher dem Parameter κ , der allgemein für die Bezeichnung der Schwere des Regeltests verwendet wird, gleichzusetzen.

Relevante Trefferquote (RHR)

Der hier eingesetzte Regeltest ist identisch mit dem des Relevanzindex. Der Bewertungsindex ist aber unabhängig von der geschätzten unbedingten Wahrscheinlichkeit $\hat{p}(C)$:

$$J_{RHR} = \begin{cases} \hat{p}(C|P) & \text{falls } \hat{p}_u^\alpha(C|P) > \hat{p}_l^\alpha(C) \text{ für positive Regeln} \\ 1 - \hat{p}(C|P) & \text{falls } \hat{p}_l^\alpha(C|P) < \hat{p}_u^\alpha(C) \text{ für negative Regeln} \end{cases} \quad (\text{A.2})$$

Auch für dieses Verfahren ist α dem allgemeinen Parameter κ gleichzusetzen.

Normierte Trefferquote (NHR)

Bei diesem Ansatz sind sowohl der Regeltest als auch der Bewertungsindex unabhängig von der geschätzten unbedingten Wahrscheinlichkeit $\hat{p}(C)$:

$$J_{NHR} = \begin{cases} \frac{\hat{p}(C|P) - \theta}{1 - \theta} & \text{falls } \hat{p}(C|P) > \theta \text{ für positive Regeln} \\ \frac{\vartheta - \hat{p}(C|P)}{\vartheta} & \text{falls } \hat{p}(C|P) < \vartheta \text{ für negative Regeln} \end{cases} \quad (\text{A.3})$$

Typischerweise werden $\theta = 0.5$ und $\vartheta = 0.5$ gewählt. Es werden keine Konfidenzintervalle zu den geschätzten relativen Häufigkeiten verwendet. Hier wird die Schwere des Regeltests über θ und ϑ gesteuert, die stets identische Werte annehmen. Daher sind für dieses Verfahren θ und ϑ dem allgemeinen Parameter κ gleichzusetzen.

Konfidente normierte Trefferquote (CHR)

Der hier verfolgte Ansatz ist im Prinzip identisch zu dem der normierten Trefferquote, allerdings wird hier entsprechend die jeweils relevante Grenze des Konfidenzintervalls zu der geschätzten bedingten relativen Häufigkeit verwendet und nicht die bedingte relative Häufigkeit selbst:

$$J_{CHR} = \begin{cases} \frac{\hat{p}_u^\alpha(C|P) - \theta}{1 - \theta} & \text{falls } \hat{p}_u^\alpha(C|P) > \theta \text{ für positive Regeln} \\ \frac{\vartheta - \hat{p}_l^\alpha(C|P)}{\vartheta} & \text{falls } \hat{p}_l^\alpha(C|P) < \vartheta \text{ für negative Regeln} \end{cases} \quad (\text{A.4})$$

Typischerweise werden $\theta = 0.5$ und $\vartheta = 0.5$ gewählt. Es werden Konfidenzintervalle zu den geschätzten relativen Häufigkeiten verwendet, die abhängig sind von dem Alpha-Niveau α . Die Schwere des Regeltests wird in diesem Fall über die Parameter θ und α bzw. ϑ und α gesteuert. Im Rahmen der in dieser Arbeit durchgeführten Untersuchungen (siehe Kapitel 9) wird jedoch der Parameter α stets auf dem empfohlenen Default-Wert belassen, sodass die Schwere des Regeltests ausschließlich über die Parameter θ und ϑ kontrolliert wird, die stets identische Werte annehmen. Für dieses Verfahren sind daher θ und ϑ dem allgemeinen Parameter κ gleichzusetzen.

Inkrementelle Trefferquote (ITQ)

Der Bewertungsindex verwendet unmittelbar den Abstand zwischen der geschätzten unbedingten relativen Häufigkeit und der geschätzten bedingten relativen Häufigkeit. Eine Normierung wird nicht vorgenommen. Dieser Bewertungsindex ist in erster Linie für die Verwendung in Verbindung mit dem Konzept des Inkrementellen Relevanzindex zur Regelgenerierung (siehe Abschnitt 6.4) vorgesehen:

$$J_{ITQ} = \begin{cases} \hat{p}(C|P) - \hat{p}(C) & \text{falls } \hat{p}_u^\alpha(C|P) > \hat{p}_l^\alpha(C) \text{ für positive Regeln} \\ \hat{p}(C) - \hat{p}(C|P) & \text{falls } \hat{p}_l^\alpha(C|P) < \hat{p}_u^\alpha(C) \text{ für negative Regeln} \end{cases} \quad (\text{A.5})$$

Konfidente inkrementelle Trefferquote (KITQ)

Der hier verfolgte Ansatz ist sehr ähnlich zu dem der inkrementellen Trefferquote, allerdings wird hier entsprechend die jeweils relevante Grenze des Konfidenzintervalls zu der geschätzten unbedingten relativen Häufigkeit verwendet und nicht die unbedingte relative Häufigkeit selbst. Dieser Bewertungsindex ist ebenfalls für die Verwendung in Kombination mit dem Konzept des Inkrementellen Relevanzindex zur Regelgenerierung vorgesehen:

$$J_{KITQ} = \begin{cases} \hat{p}(C|P) - \hat{p}_u^\alpha(C) & \text{falls } \hat{p}_u^\alpha(C|P) > \hat{p}_l^\alpha(C) \text{ für positive Regeln} \\ \hat{p}_l^\alpha(C) - \hat{p}(C|P) & \text{falls } \hat{p}_l^\alpha(C|P) < \hat{p}_u^\alpha(C) \text{ für negative Regeln} \end{cases} \quad (\text{A.6})$$

Anhang B

Vorstellung der untersuchten Anwendungen und Benchmarkprobleme

Im Folgenden werden die Anwendungen und Benchmarkprobleme, die im Rahmen der systematischen Exploration der Einstellmöglichkeiten untersucht werden, getrennt nach Approximations- und Klassifikationsproblemen vorgestellt¹. Bei den nachfolgend vorgestellten Benchmarkproblemen bezeichnet N die Anzahl der Eingangsgrößen und D die Anzahl der Datenpunkte des gesamten zur Verfügung stehenden Datensatzes. Zu jedem Benchmarkproblem wird eine kurze Beschreibung der jeweiligen Modellierungsaufgabe gegeben und es werden ausgewählte charakteristische Merkmale des jeweiligen Datensatzes aufgeführt. Weiterhin werden für jede der Eingangs- und für die Ausgangsgröße ein Box-Plot und ein Histogramm-Plot der Daten für die jeweilige Größe gezeigt.

Die Darstellungen sind dabei zeilenweise aufgebaut: Die Grafik in der linken oberen Ecke bezieht sich jeweils auf die Eingangsgröße X_1 , die Grafik rechts daneben auf X_2 und so fort. Die letzte Grafik in der untersten Zeile bezieht sich auf die Ausgangsgröße Y . Den Darstellungen ist die Partitionierung überlagert, die durch das unidimensionale Clusterverfahren für die jeweilige Größe erzeugt wird. Wie in Abschnitt 9.2 ausgeführt, wird in erster Linie diese Partitionierung für die systematische Exploration der Einstellmöglichkeiten verwendet. Zur besseren Übersicht und Vergleichbarkeit sind die Größen auf den Wertebereich $[0, 1]$ normiert. Daher wird auch auf eine Beschriftung der Achsen verzichtet.

Ein Box-Plot bzw. Box-Whisker-Plot erlaubt eine kompakte Darstellung der Verteilung von Werten und eine Aussage über eventuell vorhandene Ausreißer. Der Aufbau ist wie folgt: Die Box wird durch das erste Quartil P_{25} (25. Perzentil) und das dritte Quartil P_{75} (75. Perzentil) der Messwerte einer Größe definiert. Innerhalb dieser Box liegt der Median (zweites Quartil P_{50} bzw. 50. Perzentil). Der Abstand zwischen dem dritten und dem

¹Die Datensätze hierzu finden sich beispielsweise unter <http://neural.cs.nthu.edu.tw/jang/benchmark/>, <http://lib.stat.cmu.edu/datasets/> und <http://www.ics.uci.edu/~mlearn/MLSummary.html>

ersten Quartil wird als Inter-Quartile-Range (IQR) bezeichnet. Dieser Wert liefert eine Aussage über die Spannweite der gemessenen Werte und ist robust gegenüber eventuell vorhandenen Ausreißern, da die untersten und obersten 25% der Messwerte hierbei nicht beachtet werden. Als Abschätzung für Ausreißer werden dabei folgende Grenzen festgelegt: $L = P_{25} - \kappa \cdot IQR$ als untere Grenze und $U = P_{75} + \kappa \cdot IQR$ als obere Grenze. Hier wird $\kappa = 1.5$ gewählt. Von den Enden der Box aus werden die so genannten Whiskers² zur Abschätzung bezüglich Ausreißern nach unten und oben aufgetragen. Diese gehen bis zu den extremsten Werten in den Messwerten, die noch innerhalb des oben genannten Bereichs liegen, in dem Datenpunkte nicht als Ausreißer eingestuft werden. Datenpunkte, die außerhalb dieses Bereichs liegen und daher als Ausreißer angesehen werden, sind durch Kreuze dargestellt.

Zu jedem Benchmarkproblem werden charakteristische Merkmale des jeweils zu Grunde liegenden Datensatzes angegeben. Entsprechende Merkmale werden zunächst für jede Größe individuell berechnet. Die Berechnung der Merkmale erfolgt dabei anhand normierter Daten (siehe obige Ausführungen bzgl. der Visualisierung), um eine spätere Aggregation durchführen zu können und um eine Vergleichbarkeit zwischen verschiedenen Größen, aber auch zwischen verschiedenen Problemen zu ermöglichen. In Abschnitt 9.2 wird ausgeführt, dass die Anzahl der so bestimmten Merkmale abhängig ist von der Anzahl der Eingangsgrößen eines Modellierungsproblems. Dies wird für die Entwicklung eines Moduls zur Prognose des Modellierungsfehlers als ungünstig angesehen (siehe Abschnitt 9.6). Daher werden die Merkmale über alle Eingangsgrößen aggregiert, sodass sich unabhängig von der Anzahl an Eingangsgrößen eines Datensatzes stets die gleiche Anzahl an Merkmalen ergibt (siehe Abschnitt 9.5). Die Aggregation geschieht durch Bildung des Median.

Folgende Merkmale werden in den nachfolgenden Abschnitten aufgeführt: MedianAD, Std, Skewness, IQR, RatioDataLow, RatioDataHigh, In2InCorr bzw. In2OutCorr. Dabei steht MedianAD für Median Average Deviation und Std für Standard Deviation. Beides sind Maße für die Streuung der Werte innerhalb einer Stichprobe. Skewness ist ein Maß für die Symmetrie der Verteilung von Werten einer Stichprobe. Die Abkürzung IQR steht für Inter-Quartile-Range. Dieser Begriff wird in Zusammenhang mit der Darstellungsform des Box-Plots erläutert (siehe oben). Das Verhältnis der Anzahl an Daten einer Stichprobe, deren Wert unterhalb des ersten Quartils P_{25} der Stichprobe liegen, zu der Gesamtgröße der Stichprobe wird mit RatioDataLow bezeichnet. Dementsprechend wird das Verhältnis der Anzahl an Daten einer Stichprobe, deren Wert oberhalb des dritten Quartils P_{75} der Stichprobe liegen, zu der Gesamtgröße der Stichprobe mit RatioDataHigh bezeichnet. Diese beiden Maße beschreiben den Anteil an Ausreißern innerhalb einer Stichprobe. Für die Eingangsgrößen wird das Maß In2InCorr angegeben. Dieses ist die gemittelte Korrelation zwischen den Eingangsgrößen. Für die Ausgangsgröße wird das Maß In2OutCorr angegeben. Dieses ist die gemittelte Korrelation zwischen einer Eingangs- und der Ausgangsgröße.

²Im Deutschen am ehesten mit *Schnurrhaare* zu übersetzen.

Approximationsprobleme

Wärmetauscher ($N = 3$, $D = 1126$)

In einem chemischen Prozess soll die Temperatur innerhalb eines Reaktors mithilfe eines Wärmetauschers konstantgehalten werden [Kra02]. Die Temperatur des Öls innerhalb des Wärmetauschers ist abhängig von der Reaktortemperatur, dem Reaktordruck und der Stellgröße für den Wärmetauscher. Die Modellierungsaufgabe besteht darin, die Temperatur des Öls anhand der drei genannten Größen vorherzusagen¹.

Anhand der grafischen Darstellungen ist zu erkennen, dass die Werte aller drei Eingangsgrößen eine eher schiefe Verteilung zeigen und Ausreißer auftreten. Gleiches gilt für die Ausgangsgrößenwerte. Die unsymmetrischen Verteilungen spiegeln sich in entsprechend hohen Werten für die Skewness wider. Die Partitionierung der Eingangsgrößen durch das unidimensionale Clusterverfahren gestaltet sich entsprechend schwierig und das Ergebnis erscheint hier verbesserungsfähig. Insbesondere bei der Eingangsgröße X_2 würde man eher erwarten, dass der Wertebereich, in dem sich die meisten Daten konzentrieren, feiner aufgelöst wird als dies hier der Fall ist. Für die Eingangsgröße X_3 werden lediglich drei linguistische Werte vorgesehen. Bei einer wissensbasierten Partitionierung würde man wohl eher mehr linguistische Werte verwenden.

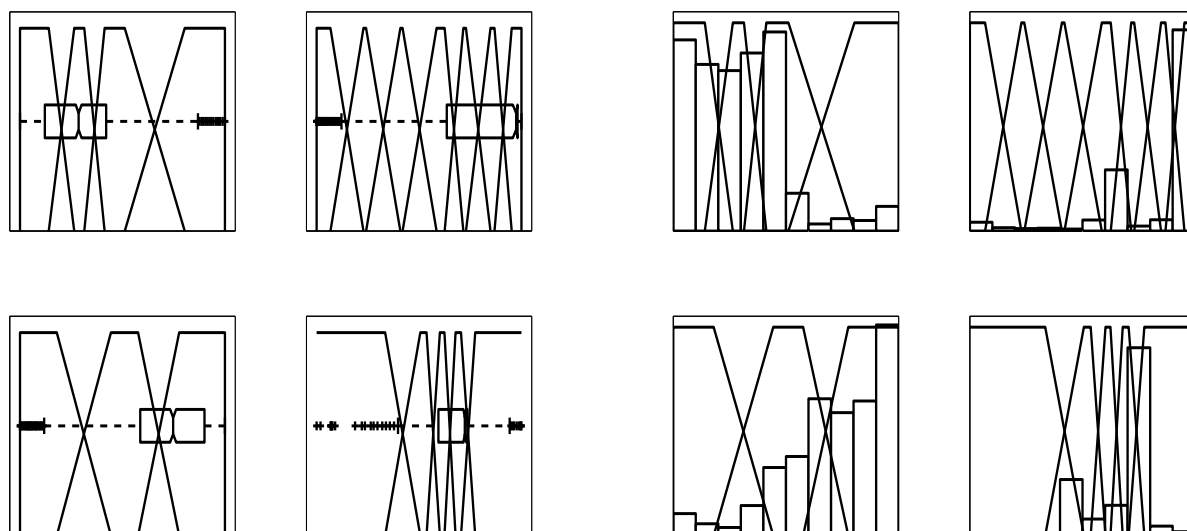


Abbildung B.1: Box-Plots (links) und Histogramm-Plots (rechts) für das Benchmarkproblem Wärmetauscher.

¹Dieser Datensatz stammt aus einem Industrieprojekt und ist daher nicht frei zugänglich.

MedianAD	0.150	MedianAD	0.019
Std	0.229	Std	0.132
Skewness	-0.908	Skewness	-1.321
IQR	0.314	IQR	0.141
RatioDataLow	0.029	RatioDataLow	0.016
RatioDataHigh	0.000	RatioDataHigh	0.007
In2InCorr	0.821	In2OutCorr	0.522

Tabelle B.1: Aggregierte Merkmale für die Eingangsgrößen (links) und für die Ausgangsgröße (rechts) des Benchmarkproblems Wärmetauscher.

Mackey-Glass-Zeitreihe ($N = 4$, $D = 1000$)

Die Mackey-Glass-Zeitreihe [MG77] ist ein in der Literatur häufig zum Testen von Modellierungsverfahren eingesetztes Benchmarkproblem. Es handelt sich hierbei um eine chaotische Zeitreihe, deren diskrete Form wie folgt gegeben ist:

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

Als Parameter werden $a = 0.1$, $b = 0.2$ und $\tau = 17$ gewählt. Es ist $x(t < 0) = 0$ und $x(t = 0) = 1.2$. Die Modellierungsaufgabe besteht in diesem Fall darin, den zukünftigen Wert $x(t+6)$ anhand des aktuellen Wertes $x(t)$ und dreier zeitlich zurückliegender Werte $x(t-18)$, $x(t-12)$ und $x(t-6)$ zu prognostizieren.

Man erkennt in den grafischen Darstellungen deutlich, dass sich die Eingangsgrößen sehr stark ähneln, da sie die gleichen Werte, nur zeitlich versetzt, beinhalten. Daher sind auch die erzeugten Partitionierungen nahezu identisch. Dies zeigt sich auch anhand der hohen Korrelation der Eingangsgrößen untereinander. Für die Ausgangsgröße werden durch das Clusterverfahren vier linguistische Werte erzeugt. Ein Vergleich der Modellierungsergebnisse, die mit dieser automatischen Partitionierung erzielt werden, mit denen verschiedener anderer Standard-Partitionierungen zeigt, dass diese Anzahl offenbar nicht ausreicht, um eine gute Modellierung zu erzielen (siehe Abschnitt 9.3). Bessere Modellierungsergebnisse stellen sich ein, wenn die Anzahl der linguistischen Ausgangsgrößenwerte auf fünf oder sieben erhöht wird.

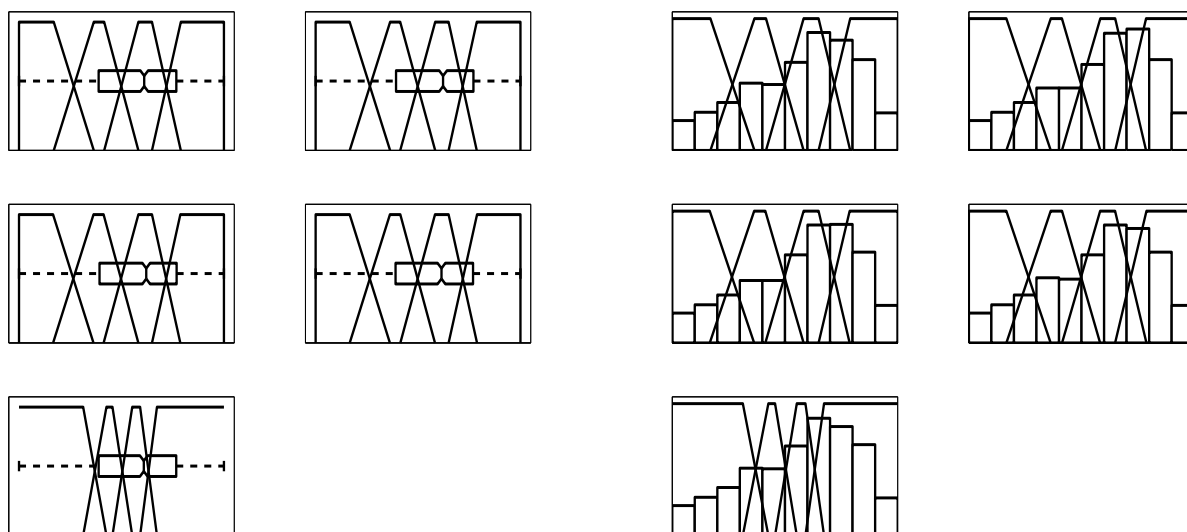


Abbildung B.2: Box-Plots (links) und Histogramm-Plots (rechts) für das Benchmarkproblem Mackey-Glass-Zeitreihe.

MedianAD	0.173	MedianAD	0.177
Std	0.241	Std	0.241
Skewness	-0.478	Skewness	-0.441
IQR	0.378	IQR	0.379
RatioDataLow	0.000	RatioDataLow	0.000
RatioDataHigh	0.000	RatioDataHigh	0.000
In2InCorr	0.589	In2OutCorr	0.590

Tabelle B.2: Aggregierte Merkmale für die Eingangsgrößen (links) und für die Ausgangsgröße (rechts) des Benchmarkproblems Mackey-Glass-Zeitreihe.

Operator ($N = 5$, $D = 70$)

Bei diesem Benchmarkproblem [SY93] besteht die Aufgabe darin, die Eingriffe eines menschlichen Experten in einen Prozess nachzubilden. Das Problem wird durch fünf Eingangsgrößen beschrieben. Der Datensatz beinhaltet lediglich 70 Datenpunkte, sodass bei einer Kreuzvalidierung nur 35 Lerndatenpunkte für die Erstellung eines Modells genutzt werden. Für die Eingangsgröße X_1 werden lediglich drei linguistische Werte vorgesehen. Auffällig ist auch, dass für die Eingangsgrößen X_4 und X_5 sieben bzw. fünf linguistische Werte erzeugt werden, die entsprechenden Zugehörigkeitsfunktionen hierzu sich aber nicht überlappen. Anhand der Histogramm-Darstellung zeigt sich aber, dass diese Partitionierung angebracht ist. Sowohl die Korrelation der Eingangsgrößen untereinander als auch deren Korrelation mit der Ausgangsgröße sind gering.

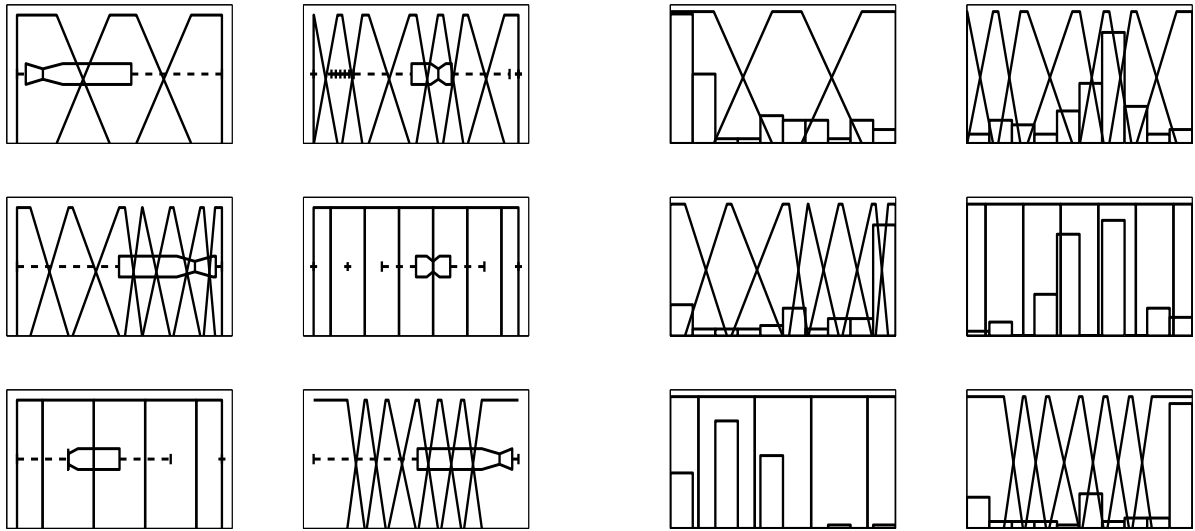


Abbildung B.3: Box-Plots (links) und Histogramm-Plots (rechts) für das Benchmarkproblem Operator.

MedianAD	0.103	MedianAD	0.092
Std	0.214	Std	0.345
Skewness	-0.161	Skewness	-0.843
IQR	0.250	IQR	0.462
RatioDataLow	0.000	RatioDataLow	0.000
RatioDataHigh	0.014	RatioDataHigh	0.000
In2InCorr	0.201	In2OutCorr	0.334

Tabelle B.3: Aggregierte Merkmale für die Eingangsgrößen (links) und für die Ausgangsgröße (rechts) des Benchmarkproblems Operator.

Roboterkinematik ($N = 8$, $D = 8192$)

Die Aufgabe besteht darin, den Abstand des Endeffektors eines 8-Gelenk-Roboterarms von einem vorgegebenen Zielpunkt vorherzusagen. Für dieses Problem stehen insgesamt vier unterschiedliche Datensätze zur Verfügung, die sich dahingehend unterscheiden, inwieweit Nichtlinearitäten des Roboters und Messrauschen beachtet werden². Die Beachtung dieser Aspekte erschwert in der Regel die Modellierung. Die Ausführungen beziehen sich im Folgenden auf den Datensatz *KinFH*, der nur geringe Nichtlinearitäten und wenig Rauschen beinhaltet.

²Detaillierte Angaben finden sich unter <http://www.cs.toronto.edu/delve/data/pumadyn/desc.html>

Es zeigen sich deutliche Parallelen in den generierten Partitionierungen für die Eingangsgrößen X_1 und X_8 , für X_2 , X_3 und X_7 sowie für X_4 , X_5 und X_6 . Auffällig ist, dass die Anzahl der linguistischen Werte von Gruppe zu Gruppe verschieden ist, obwohl die Werte aller Eingangsgrößen näherungsweise gleichverteilt sind. Es ist daher eher zu erwarten, dass für alle Eingangsgrößen die gleiche Anzahl linguistischer Werte erzeugt wird. Die unterschiedliche Granularisierung erscheint wenig plausibel. Die Ausgangsgrößenwerte sind näherungsweise normalverteilt und die erzeugte Partitionierung erscheint im Prinzip angemessen, da der Bereich um den Mittelwert feiner aufgelöst wird als die an den Rändern liegenden Bereiche. Die Diskussion der Modellierungsergebnisse in Abschnitt 9.3 zeigt aber, dass es von Vorteil ist, die Randbereiche ebenfalls feiner zu granularisieren.

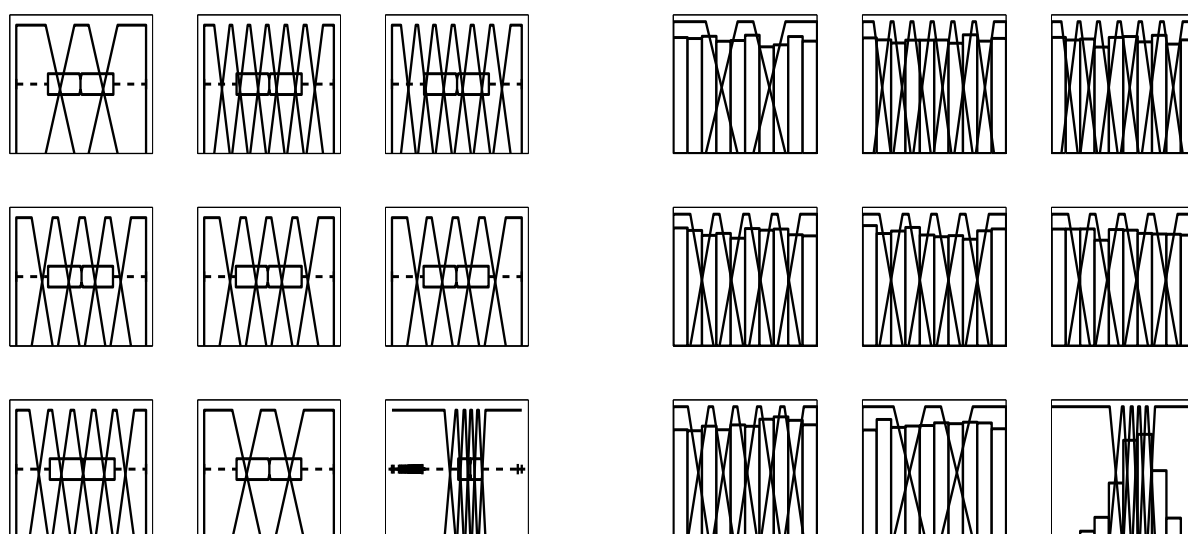


Abbildung B.4: Box-Plots (links) und Histogramm-Plots (rechts) für das Benchmarkproblem Roboterkinematik (KinFH).

MedianAD	0.251	MedianAD	0.092
Std	0.289	Std	0.138
Skewness	-0.005	Skewness	-0.400
IQR	0.501	IQR	0.185
RatioDataLow	0.000	RatioDataLow	0.011
RatioDataHigh	0.000	RatioDataHigh	0.000
In2InCorr	0.007	In2OutCorr	0.287

Tabelle B.4: Aggregierte Merkmale für die Eingangsgrößen (links) und für die Ausgangsgröße (rechts) des Benchmarkproblems Roboterkinematik (KinFH).

Box-Jenkins-Zeitreihe ($N = 10$, $D = 290$)

Die Aufgabe besteht in diesem Fall darin, die aktuelle CO_2 -Konzentration anhand zeitlich zurückliegender Konzentrationswerte und des Gaszuflusses zu prognostizieren [BJ70].

Man erkennt deutliche Parallelen in den generierten Partitionierungen für die Eingangsgrößen X_1 bis X_4 , und für X_5 bis X_{10} . Die erste Gruppe bilden die zurückliegenden Gaskonzentrationen, die zweite Gruppe bilden die protokollierten Gaszuflüsse. Sowohl die Eingangsgrößen- als auch die Ausgangsgrößenwerte sind näherungsweise normalverteilt. Es zeigt sich eine hohe Korrelation sowohl zwischen den Eingangsgrößen als auch zwischen Eingangs- und Ausgangsgröße.

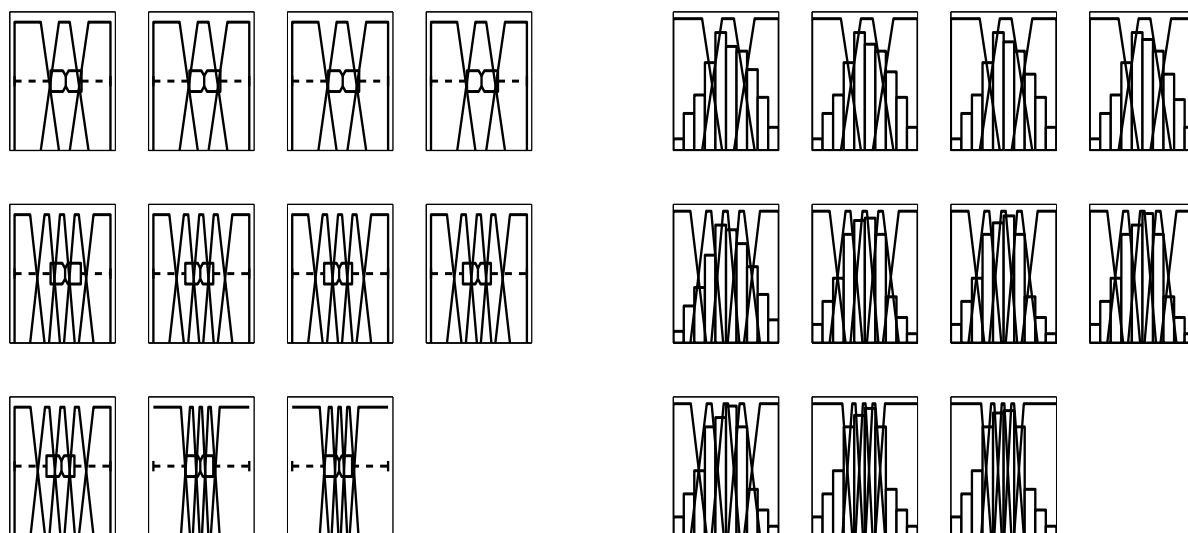


Abbildung B.5: Box-Plots (links) und Histogramm-Plots (rechts) für das Benchmarkproblem Box-Jenkins-Zeitreihe.

MedianAD	0.154	MedianAD	0.143
Std	0.205	Std	0.195
Skewness	-0.046	Skewness	-0.047
IQR	0.303	IQR	0.291
RatioDataLow	0.000	RatioDataLow	0.000
RatioDataHigh	0.000	RatioDataHigh	0.000
In2InCorr	0.733	In2OutCorr	0.843

Tabelle B.5: Aggregierte Merkmale für die Eingangsgrößen (links) und für die Ausgangsgröße (rechts) des Benchmarkproblems Box-Jenkins-Zeitreihe.

Stock Prices ($N = 10, D = 100$)

Bei diesem Problem soll die Entwicklung eines Aktienkurses anhand von zehn aussagekräftigen Merkmalen prognostiziert werden [SY93].



Abbildung B.6: Box-Plots (links) und Histogramm-Plots (rechts) für das Benchmarkproblem Stock Prices.

MedianAD	0.117	MedianAD	0.088
Std	0.199	Std	0.200
Skewness	-0.232	Skewness	-0.161
IQR	0.223	IQR	0.282
RatioDataLow	0.015	RatioDataLow	0.000
RatioDataHigh	0.000	RatioDataHigh	0.010
In2InCorr	0.307	In2OutCorr	0.380

Tabelle B.6: Aggregierte Merkmale für die Eingangsgrößen (links) und für die Ausgangsgröße (rechts) des Benchmarkproblems Stock Prices.

Boston Housing ($N = 13, D = 506$)

Die ursprüngliche Aufgabenstellung bei diesem Problem besteht darin, anhand von Merkmalen sowohl die Grundstückspreise als auch die Schadstoffbelastungen in verschiedenen Stadtteilen von Boston zu modellieren [HR78]. Es wird hier jedoch nur die Aufgabe der Prognose von Grundstückspreisen betrachtet. Die Merkmale erfassen dabei die Kriminali-

tätsrate, den Anteil der Industrie und den Armutsanteil der Bevölkerung für den jeweiligen Stadtteil sowie weitere für den Preis relevante Kenngrößen.

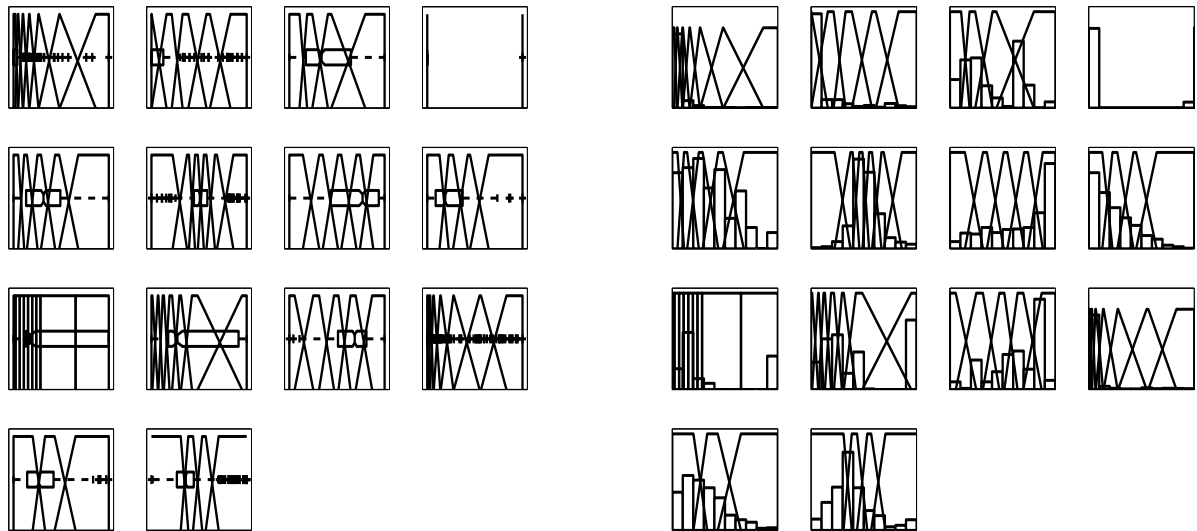


Abbildung B.7: Box-Plots (links) und Histogramm-Plots (rechts) für das Benchmarkproblem Boston Housing.

MedianAD	0.117	MedianAD	0.089
Std	0.233	Std	0.204
Skewness	0.904	Skewness	1.105
IQR	0.283	IQR	0.178
RatioDataLow	0.000	RatioDataLow	0.004
RatioDataHigh	0.010	RatioDataHigh	0.075
In2InCorr	0.384	In2OutCorr	0.388

Tabelle B.7: Aggregierte Merkmale für die Eingangsgrößen (links) und für die Ausgangsgröße (rechts) des Benchmarkproblems Boston Housing.

Klassifikationsprobleme

Iris ($N = 4$, $D = 150$)

Die Aufgabe besteht in diesem Fall darin, drei Arten von Schwertlilien anhand von vier charakteristischen Merkmalen zu unterscheiden [And35, PPB97]. Dieses Problem findet sich häufig in der Literatur zur Erprobung von Klassifikationsverfahren. Aufgrund der relativ

geringen Anzahl an Eingangsgrößen lassen sich die Daten in diesem Fall vergleichsweise gut visualisieren und analysieren.

Auffällig ist in erster Linie die vergleichsweise geringe Anzahl linguistischer Werte, die für die Eingangsgrößen vorgesehen werden. Bei einer manuellen Partitionierung würde man wohl eher fünf statt drei linguistische Werte verwenden. Die Ausgangsgröße wird hier manuell partitioniert. Dabei werden gemäß der Anzahl an Klassen drei linguistische Werte und entsprechende Zugehörigkeitsfunktionen hierzu in Form von Singletons festgelegt. Die Korrelation unter den Eingangsgrößen ist vergleichsweise hoch, insbesondere aber die Korrelation zwischen Eingangs- und Ausgangsgröße.

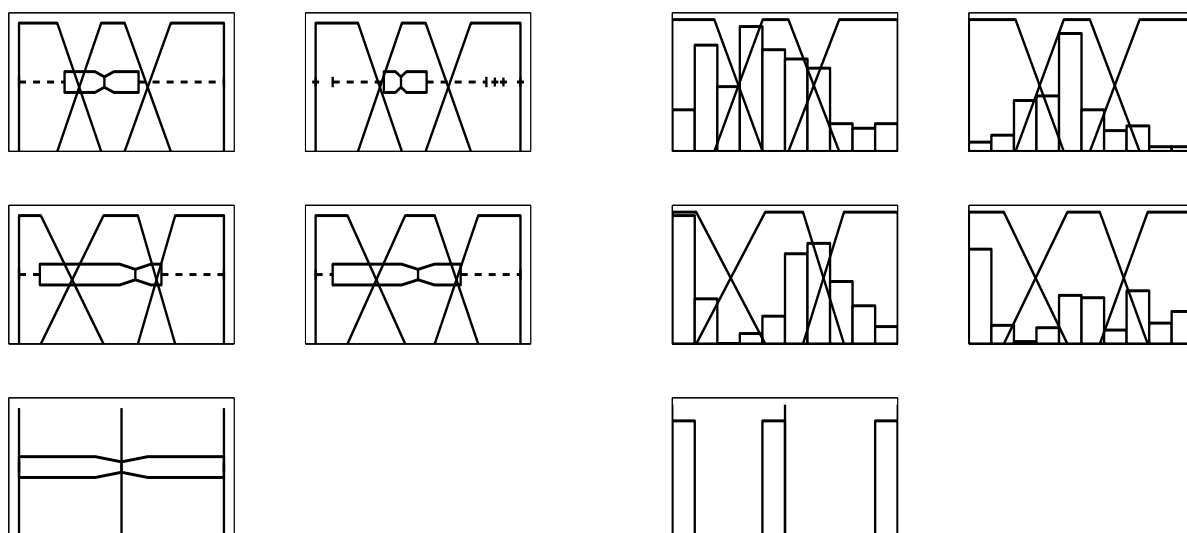


Abbildung B.8: Box-Plots (links) und Histogramm-Plots (rechts) für das Benchmarkproblem Iris.

MedianAD	0.203	MedianAD	0.500
Std	0.265	Std	0.410
Skewness	0.104	Skewness	0.000
IQR	0.477	IQR	1.000
RatioDataLow	0.000	RatioDataLow	0.000
RatioDataHigh	0.000	RatioDataHigh	0.000
In2InCorr	0.619	In2OutCorr	0.866

Tabelle B.8: Aggregierte Merkmale für die Eingangsgrößen (links) und für die Ausgangsgröße (rechts) des Benchmarkproblems Iris.

Wine ($N = 13$, $D = 178$)

In diesem Problem ist die regionale Herkunft italienischer Weine anhand ihrer chemischen Zusammensetzung zu bestimmen [CS94]. Drei unterschiedliche Herkunftsregionen werden dabei unterschieden.

Bei den erzeugten Partitionierungen für die Eingangsgrößen fällt auf, dass für neun der dreizehn Größen nur drei bzw. vier linguistische Werte vorgesehen werden. Für die übrigen vier Eingangsgrößen sind es fünf bis sieben linguistische Werte. Die Ausgangsgröße wird manuell partitioniert. Es werden drei linguistische Werte und entsprechende Zugehörigkeitsfunktionen hierzu in Form von Singletons festgelegt.

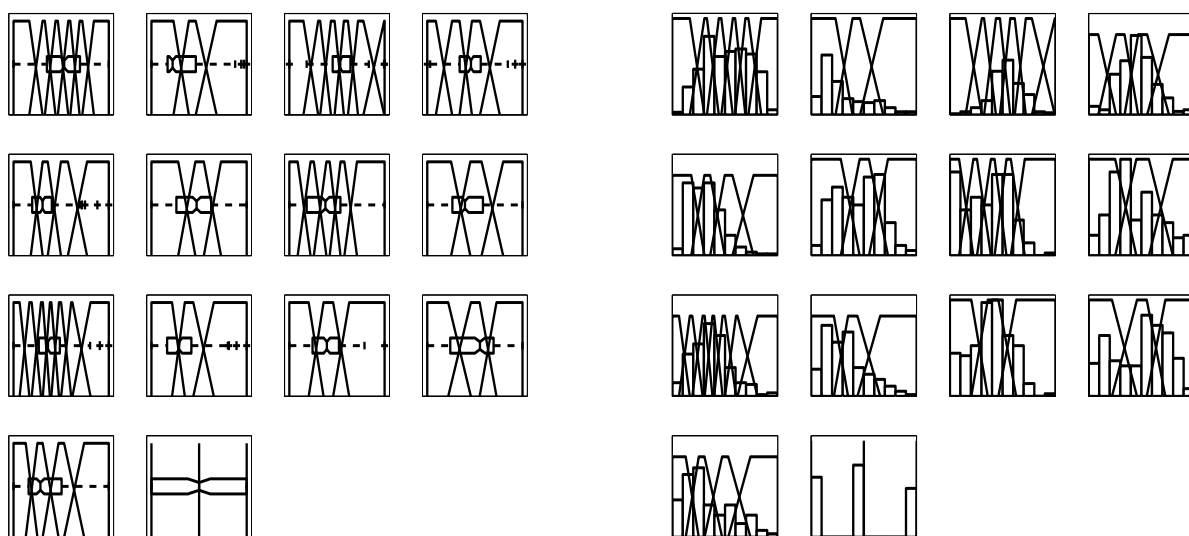


Abbildung B.9: Box-Plots (links) und Histogramm-Plots (rechts) für das Benchmarkproblem Wine.

MedianAD	0.134	MedianAD	0.500
Std	0.211	Std	0.388
Skewness	0.211	Skewness	0.107
IQR	0.296	IQR	1.000
RatioDataLow	0.000	RatioDataLow	0.000
RatioDataHigh	0.006	RatioDataHigh	0.000
In2InCorr	0.282	In2OutCorr	0.499

Tabelle B.9: Aggregierte Merkmale für die Eingangsgrößen (links) und für die Ausgangsgröße (rechts) des Benchmarkproblems Wine.

Anhang C

Maße zur Beurteilung der Modellierungsgüte

Der zu modellierende Datensatz \mathcal{D} bestehe aus D Datenpunkten. Für jeden Datenpunkt $\mathbf{d}_i = [\mathbf{x}_i, y_i]$ wird anhand des zugehörigen Eingangsgrößenvektors \mathbf{x}_i mithilfe eines wie auch immer gearteten Modells eine Prognose $\hat{y}_i = f(\mathbf{x}_i)$ für den zugehörigen Ausgangsgrößenwert y_i erzeugt.

Der relative Klassifikationsfehler (RCE) des Modells bezüglich des zu modellierenden Datensatzes berechnet sich zu

$$RCE = \frac{1}{D} \cdot \sum_{i=1}^D \Delta_\epsilon(y_i - \hat{y}(\mathbf{x}_i)) \quad , \quad (\text{C.1})$$

wobei die Fehlerfunktion Δ_ϵ wie folgt definiert ist:

$$\Delta_\epsilon(y_i - \hat{y}(\mathbf{x}_i)) = \begin{cases} 0, & \text{falls } y_i - \hat{y}(\mathbf{x}_i) = 0, \\ 1, & \text{falls } y_i - \hat{y}(\mathbf{x}_i) \neq 0 \end{cases} \quad . \quad (\text{C.2})$$

Der mittlere absolute Fehler (MAE) ergibt sich zu:

$$MAE = \frac{1}{D} \cdot \sum_{i=1}^D |y_i - \hat{y}(\mathbf{x}_i)| \quad . \quad (\text{C.3})$$

Der mittlere quadratische Fehler (MSE) ergibt sich zu:

$$MSE = \frac{1}{D} \cdot \sum_{i=1}^D (y_i - \hat{y}(\mathbf{x}_i))^2 \quad . \quad (\text{C.4})$$

Die Fehlermaße MAE und MSE sind problemspezifisch zu bewerten, da sie im Kontext der Größenordnung zu verstehen sind, in der sich die Ausgangsgrößenwerte bewegen. Alle

oben genannten Fehlermaße haben den Nachteil, dass sie keine Aussage in dem Sinne erlauben, wie groß der erzielte Nutzen im Verhältnis zu dem betriebenen Aufwand ist. Um diese Nachteile zu überwinden, werden nachfolgend normierte Fehlermaße vorgestellt, die auch einen Vergleich der erzielten Ergebnisse auf verschiedenen Benchmarkproblemen vereinfachen.

Die normierten Fehlermaße setzen die für ein Modell erzielten Fehlerwerte RCE bzw. MAE und MSE in das Verhältnis zu den Fehlerwerten, die ein denkbar einfaches Modell, ein so genanntes Trivialmodell, erzielt. Dieses Trivialmodell gibt zu einem beliebigen Eingangsgrößenvektor \mathbf{x} stets die gleiche konstante Ausgabe \hat{y}_{def} und wird in der Literatur auch als *Baseline-Prognose* bezeichnet. Ein normiertes Fehlermaß von Eins bedeutet somit, dass ein wie auch immer geartetes Modell den gleichen Fehler erzielt wie die Baseline-Prognose. Anhand des normierten Fehlermaßes kann so unmittelbar der Zugewinn an Modellgenauigkeit (also der erzielte Nutzen) abgelesen werden. Für den Fall, dass $\hat{y}_{def} = \bar{y}$ gewählt wird, werden diese Fehlermaße auch als *goodness of fit* bezeichnet.

$$RCE_{norm} = \frac{\sum_{i=1}^D \Delta_{\epsilon}(y_i - \hat{y}(\mathbf{x}_i))}{\sum_{i=1}^D \Delta_{\epsilon}(y_i - \hat{y}_{def})} \quad (\text{C.5})$$

$$MAE_{norm} = \frac{\sum_{i=1}^D |y_i - \hat{y}(\mathbf{x}_i)|}{\sum_{i=1}^D |y_i - \hat{y}_{def}|} \quad (\text{C.6})$$

$$MSE_{norm} = \frac{\sum_{i=1}^D (y_i - \hat{y}(\mathbf{x}_i))^2}{\sum_{i=1}^D (y_i - \hat{y}_{def})^2} \quad (\text{C.7})$$

Neben den oben genannten sind oftmals auch solche Maße von Interesse, die das Modellverhalten auf unbekanntem Daten (Validierungsdaten) charakterisieren. Hierzu zählt beispielsweise das so genannte *unbiased criterion* ϵ_{UC} [SK88]:

$$\epsilon_{UC} = \sqrt{\frac{1}{D_A} \sum_{i=1}^{D_A} \Delta_{\epsilon}(\hat{y}^A(\mathbf{x}_q^A) - \hat{y}^B(\mathbf{x}_q^A)) + \frac{1}{D_B} \sum_{i=1}^{D_B} \Delta_{\epsilon}(\hat{y}^B(\mathbf{x}_q^B) - \hat{y}^A(\mathbf{x}_q^B))} \quad (\text{C.8})$$

Hierfür werden zwei Beobachtungsdatensätze \mathcal{D}_A und \mathcal{D}_B des gleichen Systems erzeugt und so zwei unterschiedliche Modelle generiert. Beide Datensätze werden an beiden Modellen ausgewertet, wobei jeweils einer der Datensätze nicht zum Erlernen des jeweils betrachteten Modells eingesetzt wird und so für das Modell unbekannte Daten darstellt. So kann geschätzt werden, wie unvoreingenommen sich die Modelle auf unbekanntem Daten verhalten.

Weiterhin wird häufig auch die Korrelation r zwischen den gemessenen und den modellierten Ausgangsgrößenwerten als Maß für die Qualität der Modellierung verwendet. Der Vorteil dieses Maßes liegt darin, dass es unabhängig von dem Wertebereich der Ausgangsgrößenwerte zu interpretieren ist:

$$r = \frac{\sum_{i=1}^D (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sum_{i=1}^D \sqrt{(y_i - \bar{y})^2} \sqrt{(\hat{y}_i - \bar{\hat{y}})^2}} \quad . \quad (\text{C.9})$$

Literaturverzeichnis

- [Abd80] ABDELMALEK, N. N.: L_1 solution of overdetermined systems of linear equations. In: *ACM Transactions on Mathematical Software* Bd. 6, 1980, S. 220–227
- [And35] ANDERSON, E.: The IRISes of the Gaspé Peninsula. In: *Bull. Amer. IRIS Soc.* 59 (1935), S. 2–5
- [And73] ANDERBERG, M. R.: *Cluster analysis for applications*. New York, USA : Academic Press, 1973
- [BE99] BLOCHWITZ, S. ; EIGERMANN, J.: Creation and Application of Fuzzy Rules to Access the Creditworthiness of Enterprises. In: *Proceedings of the Fourth European Congress on Intelligent Techniques and Soft Computing (EUFIT '99)* Bd. CD-ROM. Aachen : Verlag Mainz, 1999
- [Bel61] BELLMAN, R. E.: *Adaptive control processes: A guided tour*. Princeton : Princeton University Press, 1961
- [Bez81] BEZDEK, J. C.: *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York, USA : Plenum Press, 1981
- [BJ70] BOX, G. E. P. ; JENKINS, G. M.: *Time Series Analysis, Forecasting and Control*. San Francisco, CA : Holden Day, 1970
- [BSMM99] BRONSTEIN, I. N. ; SEMENDJAJEW, G. ; MUSIOL, G. ; MÜHLIG, H.: *Taschenbuch der Mathematik*. 4. überarbeitete und erweiterte Auflage der Neubearbeitung. Frankfurt am Main, Thun : Harri Deutsch, 1999
- [CD96] CASTRO, J. ; DELGADO, M.: Fuzzy Systems with Defuzzification are Universal Approximators. In: *IEEE Transactions on Systems, Man, and Cybernetics* 26 (1996), Nr. 1, S. 149–152
- [CDM91] COLORNI, A. ; DORIGO, M. ; MANIEZZO, V.: Distributed optimization by ant colonies. In: *Proceedings First European Conference on Artificial Life*. Cambridge, MA : MIT Press, 1991, S. 134–142

- [CS94] CORCORAN, A. L. ; SANDIP, S.: *Using Real-Valued Genetic Algorithms to Evolve Rule Sets for Classifications*. Bd. 1. Piscataway, NJ : IEEE Press, 1994. – 120–124 S
- [CW82] COOK, R. D. ; WEISBERG, S.: *Residuals and Influence in Regression*. London : Chapman and Hall, 1982
- [DMC96] DORIGO, M. ; MANIEZZO, V. ; COLORNI, A.: The Ant System: Optimization by a Colony of Cooperating Agents. In: *IEEE Transactions on Systems, Man, and Cybernetics* B 26 (1996), S. 29–46
- [DO89] DUMOUCHEL, W. H. ; O'BRIEN, F. L.: Integrating a Robust Option into a Multiple Regression Computing Environment. In: *Computer Science and Statistics: Proceedings of the 21st Symposium on the Interface* American Statistical Association, Alexandria, VA, 1989
- [Gos08] GOSSET, W. S.: The probable error of a mean. In: *Biometrika* 6 (1908), Nr. 1, S. 1–25. – unter dem Pseudonym *Student* veröffentlicht
- [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
- [Hae03] HAENDEL, L.: *Clusterverfahren zur datenbasierten Generierung interpretierbarer Regeln unter Verwendung lokaler Entscheidungskriterien*, Universität Dortmund, Diss., 2003. – <http://hdl.handle.net/2003/2834>
- [Har95] HARTUNG, J.: *Statistik: Lehr- und Handbuch der angewandten Statistik*. München : Oldenbourg, 1995 (10. Auflage)
- [HEK95] HARTUNG, J. ; ELPELT, B. ; KLOESENER, K.-H.: *Statistik*. München : Oldenbourg, 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
- [Hol75] HOLLAND, J. H.: *Adaptation in Natural and Artificial Systems*. Michigan, USA : Universität Michigan Press, Ann Arbor, 1975
- [HR78] HARRISON, D. ; RUBINFELD, D. L.: Hedonic Prices and the Demand for Clean Air. In: *Economics & Management* 5 (1978), S. 81–102
- [Hub81] HUBER, P. J.: *Robust statistics*. New York, USA : John Wiley & Sons, 1981
- [HW77] HOLLAND, P. W. ; WELSCH, R. E.: Robust regression using iteratively reweighted least-squares. In: *Communications in Statistics - Theory and Methods* A6 (1977), S. 813–827

- [HW79] HARTIGAN, J. A. ; WONG, M. A.: Algorithm AS 136: A K -Means Clustering Algorithm. In: *Applied Statistics* 28 (1979), Nr. 1, S. 100–108
- [Jan98] JANIKOW, C. Z.: Fuzzy Decision Trees: Issues and Methods. In: *IEEE Transactions on Systems, Man and Cybernetics, Part B* 28 (1998), S. 1–14
- [Jes99] JESSEN, H.: A Mean-Value Based Test- and Rating Strategy for Automatic Fuzzy Rule Generation and Application to Load Prediction. In: *International Conference on Computational Intelligence for Modelling Control, Wien, Österreich*. Amsterdam, Niederlande : IOS Press, 1999
- [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
- [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
- [JS98a] JESSEN, H. ; SLAWINSKI, T: Mittelwertbasierter Regeltest und Bewertung für das Fuzzy-ROSA-Verfahren und Anwendung zur Lastprognose. In: *Tagungsband des 8. Workshop Fuzzy Control VDI/VDE GMA-FA 5.22*, Forschungsbericht der Fakultät für Elektrotechnik, Nr. 0298, Universität Dortmund, 1998, S. 67–81
- [JS98b] JESSEN, H. ; SLAWINSKI, T.: Test- and Rating Strategies for Data-based Rule Generation. CI-39/98, Sonderforschungsbereich 531, Universität Dortmund, 1998. – Forschungsbericht
- [KB78] KOENKER, R. ; BASSETT, G.: Regression Quantiles. In: *Econometrica* 46 (1978), Nr. 1, S. 33–50
- [KBT96] KRONE, A. ; BÄCK, T. ; TEUBER, P.: Evolutionäres Suchkonzept zum Aufstellen signifikanter Fuzzy-Regeln. In: *at – Automatisierungstechnik* 44 (1996), Nr. 8, S. 405–411
- [KE95] KENNEDY, J. ; EBERHART, R.: Particle Swarm Optimization. In: *Proceedings of the IEEE International Conference on Neural Networks* Bd. IV. Piscataway, NJ, USA : IEEE-Press, 1995, S. 1941–1948

- [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
- [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.: Invarianzforderungen für Inferenzfilter. In: *Tagungsband des 4. Workshops Fuzzy Control des GMA-FA 1.4.2, Dortmund*. Dortmund : Forschungsbericht der Fakultät für Elektrotechnik, Nr. 0194, Universität Dortmund, 1994, S. 1–12
- [Kie97a] KIENDL, H.: *Fuzzy Control methodenorientiert*. München, Wien : 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
- [Kie00] KIENDL, H.: Implizite Modellierung, inkrementeller Relevanzindex und Rauigkeitsmaß: neue Strategieelemente für die datenbasierte Modellierung. In: *Proceedings 10. Workshop Fuzzy Control des GMA-FA 5.22*. Karlsruhe : Wissenschaftliche Berichte Forschungszentrum Karlsruhe (FZKA 6509), 2000, S. 1–14
- [Kin94] KINNEBROCK, W.: *Neuronale Netze*. 2. München Wien : Oldenbourg, 1994
- [KJ97] KOHAVI, R. ; 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
- [KK01] KIENDL, H. ; KRAUSE, P.: Advanced Inference Filter Defuzzification. In: REUSCH, B. (Hrsg.): *Computational Intelligence: Theory and Applications (Proceedings of 7th Fuzzy-Days), Dortmund*. Berlin : Springer, 2001, S. 249–260
- [KKR06] KIENDL, H. ; KISELIOVA, T. ; RAMBINITSOA, Y.: Use of Fuzzy Strategies for Interpretation of Music. In: *Journal of Multiple-Valued Logic and Soft Computing, Special Issue Dedicated to the Memory of Professor Dr. Helmut Thiele (1926-2003)* 12 (2006), Nr. 1-2
- [KKS00a] KRAUSE, P. ; KRONE, A. ; SLAWINSKI, T.: Fuzzy System Identification by Generating and Evolutionary Optimizing Fuzzy Rule Bases Consisting of Relevant Fuzzy Rules / SFB 531, Universität Dortmund. 2000. – Forschungsbericht. Reihe CI 84/00
- [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
- [KKSS02a] KIENDL, H. ; KRAUSE, P. ; SLAWINSKI, T. ; SCHAUTEN, D.: Data-Based Fuzzy Modeling for Complex Applications. In: SCHWEFEL, H.-P. (Hrsg.) ; WEGENER, I. (Hrsg.) ; WEINERT, K. (Hrsg.): *Advances in Computational Intelligence*. Berlin : Springer, 2002, S. 46–77
- [KKSS02b] KIENDL, H. ; KRAUSE, P. ; SLAWINSKI, T. ; SCHAUTEN, D.: Fuzzy Modeling to Cope with Ambiguities. In: SCHWEFEL, H.-P. (Hrsg.) ; WEGENER, I. (Hrsg.) ; WEINERT, K. (Hrsg.): *Advances in Computational Intelligence*. Berlin : Springer, 2002, S. 78–103
- [Koe05] KOENKER, R.: *Quantile regression*. Cambridge, England : Cambridge University Press, 2005

- [Kos94] KOSKO, B.: Fuzzy Systems as Universal Approximators. In: *IEEE Transactions on Computers* 43 (1994), Nr. 11, S. 1329–1333
- [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)
- [Kra00] KRAUSE, P.: Generierung von Takagi–Sugeno–Fuzzy–Systemen aus relevanten Fuzzy–Regeln. In: *Proceedings 10. Workshop Fuzzy Control des GMA–FA 5.22*. Karlsruhe : Wissenschaftliche Berichte Forschungszentrum Karlsruhe (FZKA 6509), 2000, S. 84–97
- [Kra02] KRAUSE, P.: *Datenbasierte Generierung von transparenten und genauen Fuzzy–Modellen für mehrdeutige Daten und komplexe Systeme*. Düsseldorf : VDI Verlag, 2002 (Fortschritt-Berichte VDI, Reihe 10, Nr. 691)
- [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)
- [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
- [KSS06] KROHLING, R. ; SCHAUTEN, D. ; STERNKE, M.: Particle Swarm Optimization Applied to a Control System of a Hydraulic Valve. In: *Proceedings of the IEEE Swarm Intelligence Symposium (SIS 2006)* IEEE Computational Intelligence Society and IEEE Central Indiana Section, 2006, S. 32–37
- [KSSH05] KROHLING, R. ; STERNKE, M. ; SCHAUTEN, D. ; HOFFMANN, F.: Partikel Schwarm Optimierung eines Regelungssystems für ein industrielles Hydraulikventil. In: *Proceedings 15. Workshop Computational Intelligence*. Karlsruhe : Universitätsverlag Karlsruhe, 2005, S. 227–240
- [KT01] KRONE, A. ; TAEGER, H.: Data–based fuzzy rule test for fuzzy modelling. In: *Fuzzy Sets and Systems* 126 (2001), S. 343–358
- [KW00] KOJADINOVIC, I. ; WOTTKA, T.: Comparison between a filter and a wrapper approach to variable subset selection in regression problems. In: *Proceedings of the European Symposium on Intelligent Techniques (ESIT)*, 2000
- [Mat96] MATHAR, R.: *Informationstheorie*. Stuttgart : Teubner Verlag, 1996
- [McQ67] MCQUEEN, J.: Some methods for classification and analysis of multivariate observations. In: *Proceedings 5th Berkely Symposium on Mathematical Statistics and Probability*, 1967, S. 281–298

- [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
- [Moo20] MOORE, E. H.: On the reciprocal of the general algebraic matrix. In: *Bulletin of the American Mathematical Society* Bd. 26, 1920, S. 394–395
- [Nie96] NIEWELS, F.: Analytische Formeln für das Inferenzfilter zur Berechnung gefilterter Zugehörigkeitsfunktionen. Forschungsbericht der Fakultät für Elektrotechnik, Nr. 0896, Universität Dortmund, 1996. – Forschungsbericht
- [Pen55] PENROSE, R.: A generalized inverse for matrices. In: *Proceedings of the Cambridge Philosophy Society* Bd. 51, 1955, S. 406–413
- [PJK⁺02] PFEIFFER, B.-M. ; JÄKEL, J. ; KROLL, A. ; KUHN, C. ; KUNTZE, H.-B. ; LEHMANN, U. ; SLAWINSKI, T. ; TEWS, V.: Erfolgreiche Anwendungen von Fuzzy Logik und Fuzzy Control (Teil 1/2). In: *at – Automatisierungstechnik* 10/11 (2002), Nr. 50/51, S. 461–471/511–521
- [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
- [Pra99] PRACZYK, J.: *Entwurf von Fuzzy-Gütemaßen zur Prozessbewertung*. Düsseldorf : VDI Verlag, 1999 (Fortschritt–Berichte VDI, Reihe 8, Nr. 796)
- [PTVF99] PRESS, W. H. ; TEUKOLSKY, S. A. ; VETTERLING, W. T. ; FLANNERY, B. P.: *Numerical Recipes in C. 2*. Cambridge, England : Cambridge University Press, 1999
- [Qui86] QUINLAN, J. R.: Induction of Decision Trees. In: *Machine Learning* 1 (1986), S. 81–106
- [Qui93] QUINLAN, J. R.: *C4.5: Programs for Machine Learning*. San Mateo : Morgan Kaufman, 1993
- [Rec94] RECHENBERG, I.: *Evolutionstrategie '94, Werkstatt Bionik und Evolutionstechnik*. Bd. 1. Stuttgart : Frommann–Holzboog, 1994
- [RJHR05] ROBINSON, E. J. ; JACKSON, D. E. ; HOLCOMBE, M. ; RATNIEKS, F. L.: Insect communication: 'No entry' signal in ant foraging. In: *Nature* 438 (2005), S. 442

- [RL87] ROUSSEEUW, P. J. ; LEROY, A. M.: *Robust Regression and Outlier Detection*. New York, USA : John Wiley & Sons, 1987
- [Roj96] ROJAS, R.: *Theorie der neuronalen Netze*. Berlin : Springer, 1996
- [Rou84] ROUSSEEUW, P.: Least Median of Squares Regression. In: *Journal of the American Statistics Association* 79 (1984), S. 871–880
- [Rya97] RYAN, T. P.: *Modern Regression Methods*. New York, USA : John Wiley & Sons, 1997
- [Sal99] SALZBERG, S.: *On comparing classifiers: A critique of current research and methods*. Boston : Kluwer Academic Publishers, 1999
- [Sch95] SCHWEFEL, H.-P.: *Evolution and Optimum Seeking*. New York, USA : Sixth-Generation Computer Technology Series, 1995
- [Sch99] SCHWANE, U.: *Datenbasierte Generierung von Adaptionsregeln und Anwendung zur Erhöhung der Bahngenaugigkeit eines Industrieroboters*. Düsseldorf : VDI Verlag, 1999 (Fortschritt-Berichte VDI, Reihe 8, Nr. 748)
- [SJP+00] SLAWINSKI, T. ; JESSEN, H. ; PRACZYK, J. ; KRAUSE, P. ; KRONE, A. ; KIENDL, H.: Einsatz der datenbasierten Fuzzy-Modellierungen für komplexe Anwendungen. In: *Computational Intelligence im industriellen Einsatz, Tagung Baden Baden, 2000*. Düsseldorf : VDI-Verlag, 2000 (VDI-Berichte, Nr. 1526), S. 119–124
- [SK88] SUGENO, M ; KANG, G. T.: Structure Identification of Fuzzy-Model. In: *Fuzzy Sets and Systems* 28 (1988), S. 15–33
- [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
- [SKK04] STERNKE, M. ; KRAUSE, P. ; KIENDL, H.: Erweiterung des Fuzzy-ROSA-Verfahrens für die datenbasierte Regelgenerierung bei kleinen und stark gestörten Datensätzen. In: *Proceedings 14. Workshop Fuzzy-Systeme und Computational Intelligence*. Karlsruhe : Universitätsverlag Karlsruhe, 2004, S. 55–68
- [SKK05] STERNKE, M. ; KRAUSE, P. ; KIENDL, H.: Nutzung negativer Regeln in TSK-Fuzzy-Systemen. In: *Proceedings 15. Workshop Computational Intelligence*. Karlsruhe : Universitätsverlag Karlsruhe, 2005, S. 89–103

- [SKK06] STERNKE, M. ; KRAUSE, P. ; KIENDL, H.: Formulierung problemspezifischer Einstellregeln für das Fuzzy-ROSA-Verfahren. In: *Proceedings 16. Workshop Computational Intelligence*. Karlsruhe : Universitätsverlag Karlsruhe, 2006. – erscheint 2006
- [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
- [SKKS05] STERNKE, M. ; KIENDL, H. ; KRAUSE, P. ; SCHWEFEL, H.-P.: Transferprojekt T1: Datenbasierte Generierung und Optimierung von Fuzzy-Modellen mit dem Fuzzy-ROSA-Verfahren für komplexe Systeme und Prozesse. In: *SFB 531: Design und Management komplexer technischer Prozesse und Systeme mit Methoden der Computational Intelligence, Finanzierungsantrag 2006-2008*. Bonn, Deutschland : Deutsche Forschungsgemeinschaft, 2005, S. 647–674
- [SKKS06] STERNKE, M. ; KIENDL, H. ; KRAUSE, P. ; SCHWEFEL, H.-P.: *TFB 37: Datenbasierte Generierung und Optimierung von Modellen mit Methoden der Computational Intelligence zur Lösung industrieller Anwendungsprobleme, Abschlussbericht*. Bonn, Deutschland : Deutsche Forschungsgemeinschaft, 2006
- [SKMM03] SCHAUTEN, D. ; KIENDL, H. ; MEYER, J. ; MACHE, D. H.: Rekonstruktionsbasierte Selektion relevanter Einflussgrößen / SFB 531, Universität Dortmund. 2003. – Forschungsbericht. Reihe CI 154/03
- [SKS95] SEYFARTH, R. ; KRONE, A. ; SCHWANE, U.: Fuzzy-Datenanalyse zur regelbasierten Modellierung der Achsregelgüte eines sechsachsigen Industrieroboters. In: *5. Workshop Fuzzy Control VDI/VDE GMA-UA 1.4.2*, Forschungsbericht der Fakultät für Elektrotechnik, Nr. 0295, Universität Dortmund, 1995, S. 71–84
- [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)
- [Smi80] SMITH, S. F.: *A Learning System Based on Genetic Adaptive Systems*, University of Pittsburgh, Diss., 1980
- [SNK01a] SCHAUTEN, D. ; NICOLAUS, B. ; KIENDL, H.: An Evolutionary Concept for Selecting Relevant Sets of Input Variables for Data-based Fuzzy Modeling. In: *Proceedings of the European Symposium on Intelligent Technologies, Hybrid Systems and their Implementation on Smart Adaptive Systems (EUNITE '01)* Bd. CD-ROM. Aachen : Verlag Mainz, 2001

- [SNK01b] SCHAUTEN, D. ; NICOLAUS, B. ; KIENDL, H.: Evolutionäres Verfahren zur Selektion relevanter Merkmalsätze für die datenbasierte Fuzzy-Modellierung. In: *Proceedings 11. Workshop Fuzzy Control des GMA-FA 5.22*. Karlsruhe : Wissenschaftliche Berichte Forschungszentrum Karlsruhe (FZKA 6509), 2001, S. 133–147
- [SY93] SUGENO, M. ; YASUKAWA, T.: A fuzzy-logic-based approach to qualitative modeling. In: *IEEE Transactions on Fuzzy Systems* 1 (1993), Nr. 1, S. 7–31
- [TDT99] THOMAIDIS, N. ; DOUNIAS, G. ; TSELENTIS, G.: Stock Exchange Market Analysis Using Data Engine and Winrosa. In: *Proceedings of the Fourth European Congress on Intelligent Techniques and Soft Computing (EUFIT '99)* Bd. CD-ROM. Aachen : Verlag Mainz, 1999
- [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
- [Vap95] VAPNIK, V.: *The Nature of Statistical Learning Theory*. New York : Springer, 1995
- [Vog89] VOGEL, F.: *Beschreibende und schließende Statistik*. München : Oldenbourg, 1989
- [Wil45] WILCOXON, F.: Individual Comparisons by Ranking Methods. In: *Biometrics* 1 (1945), S. 80–83
- [XW05] XU, R. ; WUNSCH, D.: Survey of Clustering Algorithms. In: *IEEE Transactions on Neural Networks* 16 (2005), Nr. 3, S. 645–678
- [YS95] YUAN, Y. ; SHAW, M. J.: Induction of Fuzzy Decision Trees. In: *Fuzzy Sets and Systems* 69 (1995), S. 125–139
- [Zad65] ZADEH, L. A.: Fuzzy Sets. In: *Information Control* 8 (1965), S. 338–353
- [Zad68] ZADEH, L. A.: Probability measures of fuzzy events. In: *Journal of Mathematical Analysis and Applications* 23 (1968), S. 421–427

Lebenslauf

Persönliche Daten

Name: Martin Sterneke
Geburtstag: 7. Mai 1975
Geburtsort: Dortmund
Familienstand: ledig

Schulbildung

1981 – 1985 Busenberg-Grundschule, Dortmund
1985 – 1994 Phoenix-Gymnasium, Dortmund
5/1994 Abitur

Zivildienst

6/1994 – 8/1995 Behindertenfahrdienst der
Dortmunder Stadtwerke AG, Dortmund

Studium

10/1995 – 8/2001 Diplomstudiengang Elektrotechnik
an der Universität Dortmund
während des Studiums Praktika bei den Firmen
– Fuzzy-Logik Systeme GmbH, Dortmund
– Robert Bosch GmbH, Schwieberdingen
8/2001 Diplom-Hauptprüfung

Berufliche Tätigkeit

10/2001 – 2/2003 wissenschaftlicher Angestellter am Lehrstuhl für
Elektrische Steuerung und Regelung
der Universität Dortmund
(Prof. em. Dr. rer. nat. H. Kiendl,
Lehrstuhlvertretung Dr. rer. nat. F. Hoffmann)
3/2003 – 12/2005 wissenschaftlicher Angestellter am Lehrstuhl für
Systemanalyse der Universität Dortmund
(Prof. a. D. Dr.-Ing. H.-P. Schwefel)
seit 1/2006 wissenschaftlicher Angestellter am Lehrstuhl für
Regelungssystemtechnik der Universität Dortmund
(Univ.-Prof. Dr.-Ing. Prof. h. c. T. Bertram)