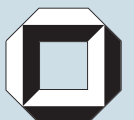
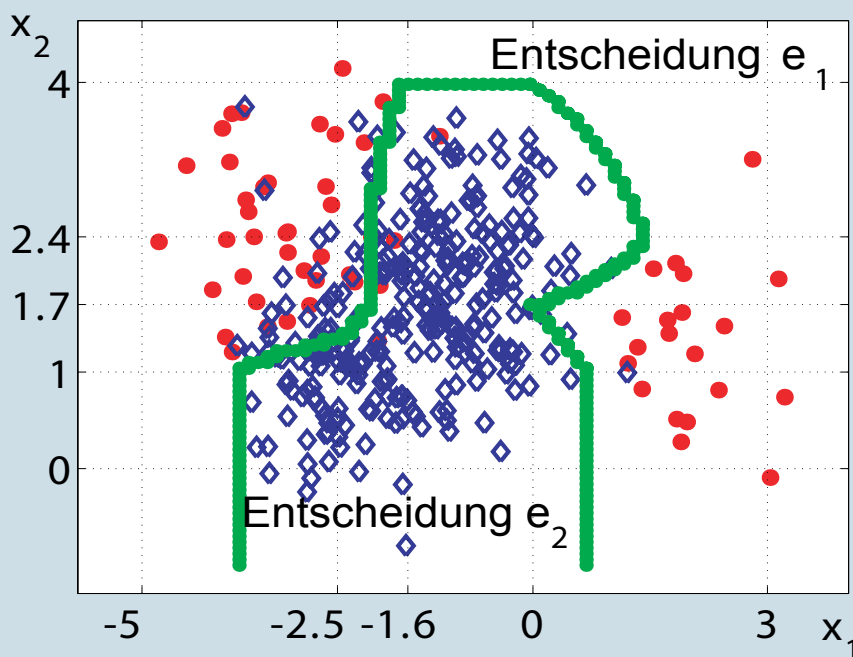


Sebastian Beck

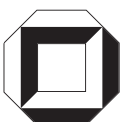
Ein Konzept zur automatischen Lösung von Entscheidungs- problemen bei Unsicherheit mittels der Theorie der unscharfen Mengen und der Evidenztheorie



Schriftenreihe des Instituts für
Angewandte Informatik / Automatisierungstechnik
Universität Karlsruhe (TH)
Band 1

Sebastian Beck

**Ein Konzept zur automatischen
Lösung von Entscheidungs-
problemen bei Unsicherheit
mittels der Theorie der unscharfen
Mengen und der Evidenztheorie**



Dissertation, Universität Karlsruhe (TH), Fakultät für Maschinenbau, 2004

Impressum

Universitätsverlag Karlsruhe
c/o Universitätsbibliothek
Straße am Forum 2
D-76131 Karlsruhe

www.uvka.de

© Universitätsverlag Karlsruhe 2005
Print on Demand

ISSN 1614-5267
ISBN 3-937300-41-4

Vorwort

Die vorliegende Dissertation entstand während der drei Jahre meiner Tätigkeit am Institut für Angewandte Informatik des Forschungszentrums Karlsruhe.

Besonders danke ich Herrn Prof. Dr.-Ing. habil. Georg Bretthauer, der die Anregung zu dieser Arbeit gab und mich dafür begeisterte, auf dem Gebiet der statistischen Entscheidungstheorie zu forschen. Er gab in vielen Gesprächen wichtige Anregungen und konstruktive Kritik zu dieser Arbeit. Seine Unterstützung bei Veröffentlichungen für wichtige Konferenzen und Tagungen war mir ebenfalls stets sicher.

Herrn Prof. Dr.-Ing. habil. Jürgen Wernstedt danke ich herzlichst für sein der Arbeit entgegen gebrachtes Interesse und die Übernahme des Korreferates.

Ein besonderer Dank gilt Herrn R. Mikut, der jederzeit bereit war, in intensiven fachlichen Diskussionen die Entwicklung der Arbeit zu begleiten. Weiterhin danke ich Herrn J. Martin, der mich bei der Umsetzung des Anwendungsbeispiels der künstlichen Hand unterstützt hat. Mein Dank gilt ebenfalls Herrn S. Schulz für die großzügige Bereitstellung von Prototypen der künstlichen Hand.

Für regen fachlichen Austausch und wertvolle Diskussionen danke ich auch meinen beiden Zimmerkollegen, Herrn M. Reischl und Herrn A. Lehmann. Dank gilt ebenso den anderen Mitarbeitern des Instituts für Angewandte Informatik, die mit Rat und Tat am Gelingen dieser Arbeit mitgewirkt haben.

Schließlich möchte ich mich auch bei meinen Eltern bedanken, ohne deren Vorbild mein erfolgreicher Weg durch Studium und Promotion nicht möglich gewesen wäre. Meiner Freundin danke ich für ihre Geduld und ihr Verständnis während der Fertigstellung der Arbeit.

Karlsruhe, im Januar 2005

Sebastian Beck

Kurzfassung

Viele verschiedene Verfahren ermöglichen den automatischen Entwurf von Entscheidungssystemen bei Unsicherheit. Dazu gehören neben Methoden des maschinellen Lernens auch neuronale Netze und statistische Verfahren. Unterschiedliche Konsequenzen von Fehlentscheidungen werden dabei nur in wenigen Fällen berücksichtigt. Die Möglichkeit, Entscheidungen bei widersprüchlichen oder unzureichenden Informationen an einen Experten zu delegieren ist ebenfalls in den meisten Entwurfsverfahren nur ansatzweise abgebildet.

Diese Arbeit trägt zur Weiterentwicklung von Verfahren zur automatischen Generierung von Fuzzy-Entscheidungssystemen bei. Die Bewertung der Entscheidungssysteme erfolgt dabei auf der Basis der Entscheidungstheorie. Über eine Kostenmatrix werden die Konsequenzen von Fehlentscheidungen modelliert. Die Kosten der Merkmale, die das Entscheidungssystem verwendet, werden optional in die Bewertung einbezogen. Für alle Entwurfsschritte des Fuzzy-Systems werden entscheidungstheoretische Bewertungsmaße entworfen und ihre Anwendung an anschaulichen Beispielen demonstriert. Für die Möglichkeit der Delegation von Entscheidungen an einen Experten wird ein neues Inferenzverfahren auf der Basis der Evidenztheorie entwickelt. Zur Reduktion der Merkmalskosten wird die Entwurfsmethodik für Fuzzy-Systeme um ein Verfahren zum Entwurf von mehrstufigen Regelbasen erweitert. Bei mehrstufigen Regelbasen werden nicht alle Merkmale für jede Entscheidung ausgewertet.

Die Leistungsfähigkeit der entwickelten Maße und Verfahren wird anhand von Benchmark Datensätzen und einem realitätsnahen Beispiel aus der Robotik demonstriert. Der Vorteil der Interpretierbarkeit der Fuzzy-Systeme wird an einem Beispiel gezeigt.

Abstract

Numerous methods allow for the automatic design of decision systems with respect to uncertainty. Among these are machine learning, artificial neural networks and statistical methods. Only very few cases consider the different consequences of false decisions. Most design methods also fall short of the possibility to delegate decisions to an expert in cases of contradictory or insufficient information.

This thesis contributes to the advancement of methods for the automatic design of fuzzy decision systems. These decision systems are assessed using decision theory. A cost matrix is used to model the consequences of false decisions. The attribute costs used by the decision system are considered optionally in the decision system's assessment. References derived from decision theory are developed for all design steps of fuzzy systems. Their use is demonstrated with a set of demonstrative examples. A new interference method to delegate decisions to an expert is developed using evidence theory. The design methodology for fuzzy systems is extended by a procedure to obtain multi-stage rule bases. Such rule bases do not evaluate all criteria for each decision and thus contribute to lower attribute costs.

Benchmark data sets and a realistic robotics example are used to demonstrate the method's performance. The advantages of the interpretability of fuzzy systems is explained using an example.

Inhaltsverzeichnis

Vorwort	I
Kurzfassung	III
Inhaltsverzeichnis	V
1 Einleitung	1
1.1 Bedeutung von Entscheidungssystemen	1
1.2 Darstellung des Entwicklungsstandes	2
1.2.1 Entscheidungstheorie	2
1.2.2 Evidenztheorie	8
1.2.3 Entscheidungssysteme	10
1.2.4 Fuzzy-Logik	13
1.3 Offene Probleme	14
1.4 Ziele der Arbeit	15
2 Neuartige entscheidungstheoretische Bewertung von Fuzzy-Systemen	17
2.1 Überblick	17
2.2 Entscheidungstheorie	17
2.2.1 Entscheidungskosten	18
2.2.2 Merkmalskosten	27
2.2.3 Neuartige zweidimensionale Visualisierung	31
2.3 Fuzzy-System	34
2.3.1 Struktur und Bezeichnungen	34
2.3.2 Wahrscheinlichkeiten und Entscheidungen	36
2.3.3 Fuzzy-Inferenz	44
2.3.4 Modifizierte evidenzbasierte Inferenz bei Fuzzy-Systemen	45
2.4 Neue Bewertungsmaße für Fuzzy-Einzelregeln	51
2.4.1 Optimale Fuzzy-Einzelregel	51
2.4.2 Bewertung als Regelbasis	53
2.4.3 Bewertung mit den Beispielen der Prämisse	55
2.4.4 Bewertung mit den Beispielen der Prämisse und den Fehlerkandidaten	56
2.5 Neue Bewertungsmaße für Fuzzy-Regelbasen	64
2.5.1 Vollständige Regelbasis	64
2.5.2 Unvollständige Regelbasis	64

3	Modifizierte entscheidungstheoretische Entwurfsverfahren für Fuzzy-Systeme	67
3.1	Überblick	67
3.2	Hypothesengenerierung - Einzelregeln	68
3.2.1	Eröffnungsverfahren (Entscheidungsbaum)	68
3.2.2	Verbesserungsverfahren (Pruning)	71
3.3	Hypothesengenerierung - einstufige Regelbasen	78
3.3.1	Eröffnungsverfahren	78
3.3.2	Verbesserungsverfahren - Genetischer Algorithmus	86
3.3.3	Verbesserungsverfahren - Optimierung der Zugehörigkeitsfunktionen	89
3.4	Hypothesengenerierung - mehrstufige Regelbasen	94
3.4.1	Überblick	94
3.4.2	Permutationsstrategie	99
3.4.3	Alpha-Strategie	101
3.5	Implementierung	102
4	Anwendungen	107
4.1	Überblick	107
4.2	Statlog-Datensätze	108
4.2.1	German-Credit	108
4.2.2	Heart-Disease	110
4.3	Kfz-Aggregate Benchmark	116
4.4	Kontakterkennung einer künstlichen Hand	122
4.5	Diskussion der Ergebnisse	132
5	Zusammenfassung und Ausblick	133
	Literaturverzeichnis	137
	Bildverzeichnis	149
	Tabellenverzeichnis	151
6	Anhang	153
6.1	Grundlagen der Evidenztheorie	153
6.2	Datensätze	156
6.2.1	Beispieldatensatz	156
6.2.2	German-Credit	156
6.2.3	Heart-Disease	159
6.2.4	Kfz-Aggregate	160
6.2.5	Kontakterkennung	161
6.3	Implementierung	162
	Lebenslauf	163

Symbolverzeichnis

Die wichtigsten Formelzeichen und ihre Bedeutung sind im Folgenden zusammengestellt:

Formelzeichen	Bedeutung
$a_{l,i}$	Parameter des linguistischen Terms $A_{l,i}$
$A_{l,i}$	linguistischer Term des Merkmals x_l
α	Faktor zur Modellierung der Einstellung gegenüber Risiko
α -Schnitt	α -Schnitt der Aktivierung einer Prämisse: $[0 \dots 1] \mapsto \{0, 1\}$.
b	kontinuierliche Beobachtung b
b_i	diskrete Beobachtung b_i
\mathcal{B}	Raum der Beobachtungen
C_r	Konklusion der Entscheidungsregel R_r
d	Entscheidungsfunktion
e_k	k -te Entscheidung (Handlungsalternative)
e_{def}	Default-Entscheidung
e_{opt}	optimale Entscheidung
$e(P_r)$	Entscheidung der Prämisse P_r
\mathcal{E}	Menge der Entscheidungen
E	Erwartungswert
Ev	Evidenz-Inferenz
ER	Erkennungsrate
f	Verteilungsdichtefunktion
F	Verteilungsfunktion
\mathcal{F}	zulässige Verteilungsfunktionen
G	Glaubwürdigkeit
GA	Genetischer Algorithmus
g	Genotyp
HRB	hierarchische Regelbasis
\mathbf{L}	Kostenmatrix ($N_e \times N_z$)
L_{def}	Kosten, die durch eine Default-Entscheidung entstehen
$L(e_k, z_j)$	Kosten bei der Entscheidung e_k und dem Zustand z_j
L_C	Merkmalskosten pro Entscheidung
$L_{C,l}$	Merkmalskosten pro Entscheidung des Merkmals l
\hat{L}_D	erwartete Entscheidungskosten
\hat{L}_{RW}	Kosten der Rückweisungsentscheidung
$\hat{L}_{RW,ev}$	Kosten pro Entscheidung eines Experten
\hat{L}_T	erwartete Gesamtkosten pro Entscheidung ($\hat{L}_T = \hat{L}_D + L_C$)

$L_{T,E}$	Kosten pro Entscheidung eines Experten
L_Z	Zusatzkosten für nicht erfasste Beispiele eines Umweltzustandes
m	Massefunktion (Evidenztheorie)
m_l	Anzahl linguistischer Terme eines Merkmales
M	Fuzzy-Menge
MK	Merkmalskosten beim Entwurf berücksichtigt
$\mu(P_r)$	Aktivierung der Prämisse P_r
n	Index eines Beispiels
N	Anzahl Beispiele eines Datensatzes
N_a	Anzahl aktivierter Terme einer Teilprämisse
N_b	Anzahl der möglichen Beobachtungen b_i
N_c	Anzahl Datensatzteile Kreuzvalidierung
N_C	Anzahl Beispiele zur Validierung bei Kreuzvalidierung
N_e	Anzahl der möglichen Entscheidungen e_k
N_P	Anzahl Genotypen in Startpopulation
N_{P_r}	Anzahl möglicher Prämissen
N_R	Anzahl Regeln in einer Regelbasis
N_{RB}	Anzahl möglicher Regelbasen
$N_{R,ist}$	Anzahl ausgewählter Regeln
$N_{R,P}$	Anzahl der Regeln nach dem Pruning
$N_{R.soll}$	Anzahl auszuwählender Regeln
N_x	Anzahl der Eingangsgrößen / Merkmale x_l
N_z	Anzahl der möglichen Zustände z_j
$N(z_j)$	Anzahl der Beispiele des Zustandes z_j im Datensatz
\hat{P}	Schätzung einer Wahrscheinlichkeit
$P(b_i z_j)$	bedingte Wahrscheinlichkeit der Beobachtung b_i beim gegebenen Zustand z_j
$P(z_j)$	a-priori Wahrscheinlichkeit des Zustandes z_j
P_r	Prämisse der Entscheidungsregel R_r
P_{RB}	alle Prämissen einer Regelbasis
R_r	Entscheidungsregel
r	Index zur Nummerierung einer Entscheidungsregel R
rh	Ebene einer hierarchischen Regelbasis
RB	Regelbasis
$R_b(e_k b_i)$	Bayessches Risiko der Entscheidung e_k bei der Beobachtung b_i
U	Ungewissheit
U_{max}	maximal zulässige Ungewissheit
W	Widerspruch aus zwei Quellen
W_m	maximal auftretender Widerspruch aus mehreren Quellen
W_{max}	maximal zulässiger Widerspruch
x_a	Merkmal zur Auftrennung des Knotens eines Baumes
x_l	l -tes Merkmal
z_j	j -ter Zustand
z_{C_r}	Zustände, für die $e(P_r)$ die beste Entscheidung ist

1 Einleitung

1.1 Bedeutung von Entscheidungssystemen

Der Einsatzbereich von Systemen, die automatisch Entscheidungen treffen oder Handlungsempfehlungen geben, erstreckt sich unter anderem auf technische, medizinische oder auch betriebswirtschaftliche Probleme. Zu dem ersten Bereich gehören unter anderem Probleme der Fehlerdiagnose von Prozessen [59, 110, 111]. Als Beispiel sei hier die Fehlerdiagnose eines PKW Diesel-Motors genannt [112]. Zu den betriebswirtschaftlichen Problemen zählen die Investitionsentscheidungen eines Unternehmens oder die Bewertung des Kunden einer Bank, der einen Kredit beantragt. Aber auch im Bereich der Medizin befinden sich Systeme zur Entscheidungsunterstützung in der Entwicklung [92, 129]. Der Arzt soll bei der Bestimmung der Diagnose unterstützt werden.

Das Ziel für den Einsatz eines Entscheidungssystems ist es, einen Prozess gemessen an einem Gütekriterium optimal zu betreiben. 'Ein Prozess ist eine Gesamtheit von aufeinander einwirkenden Vorgängen in einem System, durch die Materie, Energie oder auch Information umgeformt, transportiert oder auch gespeichert wird' [49]. Diese Definition ist nicht als Beschränkung des Begriffs Prozess auf technische Prozesse zu verstehen. Der Begriff schließt unter anderem auch biologische, soziale oder ökonomische Prozesse ein. Das Gütekriterium kann zum Beispiel darin bestehen, die Anzahl der Fehlentscheidungen zu minimieren oder das Ergebnis je Entscheidung zu optimieren. Mögliche Ergebnisse sind Kosten, Einnahmen oder andere metrische Größen.

Die Entscheidungstheorie bietet eine Basis für die Bewertung eines Entscheidungssystems nach den Konsequenzen der getroffenen Entscheidungen. Dies bringt insbesondere dann Vorteile, wenn ein gegebenes Problem durch das Entscheidungssystem nur mit Restfehlern lösbar ist und die günstigsten Fehlentscheidungen gesucht sind. Wenn viele unterschiedliche Entscheidungssysteme zu ähnlichen Ergebnissen kommen und das System gesucht ist, das die geringsten Gesamtkosten verursacht, ist eine entscheidungstheoretische Bewertung ebenfalls vorteilhaft.

Neben den gemessen am Gütekriterium besten Entscheidungen ist es eine weitere Herausforderung, Entscheidungssysteme so zu entwerfen, dass die repräsentierten Zusammenhänge nachvollziehbar sind. Dies trägt dazu bei, die Akzeptanz des Entscheidungssystems bei seinen Benutzern zu erhöhen. Das Ziel dieser Arbeit ist ein Beitrag zur Integration der entscheidungstheoretischen Bewertung in ein Verfahren zum Entwurf von Entscheidungssystemen zu leisten, deren Zusammenhänge nachvollziehbar sind.

1.2 Darstellung des Entwicklungsstandes

1.2.1 Entscheidungstheorie

In diesem Abschnitt werden die wesentlichen Grundlagen der Entscheidungstheorie skizziert. Insbesondere werden das Grundmodell einer Entscheidung [8, 28, 83, 88] und Prinzipien, nach denen Entscheidungen im Rahmen dieses Modells getroffen werden können, vorgestellt.

Die Entscheidungstheorie beschäftigt sich mit Entscheidungen und deren Entstehung. Dabei sind zwei Richtungen zu unterscheiden: die deskriptive Entscheidungstheorie und die präskriptive Entscheidungstheorie. Die deskriptive Entscheidungstheorie [54] versucht durch Beobachtungen, Befragungen und Experimente das tatsächliche Entscheidungsverhalten aus psychologischer Sicht zu erklären. Dagegen will die präskriptive Entscheidungstheorie Hilfestellungen (mathematische Modelle und Operationen) geben, wie in einer gegebenen Situation eine Entscheidung am besten getroffen werden soll [8, 22, 28, 34, 83, 88]. Die weiteren Ausführungen beschäftigen sich mit der präskriptiven Entscheidungstheorie.

Grundmodell

Das Grundmodell der Entscheidungstheorie besteht aus der Zielfunktion eines Entscheidungsträgers und dem Entscheidungsfeld, das vom betrachteten Entscheidungsproblem abhängt. Der Entscheidungsträger ist dabei eine Person oder eine Instanz, die Entscheidungen trifft. Die Zielfunktion des Entscheidungsträgers muss eine Bewertung der Handlungsalternativen zulassen. Eine mögliche Zielfunktion besteht zum Beispiel in der Maximierung des Gewinns. Das Entscheidungsfeld [28, 83, 88, 95, 122, 126] besteht aus den:

- Handlungsalternativen beziehungsweise Entscheidungen,
- Umweltzuständen,
- Ergebnissen für die Kombinationen aus Umweltzuständen und Entscheidungen (Ergebnismatrix).

Liegen mindestens zwei sich einander ausschließende Handlungsalternativen vor, kann von einem Entscheidungsproblem gesprochen werden. Die Basiselemente des Grundmodells sind in Bild 1.1 zusammengestellt.

Das folgende Beispiel beschreibt ein einfaches Entscheidungsproblem. Ein Entscheidungsträger steht vor den Handlungsalternativen für einen Einsatz E an einem Glücksspiel teilzunehmen e_1 oder nicht e_2 . Es wird eine Kugel aus einer Urne mit roten und schwarzen Kugeln gezogen. Er gewinnt den Betrag G , wenn eine rote Kugel gezogen wird. Er gewinnt nichts, wenn eine schwarze Kugel gezogen wird. Die Zielfunktion des Entscheidungsträgers sei es, seinen Geldbestand zu maximieren. Die Umweltzustände sind dabei Gewinn z_1 (es wird eine rote Kugel gezogen) oder kein Gewinn z_2 (es wird eine schwarze Kugel gezogen). Die Ergebnisse sind die Änderungen des Geldbestandes des Entscheidungsträgers. Nimmt er nicht an dem Glücksspiel teil, ändert sich sein Geldbestand unabhängig vom Umweltzustand nicht. Nimmt er teil, so erhöht sich sein Geldbestand im Falle eines Gewinns um $G - E > 0$. Gewinnt er nicht, so verringert sich sein Geldbestand um E . Tabelle 1.1 zeigt die Ergebnismatrix für dieses Beispiel.

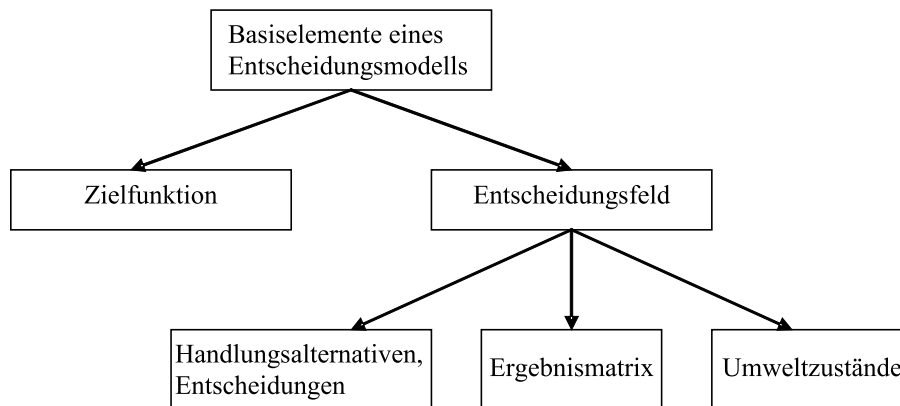


Bild 1.1: Basiselemente eines Entscheidungsmodells (Grundmodell der Entscheidung) [88].

	z_1	z_2
e_1	$G - E$	$-E$
e_2	0	0

Tabelle 1.1: Ergebnismatrix am Beispiel des Glücksspiels.

Ist dem Entscheidungsträger der Umweltzustand bekannt, d.h. die Ziehung wurde bereits durchgeführt und kann er seine Entscheidung noch immer treffen, so ist diese trivial. Es handelt sich dann um eine Entscheidung bei Sicherheit. Das Bestimmen der besten Handlungsalternative entspricht dann einem Optimierungsproblem über alle Handlungsalternativen für den gegebenen Umweltzustand. Ist der Umweltzustand nicht bekannt, liegt eine Entscheidung bei Unsicherheit vor. Es ist weiterhin zu unterscheiden, ob der Entscheidungsträger Informationen über die Anzahl der Kugeln und ihre Farben hat oder nicht. Es handelt sich dann um eine Entscheidung bei Risiko oder Ungewissheit (Bild 1.2). Die vorliegenden Informationen bezüglich der Umweltzustände sind Bestandteil des Grundmodells.

Unsicherheit, Ungewissheit, Unschärfe und Risiko

Da unterschiedliche Definitionen der Begriffe Unsicherheit, Ungewissheit und Unschärfe im Schrifttum existieren, werden die Begriffe für die Verwendung in dieser Arbeit erläutert. Dabei werden die Begriffe Unsicherheit und Ungewissheit zunächst voneinander abgegrenzt und die Unterschiede zu dem Begriff Unschärfe aufgezeigt.

Mit der Thematik des logischen Schließens unter Ungewissheit waren bereits LEIBNIZ (1646-1716), JAKOB BERNOULLI (1654-1705) und andere Zeitgenossen beschäftigt. Das Thema hat aber bis heute nicht an Aktualität verloren, wie zahlreiche Veröffentlichungen und Bücher zu dem Thema zeigen [54, 77, 88, 150]. Ein weit verbreiteter Ansatz zur

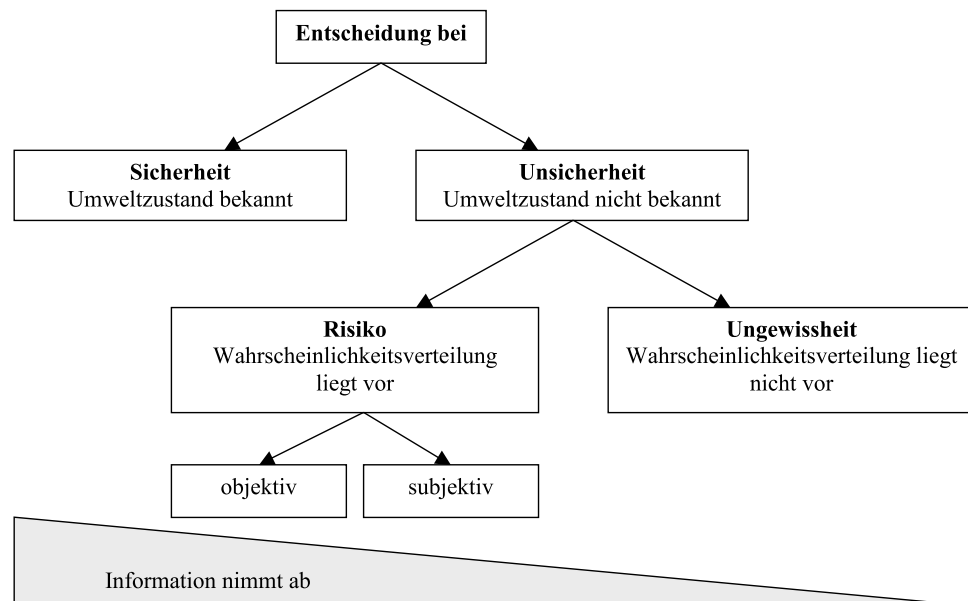


Bild 1.2: Sicherheit, Unsicherheit, Ungewissheit und Risiko. Nach [88] (erweitert).

Modellierung von Ungewissheit ist die Wahrscheinlichkeitstheorie. Als maßgebliche Begründer dieser Theorie gelten PASCAL, FERMAT [131] und HUYGENS [71]. Es wird dabei auf einen Briefwechsel zwischen FERMAT und PASCAL aus dem Jahr 1654 verwiesen. Jedoch gab es bereits im 14. Jahrhundert erste Versicherungen, vornehmlich für Schiffe, die ihre Prämien auf der Basis von Risiken festlegten. Die Wahrscheinlichkeitstheorie wurde ursprünglich für die Untersuchung von Glücksspielen und somit für die Untersuchung von Risikosituationen entwickelt.

GERKE unterscheidet in Abhängigkeit von der verwendeten Wahrscheinlichkeitsverteilung der Umweltzustände zwischen Entscheidung bei Risiko oder bei Unsicherheit [62]. Risiko liegt dann vor, wenn auf der Basis von objektiven Wahrscheinlichkeiten entschieden wird. Handelt es sich hingegen um subjektive Wahrscheinlichkeiten, wird von Unsicherheit gesprochen. Die objektive Wahrscheinlichkeit eines Ereignisses entspricht dem Grenzwert der relativen Häufigkeit dieses Ereignisses in einer Stichprobe, wenn die Stichprobengröße gegen unendlich strebt (Wahrscheinlichkeit für Kopf beim Werfen einer idealen Münze). Die Definition subjektiver Wahrscheinlichkeiten geht auf SAVAGE zurück [128]. Sie entsprechen subjektiv wahrgenommener Wettchancen und beruhen somit auf Schätzungen oder dem Glauben, dass ein Ereignis eintritt (Pferd A gewinnt das Rennen).

LAUX setzt für eine Entscheidung unter Risiko ebenfalls die Kenntnis einer Wahrscheinlichkeitsverteilung voraus, ob es sich dabei um eine subjektive oder objektive Verteilung handelt, wird nicht unterschieden [88]. Ist keine Aussage über die Wahrscheinlichkeiten der Umweltzustände möglich, handelt es sich um Unsicherheit im engeren Sinne. Der Begriff Unsicherheit wird als Oberbegriff für Unsicherheit im engeren Sinne und Risiko verwendet.

Bei REUCHER liegt wie bei LAUX Unsicherheit dann vor, wenn die vorhandenen Informationen nicht ausreichen, um den tatsächlichen Umweltzustand zu bestimmen [124]. Dies ist bei einem Zufallsexperiment stets der Fall. Ist zusätzlich zur Unsicherheit über den Umweltzustand die Information über den Zufallsmechanismus – die dem Experiment zu Grunde liegende Wahrscheinlichkeitsverteilung – unvollständig, liegt Ungewissheit vor. Die Ungewissheit resultiert aus der Größe des Zulässigkeitsraumes \mathcal{F} für die Wahrscheinlichkeitsverteilung der Ereignisse des Experiments.

Als Beispiel wird ein sechsseitiger Würfel betrachtet. Liegt die Information vor, dass es sich um einen idealen Würfel handelt, so besteht \mathcal{F} lediglich aus der Gleichverteilung $P(w) = \frac{1}{6}$ mit $w = 1, 2, \dots, 6$. Aus $|\mathcal{F}| = 1$ folgt, dass keine Ungewissheit vorliegt. Ohne die Information, dass es sich um einen idealen Würfel handelt, sind beliebig viele Verteilungen der Wahrscheinlichkeiten der Augenzahlen des Würfels möglich $|\mathcal{F}| = \infty$. Es liegt demnach Ungewissheit über die Wahrscheinlichkeiten vor.

In dieser Arbeit werden die Begriffe Unsicherheit und Ungewissheit wie folgt verwendet: Unsicherheit resultiert aus der Existenz einer Wahrscheinlichkeitsverteilung der Umweltzustände. Entscheidung bei Ungewissheit liegt vor, wenn diese Wahrscheinlichkeitsverteilung aufgrund unvollständiger Information nicht eindeutig formuliert ist. Wenn die Wahrscheinlichkeitsverteilung der Umweltzustände als objektive Wahrscheinlichkeiten oder subjektive Wahrscheinlichkeiten (zum Beispiel datenbasierte Schätzungen) gegeben ist, wird von Entscheidung unter Risiko gesprochen. Bild 1.2 stellt diese Zusammenhänge grafisch dar.

Im Gegensatz zu den Begriffen Unsicherheit und Ungewissheit bezieht sich der Begriff Unschärfe auf die Darstellung von Messwerten. Dabei wird das Ergebnis der Messung in einem Intervall angegeben z.B. $l = 1 \pm 0.01 [m]$ [87]. Die Unschärfe bezieht sich dann auf die Breite des Intervalls, in dem der tatsächliche Wert liegt. Durch genaueres (aufwändigeres) Messen lässt sich die Unschärfe gemäß der Unschärferelation nach HEISENBERG bis zu einer unteren Schranke verringern. Die Unschärfe entsteht nicht zufällig, sondern durch die Art der Bestimmung des Messwertes.

Die Breite des Intervalls kann zur Zusammenfassung von Bereichen oder zur Abstraktion bewusst gewählt sein. Wenn eine höhere Genauigkeit nicht notwendig ist, kann die Einteilung 'groß', 'mittel', 'klein' Vorteile bei der Interpretation von Messwerten wie zum Beispiel die Größe von Personen bieten. Die Wahrscheinlichkeit für einen Umweltzustand oder das Ergebnis einer Handlungsalternative können unscharf formuliert werden. Zur weiteren Abgrenzung zwischen Unschärfe und Unsicherheit sei auf [53] verwiesen.

Entscheidungsregel und Entscheidungsprinzip

In Verbindung mit den Informationen aus dem Grundmodell kann mittels einer Entscheidungsregel eine Handlungsempfehlung abgeleitet werden. Eine Entscheidungsregel beschreibt eindeutig, wie aus den vorliegenden Informationen die gemessen an der Zielfunktion beste Handlungsalternative bestimmt wird. Eine Entscheidungsregel kann beispielsweise vorschreiben, die Handlungsalternative mit dem größten oder dem kleinsten Wert einer Kennzahl zu wählen. Ein Entscheidungsprinzip gibt dagegen lediglich Richtlinien, wie die vorliegenden Informationen zu Kennzahlen der einzelnen Handlungsalternativen aggregiert werden sollen. Ein Entscheidungsprinzip liefert keine eindeutige Entscheidungsregel [88]. Das μ - σ -Prinzip gibt beispielsweise vor, dass der Erwartungswert μ und die Standardab-

	z_1	z_2	MinMax
e_1	$G - E$	$-E$	$-E$
e_2	0	0	0

Tabelle 1.2: Ergebnismatrix am Beispiel des Glücksspiels mit MinMax-Regel.

weichung σ für die Entscheidungsregel verwendet werden sollen. Eine konkrete Ausgestaltung wird nicht vorgeschrieben. Beim Beispiel des Glücksspiels ist die Zielfunktion die Maximierung des Gewinns. Liegen Informationen über die Zusammensetzung der Kugeln in der Urne vor, kann über die μ -Regel der Erwartungswert der beiden Handlungsalternativen bestimmt werden. Die Entscheidungsregel schreibt dann vor, die Alternative zu wählen, deren Erwartungswert am größten ist. Im Folgenden werden einige wichtige Entscheidungskriterien vorgestellt und diskutiert. Der Begriff Entscheidungskriterium entspricht entweder einer Entscheidungsregel oder einem Entscheidungsprinzip [88].

Die MinMax-Regel ist für alle Entscheidungen bei Ungewissheit und Risiko geeignet, da sie Informationen über die Wahrscheinlichkeitsverteilung der Zustände ignoriert. Die μ -Regel und das μ - σ -Prinzip eignen sich nur für Entscheidungen unter Risiko, da Wahrscheinlichkeiten zur Berechnung der Kennzahlen erforderlich sind. Das Bernoulli-Prinzip geht auf das Nutzenkonzept von DANIEL BERNOULLI zurück und kann mit den erwähnten Kriterien kombiniert werden. Die Kriterien sind in der Literatur zur Entscheidungstheorie ausführlich dargestellt [28, 83, 88, 95].

MinMax-Regel

Die MinMax-Regel erfordert keine Informationen über die Wahrscheinlichkeitsverteilung der Umweltzustände. Als Kennzahl für eine Handlungsalternative wird das gemessen an der Zielfunktion schlechteste mögliche Ergebnis herangezogen. Die MinMax-Regel schreibt dann vor, die Handlungsalternative mit der gemessen an der Zielfunktion besten Kennzahl zu wählen. Für das Beispiel des Glücksspiels aus Abschnitt 1.2.1 ergibt sich stets 'nicht teilnehmen' als beste Entscheidung (Tabelle 1.2 und Bild 1.3 links). Dies ist auch dann der Fall, wenn die Information vorliegt, dass sich in der Urne 99 rote und eine schwarze Kugel befinden und der Gewinn den Einsatz weit übersteigt.

μ -Regel

Die μ -Regel setzt eine Risikosituation voraus. Als Kennzahl für die Handlungsalternativen wird der Erwartungswert des Ergebnisses berechnet. Nach der μ -Regel zu entscheiden bedeutet, die Handlungsalternative auszuwählen, deren Erwartungswert des Ergebnisses gemessen an der Zielfunktion über alle möglichen Handlungsalternativen maximal ist. Die Kennzahlen der Handlungsalternativen des Entscheidungsträgers beim Glücksspiel ändern sich durch die Berücksichtigung der Wahrscheinlichkeitsverteilung. In Abhängigkeit vom Verhältnis der schwarzen und roten Kugeln sowie dem zu zahlenden Einsatz und dem möglichen Gewinn kann der Erwartungswert des Ergebnisses positiv sein. Die reine μ -Regel setzt Risikoneutralität voraus. Risikoneutralität bedeutet hier Neutralität gegenüber Schwankungen der Ergebnisgrößen für die gewählte Handlungsalternative. Bild 1.3 links

zeigt zwei Varianten des Glücksspiels mit identischem Erwartungswert, aber unterschiedlichen Ergebnisgrößen. Zur Berechnung der Schwankung ist die Verwendung der Standardabweichung weit verbreitet, obwohl dies nur bei Normalverteilung der Ergebnisse zulässig ist [28, 88, 95].

μ - σ -Prinzip

Im Gegensatz zur μ -Regel, die von Risikoneutralität des Entscheidungsträgers ausgeht, wird beim μ - σ -Kriterium die Haltung gegenüber Risiko in die Berechnung der Kennzahl über die Standardabweichung σ miteinbezogen. Es wird Risikoaversion und Risikofreude unterschieden. Erstere entspricht dem Wunsch, Risiken zu vermeiden, während letztere ein bewusstes Eingehen von Risiken beschreibt. Für Risikoaversion wird die Kennzahl gemessen an der Zielfunktion mit steigender Standardabweichung schlechter, für Risikofreude entsprechend besser. Dies wird zum Beispiel durch Linearkombination von μ und σ mit geeigneten Koeffizienten erreicht. Bild 1.3 links zeigt für das Glücksspiel die Auswirkungen der Standardabweichung auf die Kennzahlen. Der risikofreudige Entscheidungsträger wird die Variante 2 wählen, der risikoscheue (Risikoaversion) die Variante 1. Bei Risikoneutralität gibt es keine Differenz zwischen beiden Varianten. Bild 1.3 rechts zeigt allgemein für das μ - σ -Prinzip den Zusammenhang zwischen Risikoneutralität, Risikofreude (RF) und Risikoaversion (RA) grafisch.

Variante	1	2
Einsatz E	100	100
Gewinn G	1000	10000
Rote Kugeln	10	1
Schwarze Kugeln	90	99
MinMax	-100	-100
μ	0	0
σ	300	995
RA: $\mu - \alpha \cdot \sigma$	-150	-497.5
RF: $\mu + \alpha \cdot \sigma$	150	497.5
keine Teilnahme	0	0

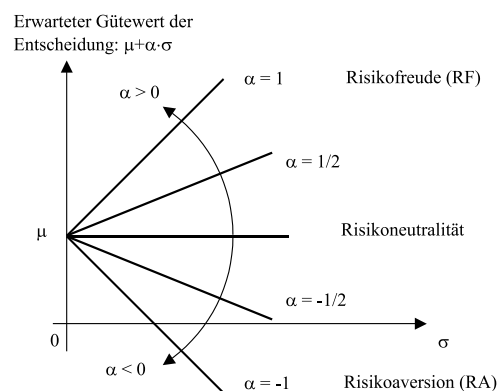


Bild 1.3: Links: Varianten des Glücksspiels, Kennzahlen der Entscheidung 'teilnehmen' für die MinMax-, μ -Regel und das μ - σ -Prinzip $\alpha = 0.5$. Rechts: Bewertung von Entscheidungen mit gleichem Mittelwert und unterschiedlichen Standardabweichungen bei Risikoneutralität, Risikofreude und Risikoaversion.

Bernoulli-Prinzip

Das Bernoulli-Prinzip von DANIEL BERNOULLI basiert auf der Annahme, dass der Nutzen eines Ergebnisses nicht proportional zu dem Ergebnis ist. Am Beispiel des Glücksspiels bedeutet dies, dass der Nutzen des Entscheidungsträgers an einem Gewinn von 10000 nicht zehnmal so hoch ist, wie der Nutzen an einem Gewinn von 1000. Die möglichen Ergebnisse der Handlungsalternativen sind zunächst in Nutzenwerte umzurechnen. Anschließend

können die Kennzahlen für das verwendete Entscheidungskriterium auf der Basis der Nutzenwerte berechnet und die beste Entscheidung bestimmt werden.

Erweiterung des Grundmodells

Für eine Entscheidung bei Unsicherheit können zusätzliche Informationen, zum Beispiel Beobachtungen, Messwerte oder Merkmale, zu einer Entscheidung bei Risiko oder auch bei Sicherheit führen. Das Entscheidungsproblem ist dann zum Beispiel mit der μ -Regel oder als Optimierungsproblem lösbar. In dieser Arbeit wird der Begriff Beobachtung als Synonym für jede zusätzliche Information über den aktuellen Umweltzustand betrachtet. Eine Beobachtung kann also zum Beispiel ein Messwert, eine aus Messwerten abgeleitete Größe, aber auch die Information sein, dass der tatsächliche Wert in einem bestimmten Intervall liegt. Wenn die zusätzlichen Informationen Kosten verursachen, sind diese bei den Kennzahlen der Entscheidungen zu berücksichtigen. Im Weiteren werden diese Kosten als Merkmalskosten bezeichnet. Eine Übersicht über weitere Kostenarten, die bei Entscheidungsproblemen auftreten können, ist in [146] zu finden.

Wird die Anzahl der möglichen Umweltzustände und die Anzahl der möglichen Entscheidungen auf unendlich viele erweitert, handelt es sich um ein Regressionsproblem [6, 148]. In dieser Arbeit werden nur Entscheidungsprobleme mit abzählbar vielen Umweltzuständen und Entscheidungen betrachtet.

1.2.2 Evidenztheorie

Die Evidenztheorie ist eine Verallgemeinerung der Wahrscheinlichkeitstheorie und geht auf die Arbeiten von DEMPSTER und SHAFER zurück. Sie dient der Modellierung von Ungewissheit. Der Ursprung liegt dabei in DEMPSTERS Veröffentlichung zu oberen und unteren Wahrscheinlichkeiten (*upper and lower probabilities*) aus dem Jahr 1967 [47]. SHAFER hat im Jahr 1976 die Arbeiten von DEMPSTER und seine eigenen in dem Buch 'A mathematical theory of evidence' [138] zusammengetragen. Zahlreiche Bücher [64, 141, 144] und Veröffentlichungen [75, 115, 116, 154, 161] sind zur Evidenztheorie und ihrer Anwendung erschienen. Eine detaillierte Zusammenfassung der Grundlagen findet sich in Anhang 6.1.

Die Evidenztheorie drückt Unterstützung für das Eintreten von Ereignissen wie die Wahrscheinlichkeitstheorie mit Zahlen zwischen Null und Eins aus [138]. Zum Beispiel kann dem Ereignis A : 'Morgen wird es regnen' eine Unterstützung (Masse m) von 0.5 zugewiesen werden. Gleichzeitig kann dem Gegenteil \bar{A} : 'Morgen wird es nicht regnen' eine Unterstützung von 0.2 zugewiesen sein. Der wesentliche Unterschied zur Wahrscheinlichkeitstheorie liegt darin, dass die Masse nicht nur den disjunkten Ereignissen (hier A und \bar{A}) einer Menge von Ereignissen zugewiesen werden darf, sondern allen Elementen der Potenzmenge dieser Menge $\{\emptyset, A, \bar{A}, A \text{ oder } \bar{A}\}$ außer der leeren Menge \emptyset . Bei obigem Beispiel bleibt noch eine Masse von 0.3 übrig. Diese ist der Verknüpfung 'A oder \bar{A} ' zuzuweisen. Aus der Verteilung der Unterstützung – diese wird auch Massenfunktion genannt – wird ein Glaubensgrad (*degree of belief*) beziehungsweise ein Plausibilitätsgrad (*degree of plausibility*) für alle Ereignisse der Potenzmenge abgeleitet. Im Glaubensgrad (*bel*) eines Ereignisses werden die Massen der Elemente der Potenzmenge summiert, die eine Teilmenge des untersuchten Ereignisses sind. Er kann als untere Abschätzung der Wahrscheinlichkeit dieses Ereignisses interpretiert werden. Im Falle des Ereignisses A be-

Wetterbericht	1	2	$1 \oplus 2$
$m(A)$	0.5	0.1	0.42
$m(\bar{A})$	0.2	0.4	0.39
$m(A \cup \bar{A})$	0.3	0.5	0.19
Widerspruch			0.22

Tabelle 1.3: Beispiel einer Kombination zweier Massefunktionen mittels Dempster's rule of combination.

steht der Glaubensgrad lediglich aus der Masse von A . Im Plausibilitätsgrad (pl) werden die Massen all jener Elemente der Potenzmenge summiert, die eine nicht leere Schnittmenge mit dem untersuchten Ereignis haben. Für den Plausibilitätsgrad des Ereignisses A ist also die Masse des Ereignisses A sowie die Masse des Ereignisses ' A oder \bar{A} ' zu addieren. Er gestattet eine Interpretation als obere Wahrscheinlichkeit dieses Ereignisses. Daraus ergibt sich, dass der Glaubensgrad eines Ereignisses stets kleiner als oder gleich dem Plausibilitätsgrad ist. Die Glaubensgrade und Plausibilitätsgrade über der betrachteten Menge von Ereignissen werden als Glaubensfunktion (*belief function*) und Plausibilitätsfunktion (*plausibility function*) bezeichnet. Bild 1.4 zeigt die Zusammenhänge zwischen Masse-, Glaubens- und Plausibilitätsfunktion nochmals grafisch.

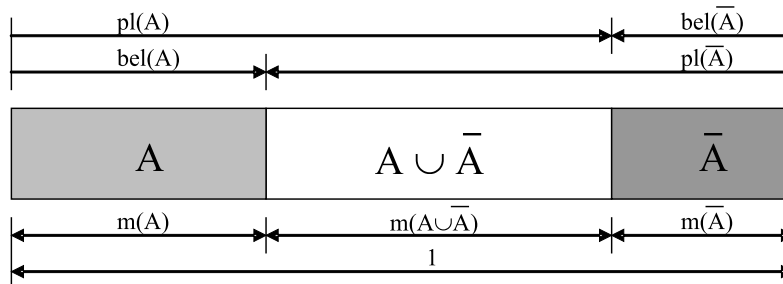


Bild 1.4: Zusammenhang Massefunktion, Glaubens- und Plausibilitätsgrad.

Ungewissheit über eine Wahrscheinlichkeitsverteilung von Ereignissen kann über die Zuweisung von Masse zu Kombinationen aus zwei oder mehr Ereignissen erfolgen. Für obiges Beispiel ergibt sich ein Intervall für die Wahrscheinlichkeit für Regen von $[0.5 \dots 0.8]$. Die Breite des Intervalls kann als Ungewissheit interpretiert werden.

Die Verknüpfung einzelner Massefunktionen, die sich auf die selbe Ereignismenge beziehen, aber aus unterschiedlichen Quellen stammen (zum Beispiel zwei Wetterberichte), ist mittels Dempster's rule of combination möglich [138] (siehe Abschnitt 2.3.4). Nach der Anwendung der Regel liegt eine neue Massefunktion und ein Maß für die Widersprüchlichkeit der beiden Quellen vor. Tabelle 1.3 zeigt die Kombination mit der Massefunktion aus einem weiteren Wetterbericht. Nach der Kombination hat sich zum Beispiel das Intervall der Wahrscheinlichkeit für Regen auf $[0.42 \dots 0.61]$ geändert. Die Ungewissheit wird durch zusätzliche Information stets kleiner.

1.2.3 Entscheidungssysteme

Die Aufgabe eines Entscheidungssystems ist es, aus einer Menge von Eingangsgrößen anhand eines Modells selbsttätig eine Entscheidung zu treffen oder eine Handlungsempfehlung zu geben.

Zunächst werden einige Merkmale diskutiert, anhand derer Entscheidungssysteme eingeteilt werden können. Anschließend erfolgt eine schematische Darstellung des Entwurfsprozesses von Entscheidungssystemen.

Unterscheidungsmerkmale

Ein Unterscheidungsmerkmal ist der Umgang mit der Antwort des Systems auf die aktuellen Eingangsgrößen. Leitet das System selbst in Abhängigkeit von seiner Entscheidung eine bestimmte Handlung ein, oder dient die Entscheidung des Systems als Empfehlung für eine weitere Instanz, die diese Empfehlung wiederum in ihre eigene Entscheidung einbindet. Der erste Fall liegt vor, wenn ein Diagnosesystem selbsttätig einen Prozess beeinflusst, der zweite Fall ist zum Beispiel bei medizinischen Entscheidungssystemen anzutreffen. Das Entscheidungssystem schlägt eine Diagnose vor, die der Arzt bestätigt, zurückweist oder ändert.

Entscheidungssysteme lassen sich auch danach unterscheiden, ob sie für Entscheidungen verwendet werden, die genau einmal getroffen werden, oder für Entscheidungen, die wiederholt zu treffen sind. Die Entscheidung über ein bestimmtes Investitionsprojekt eines Unternehmens beziehungsweise die Fehlerüberwachung eines Prozesses sind Beispiele für einmalige und mehrfache Entscheidungen. Zwar trifft ein Unternehmen in der Regel nicht nur eine Investitionsentscheidung, jedoch unterscheiden sich die einzelnen Situationen. Das Modell, auf dessen Basis eine Entscheidung getroffen wird, ist daher stets anzupassen.

Weiterhin können Entscheidungssysteme danach unterschieden werden, ob sie auf der Basis vollständiger Information, dies entspricht Entscheidungen bei Sicherheit, oder unvollständiger Information, dies entspricht Entscheidungen bei Unsicherheit (Ungewissheit oder Risiko), entscheiden (Bild 1.2). Die Gefahr einer Fehlentscheidung existiert bei vollständiger Information nicht. Liegt bezüglich des aktuellen Umweltzustandes unvollständige Information vor, sind Fehlentscheidungen möglich.

Weiterhin ist zu unterscheiden, ob die Ergebnisse der einzelnen Kombinationen aus Umweltzuständen und Entscheidungen bestimmt sind oder nicht. Zusätzlich können die Ergebnisse konstant sein oder einem zeitlichen Verlauf unterliegen.

In dieser Arbeit werden Entscheidungsprobleme betrachtet, bei denen die Ergebnisse bekannt und konstant sind, jedoch unvollständige Information bezüglich des Umweltzustandes vorliegt. Für die Verwendung der Ausgangsgröße wird keine Einschränkung vorgenommen.

Entwurf

Ziel des Entwurfs eines Entscheidungssystems ist es, ein Modell des betrachteten Prozesses zu entwerfen, das die Situation unvollständiger Information – Unsicherheit bezüglich des tatsächlichen Umweltzustandes – durch die Verarbeitung von Eingangsgrößen in eine Situation vollständiger Information – Sicherheit bezüglich des tatsächlichen Umweltzustandes – überführt. In Abhängigkeit von der Qualität der Eingangsgrößen ist dieses Ziel

vollständig oder nur teilweise erreichbar. Reichen die gegebenen Eingangsgrößen für die eindeutige Bestimmung des Umweltzustandes nicht aus, so ist die Entscheidung auf der Basis einer Schätzung der Wahrscheinlichkeiten der Umweltzustände für die gegebenen Eingangsgrößen zu treffen.

Entscheidungssysteme können manuell auf der Basis von Expertenwissen zu dem betrachteten Prozess entworfen werden. Bei vielen Prozessen übersteigt der Aufwand zur exakten und vollständigen physikalischen, chemischen, biologischen, soziologischen oder ökonomischen Modellierung den Rahmen des Möglichen. An dieser Stelle bieten datenbasierte Verfahren eine alternative Modellbildung. Datenbasierte Verfahren werden in überwachte Verfahren und nicht überwachte Verfahren eingeteilt. Bei den überwachten Verfahren stehen die tatsächlichen Umweltzustände während des Entwurfsprozesses zur Verfügung, bei nicht überwachten Verfahren entsprechend nicht. Zu den nicht überwachten Verfahren gehört zum Beispiel das Cluster-Verfahren [70]. Eine Kombination aus theoretischer (mit Expertenwissen) und experimenteller (datenbasiert) Prozessanalyse ist ebenfalls möglich. Die Wahl der Modellstruktur kann beispielsweise auf Expertenwissen basieren, während die Bestimmung der Modellparameter datenbasiert erfolgt [152]. Eine wichtige Anforderung an datenbasierte Entwurfsverfahren ist die Interpretierbarkeit der entworfenen Entscheidungssysteme. Die Validierung eines Entscheidungssystems wird erleichtert, wenn die grundlegenden Zusammenhänge für den Anwender nachvollziehbar sind.

Als Eingangsgrößen für den überwachten Entwurfsprozess eines Entscheidungssystems sind ein Lerndatensatz, eine algebraisch formulierte Zielfunktion und für eine entscheidungstheoretische Bewertung eine Ergebnismatrix sowie eine Entscheidungsregel im Sinne von Abschnitt 1.2.1 zu bestimmen. Im Folgenden wird stets von Kostenmatrizen ausgegangen. Der Lerndatensatz besteht aus den Eingangsgrößen zu einzelnen Ausprägungen des Entscheidungsproblems, im weiteren Beispiele genannt, und dem tatsächlichen Umweltzustand. Handelt es sich zum Beispiel um ein Problem der Fehlererkennung, können die Eingangsgrößen Messwerte oder daraus abgeleitete Größen zu einer Situation des Prozesses sein. Die zugeordneten Umweltzustände sind dann zum Beispiel 'Prozess in Ordnung' oder 'Prozess fehlerhaft'. Aus der beim Entwurfsprozess zu berücksichtigenden Zielfunktion und der Entscheidungsregel ist ein Bewertungsmaß für das entworfene Entscheidungssystem abzuleiten. In Bild 1.5 sind die Eingangsgrößen für den Entwurfsprozess auf der ersten Ebene dargestellt. Auf der folgenden Ebene wird das Entscheidungssystem entworfen. Nach dem Entwurfsprozess liegt ein Entscheidungssystem vor, das jedem Beispiel eine Entscheidung zuweist.

Für die Modellierung stehen eine Vielzahl von Verfahren zur Verfügung. Statistische Verfahren [148], Fuzzy-Verfahren [126], Künstliche Neuronale Netze [36], Entscheidungsbäume [33] und Cluster-Verfahren [70] seien hier als Beispiele genannt. Bei vielen dieser Verfahren besteht nicht die Möglichkeit, unterschiedliche Konsequenzen der Kombinationen aus Umweltzuständen und Entscheidungen zu berücksichtigen.

Teilweise wurde für die genannten Verfahren die Anwendung entscheidungstheoretischer Maße für beliebige Kostenmatrizen bereits untersucht [21, 55, 98, 130, 145]. Andere Ansätze versuchen, die Kosten für Fehlentscheidungen unabhängig vom verwendeten Verfahren durch Manipulation der Wahrscheinlichkeiten zu berücksichtigen. Dazu gehört die Anpassung der Häufigkeiten der einzelnen Umweltzustände im Lerndatensatz [33, 55] durch Reduzieren von Beispielen 'günstiger' Zustände oder Vervielfältigung von Beispielen 'teurer' Zustände. Beim Metakosten-Ansatz [51] wird zunächst die Wahrscheinlichkeit

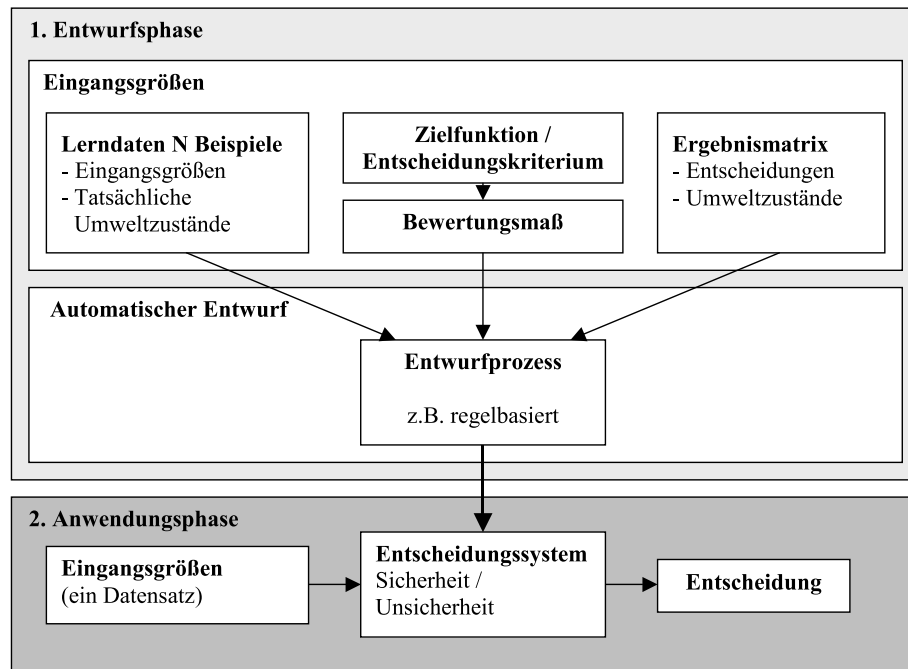


Bild 1.5: Datenbasierter überwachter Entwurf von Entscheidungssystemen.

der Umweltzustände für jedes Beispiel berechnet und darauf basierend die beste Entscheidung in Abhängigkeit von den Kosten bestimmt. Anschließend werden diese Entscheidungen den Lerndaten als tatsächliche Zustände zugewiesen und das Entscheidungssystem wird mit dem neuen Datensatz ohne Berücksichtigung von Kosten entworfen. Für den Entwurf von Fuzzy-Systemen ist bisher noch kein geschlossenes Verfahren zur entscheidungstheoretischen Bewertung vorgeschlagen worden.

Für die Berücksichtigung der Kosten von Fehlentscheidungen sind somit zwei Ansätze zu unterscheiden. Beim ersten werden mittels der gegebenen Daten Wahrscheinlichkeiten für die Umweltzustände geschätzt und anschließend über eine entscheidungstheoretische Bewertung eine Entscheidung getroffen. Beim zweiten Ansatz wird die entscheidungstheoretische Bewertung in den Entwurfsprozess zurückgekoppelt. Über die Rückkopplung werden das Datenmaterial oder die Beobachtungen verändert. Das Ziel ist die Schätzung der Umweltzustände für die Beobachtungen so zu verändern, dass die Kosten je Entscheidung sinken.

Die Merkmalskosten werden nur selten bei den Ergebnissen der Entscheidungen berücksichtigt und in den Entwurfsprozess rückgekoppelt [2, 114, 121, 145]. Die Kosten für Merkmale oder Tests lassen sich in fixe Kosten (Anschaffung eines Sensors) und variable Kosten (Durchführen eines Tests) aufteilen [56]. Die Summe dieser Kosten bezogen auf die einzelne Entscheidung ist bei den Ergebnissen zu berücksichtigen. Die variablen Kosten können zum Beispiel dadurch gesenkt werden, dass nicht alle Informationen für jede Entscheidung benötigt werden. Bei Testreihenfolgeproblemen [2, 121] werden zur Be-

stimmung des Umweltzustandes die Tests sequentiell durchgeführt. Jeder Test liefert dabei zusätzliche Informationen über den tatsächlichen Umweltzustand und verursacht bestimmte Kosten. Der nächste Test wird dabei in Abhängigkeit von den aktuellen Informationen gewählt.

1.2.4 Fuzzy-Logik

Die grundlegende Idee der 1965 von ZADEH [159] begründeten Fuzzy-Logik (Theorie der unscharfen mengen) ist es, die graduelle Zugehörigkeit zwischen Null und Eins eines Elementes zu einer Menge zu erlauben. Dies erweitert die klassische Logik, die als Zugehörigkeiten nur Null oder Eins zulässt. Die Zugehörigkeit eines Elementes zu einer Menge ist nicht als Wahrscheinlichkeit zu interpretieren. Ist über den Inhalt eines Glases bekannt, dass er aus einer Mischung mit 70% Wasser und 30% Orangensaft besteht, so handelt es sich bei den Angaben um Zugehörigkeiten zu den Mengen Wasser und Orangensaft. Besteht der Inhalt jedoch mit einer Wahrscheinlichkeit von 0.7 aus reinem Wasser und mit 0.3 aus reinem Orangensaft, so handelt es sich um ein Glas, das entweder mit Wasser oder mit Orangensaft gefüllt ist. Zur Fuzzy-Logik und den darauf basierenden Theorien Fuzzy-Regelung (Fuzzy Control) und Fuzzy Decision Making sind zahlreiche Bücher erschienen [10, 30, 38, 79, 126].

Die Fuzzy-Logik leistet einen wesentlichen Beitrag zur Nachbildung menschlichen Verhaltens, das in Form von 'WENN Voraussetzung DANN Aktion'-Regeln verarbeitet wird [11, 30, 82]. Die Regeln sind keine Entscheidungsregeln im Sinne von Abschnitt 1.2.1 sondern Entscheidungsfunktionen, die Eingangsgrößen Entscheidungen zuweisen. Fuzzy-Regeln erweitern die Möglichkeiten der Modellierung von Entscheidungssystemen. Die Verarbeitung von Eingangsgrößen zu Ausgangsgrößen innerhalb eines regelbasierten Fuzzy-Systems erfolgt in den Schritten Fuzzifizierung, Inferenz und Defuzzifizierung [79, 99]. Bild 1.6 stellt diesen Ablauf grafisch dar. Im ersten Schritt werden gemessene Größen wie zum Beispiel eine Temperatur über Zugehörigkeiten zu Mengen, die durch linguistische Terme (warm, kalt oder klein, mittel, groß) repräsentiert sind, zugeordnet und somit in eine natürlichsprachliche Situationseinschätzung transferiert. Im zweiten Schritt wird auf der Basis des Regelwerkes eine natürlichsprachliche Handlungsanweisung generiert, die abschließend in eine konkrete Handlung übersetzt wird.

Dieser Ablauf kann sowohl für ein regelbasiertes Fuzzy-Entscheidungssystem mit diskreten Ausgängen 'Prozess in Ordnung', 'Prozess fehlerhaft' als auch für Fuzzy-Regler mit einer kontinuierlichen Stellgröße angewandt werden. Überlappen sich die Voraussetzungen von zwei oder mehreren Fuzzy-Regeln, können Widersprüche auftreten, wenn die den Regeln zugeordneten Aktionen unterschiedlich sind. Es ist die Aufgabe der Inferenzmaschine, solche Widersprüche aufzulösen [79]. Ein wesentlicher Vorteil von Fuzzy-Systemen ist deren einfachere Interpretierbarkeit gegenüber statistischen Verfahren. Die wesentlichen Zusammenhänge eines Fuzzy-Systems lassen sich aus den einzelnen Regeln ablesen.

Beim Fuzzy Decision Making liegt die Unschärfe nicht in den Entscheidungsfunktionen, sondern in der Formulierung der Ziele und der Restriktionen von Optimierungsproblemen. Erste Überlegungen zum Fuzzy Decision Making wurden 1970 von BELLMAN und ZADEH veröffentlicht [20]. Diese werden in [151] weitergeführt und in [25] auf Gruppenentscheidungen erweitert. Anwendungen des Fuzzy Decision Making sind zum Beispiel in [23, 35] zu finden. Gesucht wird die Entscheidung, die sowohl die unscharfe Zielfunk-

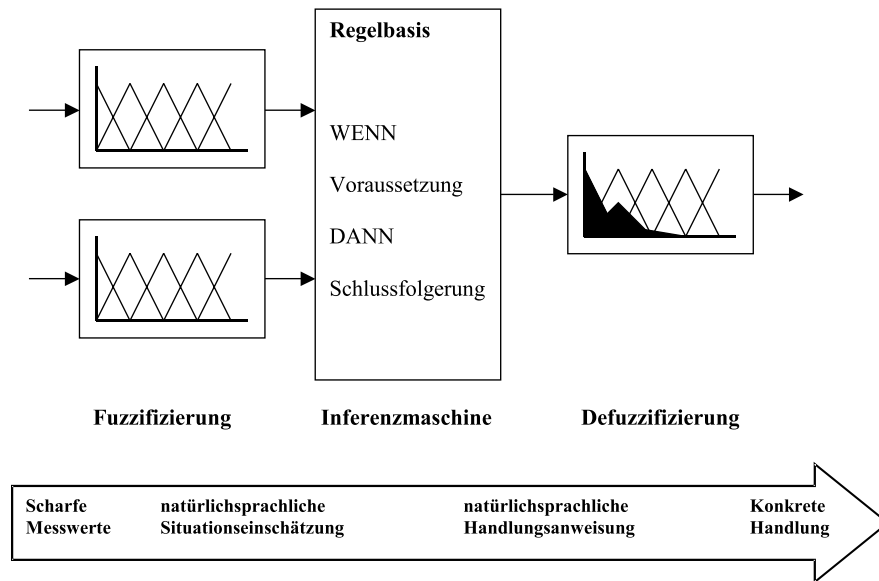


Bild 1.6: Grundprinzip eines Fuzzy-Systems aus [99].

tion als auch die unscharfen Restriktionen am besten erfüllt. Eine unscharfe Zielfunktion (Fuzzy-Goal) lautet zum Beispiel 'Die Kosten sollen nicht viel größer als C sein'. Eine unscharfe Restriktion (Fuzzy-Constraint) ist zum Beispiel: 'Es sollten nicht viel weniger Einheiten als N produziert werden'.

1.3 Offene Probleme

Es existiert eine gut entwickelte Theorie für statistische Entscheidungssysteme. Die Integration von Merkmalskosten in den Entwurfsprozess findet dagegen nur sporadisch statt. Eine Voraussetzung für den Entwurf von statistischen Entscheidungssystemen ist die Annahme spezieller Verteilungen der Daten. Die resultierenden Entscheidungssysteme sind meist nur schwer interpretierbar.

Mit den Fuzzy-Systemen steht aus Sicht der Interpretierbarkeit ein interessanter Ansatz zur Verfügung. Dieser ist aber unter dem Aspekt der entscheidungstheoretischen Bewertung noch unzureichend erforscht.

Häufig besteht die Möglichkeit, Entscheidungen, für die unzureichende oder widersprüchliche Informationen vorliegen, zurückzuweisen. Diese Entscheidungen werden dann zum Beispiel von einem Experten oder einer weiteren Ebene eines mehrstufigen Entscheidungssystems getroffen. Durch mehrstufige Entscheidungssysteme können zusätzlich Merkmalskosten eingespart werden. Für Fuzzy-Systeme, die Entscheidungen zurückweisen können beziehungsweise hierarchisch aufgebaut sind, existieren bisher keine automatisierten Entwurfsverfahren.

Für den automatischen Entwurf von Entscheidungssystemen bei unvollständiger Infor-

mation sind folgende Probleme ungelöst:

- Es existiert keine geschlossene Vorgehensweise zur Integration verschiedener Kostenarten (Entscheidungskosten, fixe und variable Merkmalskosten) in Entscheidungssysteme.
- Für Fuzzy-Systeme existiert keine entscheidungstheoretische Bewertung für den automatischen Entwurf von Entscheidungssystemen unter der Berücksichtigung beliebiger Zielfunktionen.
- Die Eignung der Evidenztheorie zur Auflösung von Konflikten und Rückweisung von Entscheidungen, die durch widersprüchliche Regeln eines Fuzzy-Systems auftreten, ist noch nicht hinreichend untersucht worden.
- Algorithmen zum automatischen Entwurf von hierarchischen Entscheidungssystemen mit dem Ziel der simultanen Minimierung der Entscheidungskosten und der Merkmalskosten sind noch nicht entwickelt.

1.4 Ziele der Arbeit

Ziel der Arbeit ist die Lösung der genannten vier offenen Probleme und damit die Erweiterung der Methoden zum automatischen Entwurf von Entscheidungssystemen.

Es wird stets von einem gegebenen Prozess ausgegangen, der durch ein automatisch entworfenes Entscheidungssystem auf der Basis von Fuzzy-Regeln bedient wird oder zu dessen Zustand eine Handlungsempfehlung beziehungsweise Entscheidung vorgeschlagen wird. Die Bedienung beziehungsweise die vorgeschlagenen Entscheidungen sollen dabei einer geforderten Genauigkeit folgen oder die Zielfunktion maximieren beziehungsweise minimieren. Als Nebenbedingung sollen möglichst geringe Kosten für die zur Bewertung der Prozesssituation verwendeten Merkmale entstehen. Bei Bedarf sollen Entscheidungssysteme entworfen werden, die in der Lage sind, bei widersprüchlichen Informationen keine Entscheidung zu treffen. Die Eignung der entworfenen Algorithmen soll durch Implementierung und Evaluierung anhand von Benchmarkdatensätzen und einem realitätsnahen Beispiel gezeigt werden.

Zur Umsetzung dieser Ziele werden die folgenden Teilprobleme gelöst:

1. Geeignete entscheidungstheoretische Bewertungsmaße für Fuzzy-Regeln und Fuzzy-Regelbasen werden identifiziert und für die jeweiligen Besonderheiten abgestimmt.
2. Die Merkmalskosten werden formalisiert und in die Bewertungsmaße integriert, um eine geschlossene Vorgehensweise bei der Berücksichtigung von Kosten zu ermöglichen.
3. Ein Entwurfsverfahren für Fuzzy-Entscheidungssysteme wird um die Bewertungsmaße erweitert. Ziel ist die Berücksichtigung der Entscheidungskosten sowie gegebenenfalls der Merkmalskosten im Entwurfsprozess.
4. Ein Inferenzverfahren für Fuzzy-Regeln wird vorgeschlagen, das auf der Evidenztheorie, insbesondere auf Dempster's rule of combination, basiert. Ziel ist das Erkennen widersprüchlicher Informationen und die quantitative Ausgabe der Widersprüchlichkeit und Unsicherheit der zugrunde liegenden Informationen.

5. Ein Algorithmus zum automatischen Entwurf von hierarchischen Entscheidungssystemen wird vorgestellt. Dabei wird für jede Ebene ein Entscheidungssystem entworfen. Ziel ist sowohl die Minimierung der Entscheidungskosten als auch der Merkmalskosten.
6. Die vorgestellten Algorithmen werden an Benchmarkdatensätzen und einem Anwendungsbeispiel aus der Robotik evaluiert.

Kapitel 2 befasst sich mit der Entwicklung neuartiger entscheidungstheoretischer Bewertungsmaße für Fuzzy-Einzelregeln und Regelbasen. In Kapitel 3 werden basierend auf den neuen Bewertungsmaßen modifizierte Verfahren vorgestellt, mit denen automatisch ein- und mehrstufige Fuzzy-Systeme zur Lösung von Entscheidungsproblemen erzeugt werden können. In Kapitel 4 sind die Ergebnisse der Anwendung der Maße und Verfahren auf Benchmarkdatensätze und ein Beispiel aus der Robotik zusammengestellt. Kapitel 5 fasst die Arbeit nochmals zusammen und gibt einen Ausblick auf mögliche zukünftige Entwicklungen.

2 Neuartige entscheidungstheoretische Bewertung von Fuzzy-Systemen

2.1 Überblick

In diesem Kapitel werden neuartige entscheidungstheoretische Bewertungsmaße für Fuzzy-Systeme vorgeschlagen. Der erste Abschnitt dient der Darstellung der erforderlichen Grundlagen der Entscheidungstheorie. Dabei wird auf verschiedene beim Entwurf von Entscheidungssystemen zu berücksichtigende Kostenarten eingegangen. Für Probleme mit zwei Umweltzuständen und zwei möglichen Entscheidungen wird eine neue Visualisierung vorgeschlagen, die die entscheidungstheoretische Bewertung verdeutlicht.

Ohne Beschränkung der Allgemeinheit werden die Ergebnisse von Entscheidungen im Weiteren stets als Kosten interpretiert. Ein Gewinn kann stets als negative Kosten dargestellt werden. Dadurch entfallen Überlegungen, ob die Kennzahl eines Entscheidungssystems zu maximieren oder minimieren ist. Ziel des Entwurfes eines Entscheidungssystems ist die Minimierung der Kosten pro Entscheidung. Im Weiteren werden die Begriffe Kostenmatrix, Entscheidungskosten, Merkmalskosten und Gesamtkosten je Entscheidung verwendet.

Der zweite Abschnitt beschreibt die betrachteten Fuzzy-Systeme, die auf WENN-DANN-Regeln basieren. Fuzzy-Entscheidungssysteme treffen auf der Basis ausgewählter Eingangsgrößen, Zugehörigkeitsfunktionen und Fuzzy-Regeln eine Entscheidung. Auf der Basis der Evidenztheorie wird für diese Fuzzy-Systeme ein neues Inferenzverfahren vorgeschlagen. Dieses Inferenzverfahren erlaubt es, bei unzureichenden oder widersprüchlichen Informationen keine Entscheidung zu treffen. Die weiteren Abschnitte erläutern schließlich die neuen entscheidungstheoretischen Bewertungsmaße für die betrachteten Fuzzy-Systeme. Die Bewertungsmaße erstrecken sich sowohl auf Fuzzy-Einzelregeln als auch auf Fuzzy-Regelbasen.

Zur Veranschaulichung der Effekte, die bei der Bewertung von regelbasierten Fuzzy-Systemen auftreten, werden gezielt gestaltete Beispiele verwendet.

2.2 Entscheidungstheorie

Dem Entscheidungssystem zu einem betrachteten Prozess steht für jede Entscheidung ein Eingangsgrößenvektor $\mathbf{x} \in \mathbb{R}_{N_x}$ zur Verfügung. Es kann sich dabei auch um einen durch neue Messwerte aktualisierten Vektor handeln. Das Entscheidungssystem verarbeitet Beobachtungen $b \in \mathcal{B}$, die aus dem Eingangsgrößenvektor \mathbf{x} durch Abbildungen H hervorgehen [7, 57]:

$$H : \mathbb{R}_{N_x} \mapsto \mathcal{B}. \quad (2.1)$$

Die Beobachtung b kann zum Beispiel eine Eingangsgröße x_l , $l = 1 \dots N_x$, ein Intervall einer Eingangsgröße oder die Prämisse einer WENN-DANN-Regel sein. Die Werte der Beobachtungen sind entweder kontinuierlich oder diskret.

Das Entscheidungssystem ist dann eine Funktion d , die jeder Beobachtung $b \in \mathcal{B}$ eine Entscheidung e_k aus der Menge der zulässigen Entscheidungen \mathcal{E} zuweist [7]:

$$d : \mathcal{B} \mapsto \mathcal{E} . \quad (2.2)$$

In Bild 2.1 ist eine einfache Entscheidungsfunktion dargestellt. Es sind zwei diskrete Beobachtungen b_1 und b_2 auf den Eingangsgrößen x_1 und x_2 definiert. Die Beobachtung b_1 ist durch die graue Fläche markiert und alle anderen Werte der Eingangsgrößen entsprechen der Beobachtung b_2 . Die Entscheidungsfunktion lautet:

$$d : \begin{cases} b_1 \mapsto e_1 \\ b_2 \mapsto e_2 \end{cases} . \quad (2.3)$$

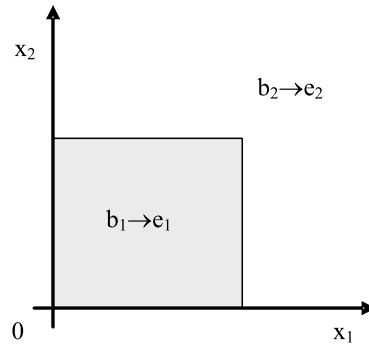


Bild 2.1: Beispiel einer Entscheidungsfunktion in \mathbb{R}_2 .

2.2.1 Entscheidungskosten

Entwurf von Kostenmatrizen

Eine wichtige Vorbereitung für den Einsatz eines entscheidungstheoretischen Verfahrens zum Entwurf eines Entscheidungsproblems ist der Entwurf der Kostenmatrix \mathbf{L} . In der Kostenmatrix werden die Kosten der N_e Entscheidungen beziehungsweise Handlungsalternativen e_k bei den N_z möglichen Umweltzuständen z_j erfasst und bewertet. In den Zeilen der Matrix stehen die einzelnen Entscheidungen, in den Spalten die Umweltzustände:

$$\mathbf{L} = \begin{pmatrix} L(e_1, z_1) & \cdots & L(e_1, z_j) & \cdots & L(e_1, z_{N_z}) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ L(e_k, z_1) & \cdots & L(e_k, z_j) & \cdots & L(e_k, z_{N_z}) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ L(e_{N_e}, z_1) & \cdots & L(e_{N_e}, z_j) & \cdots & L(e_{N_e}, z_{N_z}) \end{pmatrix} . \quad (2.4)$$

Die Kosten für die einzelnen Kombinationen aus Entscheidungen und Umweltzuständen können konstant, variabel oder Zufallsgrößen sein. Die tatsächlichen Kosten hängen bei variablen Elementen von einer weiteren Eingangsgröße q_{kj} über eine Funktion f ab:

$$L(e_k, z_j) = f(q_{kj}). \quad (2.5)$$

In [40, 69] werden Entscheidungssysteme für Kreditkartenbetrug beziehungsweise Betrug bei Telefonverbindungen untersucht. Die Kostenmatrix wird für jedes Beispiel in Abhängigkeit von den unterschiedlichen Kosten der Beispiele bestimmt. In [91] wird auf der Basis einer variablen Kostenmatrix die Lösung eines vereinfachten Zuordnungsproblems¹ vorgestellt. Die Elemente der Kostenmatrix hängen hier von der Zeit ab, die für die einzelnen Kombinationen aus Auftrag und Ausführendem verwendet wird. In [1] wird Unsicherheit über die Kosten von Fehlentscheidungen für Entscheidungsprobleme mit zwei Zuständen zugelassen.

Existieren zwei Handlungsalternativen, von denen die erste unabhängig vom Umweltzustand stets zum gleichen oder zu einem günstigeren Ergebnis führt als die zweite, so ist es nicht sinnvoll, die zweite Handlungsalternative zu wählen. Solche Dominanzbeziehungen zwischen den Handlungsalternativen führen dazu, dass dominierte Handlungsalternativen vom Entscheidungssystem nicht ausgewählt werden, da dann stets eine Handlungsalternative gewählt werden kann, die unabhängig vom Umweltzustand zum gleichen oder einem günstigeren Ergebnis führt. Eine Entscheidung e_s wird genau dann von einer anderen e_k dominiert, wenn es bei ansonsten identischen Kosten mindestens einen Umweltzustand gibt, für den die dominante Entscheidung weniger Kosten verursacht als die dominierte [28]:

$$e_k \succ e_s \mapsto \begin{array}{l} L(e_k, z_j) \leq L(e_s, z_j) \forall j \in \{1 \dots N_z\} \wedge \\ L(e_k, z_i) < L(e_s, z_i) \exists i \in \{1 \dots N_z\} \end{array} . \quad (2.6)$$

Eine Kostenmatrix soll so gestaltet sein, dass zwischen den Handlungsalternativen keine Dominanzbeziehungen bestehen.

Einfache quadratische Kostenmatrizen haben häufig Nullen auf der Hauptdiagonalen. Die anderen Elemente sind von Null verschieden. Für Klassifikationsprobleme wird häufig derselbe Ansatz verfolgt [1, 98, 130, 145]:

$$\mathbf{L} = \begin{pmatrix} 0 & \cdots & L(e_1, z_j) & \cdots & L(e_1, z_{N_z}) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ L(e_k, z_1) & \cdots & 0 & \cdots & L(e_k, z_{N_z}) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ L(e_{N_e}, z_1) & \cdots & L(e_{N_e}, z_j) & \cdots & 0 \end{pmatrix}, \quad N_e = N_j. \quad (2.7)$$

Dies ist nicht generell ein sinnvoller Ansatz. ELKAN [55] verweist auf die im Statlog-Projekt [98] verwendete Kostenmatrix für den so genannten 'German-Credit' Datensatz. Bei diesem Datensatz soll die Kreditwürdigkeit von Bankkunden vorhergesagt werden. Als Entscheidungen stehen die Möglichkeiten 'Kredit gewähren' oder 'Kredit ablehnen' den beiden Umweltzuständen 'Kunde zahlt zurück' beziehungsweise 'Kunde zahlt nicht zurück' gegenüber. Tabelle 2.1 zeigt die für das Klassifikationsproblem beim 'German-Credit' Datensatz in [98] verwendete Kostenmatrix.

¹Bei Zuordnungsproblemen ist eine Anzahl von N Jobs zu M Personen oder Maschinen kostenminimal zuzuordnen. Zuordnungsprobleme sind Optimierungsprobleme.

$L(e_k, z_j)$	Kunde zahlt zurück	Kunde zahlt nicht zurück
Kredit vergeben	0	5
Kredit ablehnen	1	0

Tabelle 2.1: Kostenmatrix für den 'German Credit' Datensatz aus [98].

$L(e_k, z_j)$	Kunde zahlt zurück	Kunde zahlt nicht zurück
Kredit vergeben	$-c \cdot r$	$c \cdot (1 + r_a)$
Kredit ablehnen	$-c \cdot r_a$	$-c \cdot r_a$

Tabelle 2.2: Alternative Kostenmatrix für den 'German Credit' Datensatz.

Kritisiert wird dabei, dass es in der Realität auf den Zahlungssaldo der Bank keinen Einfluss hat, ob ein Kunde den Kredit zurückzahlt oder nicht, wenn sein Kreditantrag abgelehnt wurde. Obige Kostenmatrix suggeriert hier jedoch einen Unterschied. ELKAN fordert daher, dass die Elemente einer Kostenmatrix relativ zu einer Basis zu bestimmen sind. Diese Basis kann zum Beispiel der Geldbestand vor der Entscheidung sein. Für obige Kostenmatrix bedeutet dies, dass die Elemente $L(e_2, z_1)$ und $L(e_1, z_1)$ identisch sein müssen. Wird der Kredit abgelehnt, kann zum Beispiel eine alternative Anlage des Kreditbetrages durch negative Kosten betrachtet werden.

Eine alternative Gestaltung der Kostenmatrix ist in der Tabelle 2.2 dargestellt. Es wird dabei von einer Kreditlaufzeit von einem Jahr und einer Rückzahlung der gesamten Summe plus Zinsen am Ende der Laufzeit ausgegangen. Dabei ist c der Kreditbetrag, r der Zinssatz, den der Kunde zahlt, und r_a der Zins einer alternativen Anlagemöglichkeit. Diese Kostenmatrix hängt lediglich linear von c ab. Dies ändert sich, wenn zusätzlich eine variable Kreditlaufzeit oder Bearbeitungskosten berücksichtigt werden. Beide Ergänzungen sind mathematisch einfach zu modellieren. Die Faktoren r_a und r werden zeitabhängig und zu jedem Element kommen die entsprechenden Bearbeitungskosten hinzu. Aufwändiger wenn nicht sogar unmöglich ist zum Beispiel die Modellierung eines potentiellen Imageschadens der Bank durch die Ablehnung eines 'guten' Kunden.

Eine exakte Bestimmung der Kostenmatrix für technische Prozesse ist nur dann möglich, wenn die Konsequenzen einer Fehlentscheidung bekannt und bewertbar sind. Dies ist zum Beispiel der Fall, wenn in der Qualitätskontrolle einer Serienfertigung ein intaktes Produkt als fehlerhaft deklariert und vernichtet wird. Hier entstehen Kosten in Höhe der Grenzkosten² der Produktion. Bei der Bewertung eines technischen Prozesses sind die Konsequenzen einer Fehlentscheidung nicht immer genau bekannt oder schwer zu bewerten. Wird ein Defekt in einem Kraftfahrzeug nicht erkannt und können deswegen weitere Schäden folgen, die zu einem Ausfall führen, so ist die Bewertung dieser Situation aufwändig. Bei der Bewertung ist auch der Ärger des Kunden zu berücksichtigen. Dieser führt aber erst mittelfristig zu schwer berechenbaren Einnahmeausfällen des Unternehmens (Verzicht auf zukünftige Käufe, negative Werbung im Bekanntenkreis usw.). Ist es für den betrachteten technischen Prozess nicht möglich, die Kosten exakt zu bestimmen, können auch die

²Die Grenzkosten sind die Kosten, die für die Erhöhung der Produktion um genau eine Einheit entstehen [153].

Verhältnisse der Konsequenzen unterschiedlicher Fehler zueinander als Kostenmatrix verwendet werden. In [111] wird eine solche Kostenmatrix für die Fehlerdiagnose in einem Verbrennungsmotor verwendet.

Ist eine Kostenmatrix für eine medizinische Anwendung zu erstellen, sind ebenfalls viele Faktoren zu berücksichtigen. Neben den Behandlungskosten spielen auch die Konsequenzen für den Patienten bei einer unterlassenen oder einer überflüssigen Behandlung eine Rolle. Besteht durch eine Fehlentscheidung das Risiko einer lebensbedrohlichen Situation, ergibt sich für die Bewertung zwangsläufig die Frage nach dem Wert des Lebens. Diese Frage soll an dieser Stelle jedoch nicht diskutiert werden. Grundsätzlich sollten Entscheidungssysteme im Bereich der Medizin zum Beispiel in der Diagnostik keine Handlungen einleiten, sondern lediglich den Entscheidungsträger durch Empfehlungen unterstützen. Dennoch sind, sofern Fehlentscheidungen unterschiedliche Folgen für den Patienten haben, die einzelnen Entscheidungen für die Umweltzustände zu bewerten. Hier kann ebenfalls von absoluten Beträgen zu einer relativen Bestimmung der Kosten oder einer qualitativen Bewertung übergegangen werden. Es geht dadurch der monetäre Bezug verloren, eine kardinale Ordnung unter verschiedenen Entscheidungssystemen ist aber dennoch möglich. Ein Verzicht auf die Bewertung mit Kosten führt hingegen zu einer Gleichbehandlung aller Fehlentscheidungen und ist somit die noch schlechtere Alternative.

Bewertungsmaße

Um unterschiedliche Entscheidungssysteme für ein gegebenes Problem zu bewerten, sind Bewertungsmaße notwendig, die eine kardinale Ordnung erlauben. Eine mögliche Bewertung ist der Erwartungswert der Kosten pro Entscheidung.

Eine Schätzung des Erwartungswerts der Kosten pro Entscheidung \hat{L}_D bei kontinuierlichen Zuständen und kontinuierlichen Beobachtungen ergibt sich zu [7]:

$$\hat{L}_D(d(b)) = \int_Z \int_{\mathcal{B}} L(z, d(b)) \cdot \hat{f}(b|z) \cdot \hat{\varphi}(z) db dz. \quad (2.8)$$

Dabei ist $\hat{f}(b|z)$ eine Schätzung der bedingten Verteilungsdichtefunktion der Beobachtungen b bei einem gegebenen Umweltzustand z . Diese Verteilung wird abhängig vom Entwurfsverfahren aus den Lerndaten geschätzt. Die Verteilungsdichtefunktion der Zustände $\hat{\varphi}(z)$ kann ebenfalls aus den Lerndaten geschätzt werden.

Beim Übergang zu N_z diskreten Umweltzuständen z_j und weiterhin kontinuierlichen Beobachtungen b geht das äußere Integral aus Gleichung (2.8) in eine Summe über:

$$\hat{L}_D(d(b)) = \sum_{j=1}^{N_z} \int_{\mathcal{B}} L(z_j, d(b)) \cdot \hat{f}(b|z_j) db \cdot \hat{P}(z_j). \quad (2.9)$$

Dabei bezeichnet $\hat{P}(z_j)$ die a-priori Wahrscheinlichkeit des Umweltzustandes z_j . Liegen nur N_b diskrete Beobachtungen b_i vor, folgt aus (2.9) eine Doppelsumme:

$$\hat{L}_D(d(b_i)) = \sum_{j=1}^{N_z} \sum_{i=1}^{N_b} L(z_j, d(b_i)) \cdot \hat{P}(b_i|z_j) \cdot \hat{P}(z_j). \quad (2.10)$$

Dabei ist $\hat{P}(b_i|z_j)$ die bedingte Wahrscheinlichkeit für die Beobachtung b_i , wenn der Zustand z_j vorliegt. Der Term

$$\hat{P}(b_i \wedge z_j) = \hat{P}(b_i|z_j) \cdot \hat{P}(z_j)$$

entspricht der Verbundwahrscheinlichkeit von Beobachtung b_i und Zustand z_j . Wird nun über alle Beobachtungen integriert beziehungsweise summiert, die zur selben Entscheidung e_k führen, ergibt sich die Verbundwahrscheinlichkeit von e_k und z_j :

$$\hat{P}(e_k \wedge z_j) = \sum_{b_i: d(b_i)=e_k} \hat{P}(b_i|z_j) \cdot \hat{P}(z_j) \quad (2.11)$$

Die Verbundwahrscheinlichkeiten können in einer Matrix mit den Entscheidungen in den Zeilen und den Umweltzuständen in den Spalten abgelegt werden. Damit kann die Entscheidungsfunktion $d(b_i)$ aus (2.10) eliminiert werden:

$$\hat{L}_D = \sum_{k=1}^{N_e} \sum_{j=1}^{N_z} L(e_k, z_j) \cdot \hat{P}(e_k \wedge z_j). \quad (2.12)$$

Diese Gleichung sagt aus, dass ein Entscheidungssystem anhand der Kostenmatrix und Schätzungen der Verbundwahrscheinlichkeiten aus Zuständen und Entscheidungen bewertet werden kann.

Bestimmung der optimalen Entscheidungen

Je nach Umfang der vorliegenden Informationen sind unterschiedliche Bewertungen zur Bestimmung der optimalen Entscheidung vorzunehmen. Es resultieren Entscheidungsfunktionen, die entweder stochastisch oder deterministisch sind. Wird lediglich die Kostenmatrix zur Bestimmung der optimalen Entscheidung verwendet, kann zum Beispiel die MinMax-Regel angewendet werden. Es ist dann die Entscheidung auszuwählen, für die die maximal möglichen Kosten minimal werden:

$$e_{opt,def} = \arg \min_{e_k} \max_{z_j} L(e_k, z_j). \quad (2.13)$$

Die Entscheidungsfunktion d ist dann eine Konstante und weist jedem Beispiel die Entscheidung $e_{opt,def}$ zu.

Diese Situation kann auch als Zwei-Personen-Nullsummenspiel aufgefasst werden. Ein Zwei-Personen-Nullsummenspiel wird durch das Tripel (\mathcal{E}, Z, L) beschrieben [7, 109]. Dabei ist \mathcal{E} die Menge der möglichen Strategien – hier die möglichen Entscheidungen – von Spieler 1. Z sind die Strategien – hier die möglichen Umweltzustände – von Spieler 2. L ist die Kostenmatrix. Aus Sicht von Spieler 1 entspricht L einer Kostenmatrix, aus Sicht von Spieler 2 einer Gewinnmatrix. Spieler 1 sei das Entscheidungssystem und Spieler 2 sei die Umwelt. Die beste Strategie für Spieler 1, das Entscheidungssystem, lässt sich mit (2.13) ermitteln.

Die Spieltheorie [7, 109] lässt auch gemischte Strategien zu. Bei einer gemischten Strategie entscheidet sich der Spieler nicht fest für eine Entscheidung. Er nutzt einen Zufallsmechanismus, der mit bekannten Wahrscheinlichkeiten $P(e_k)$ die möglichen Entscheidungen

trifft, für die Festlegung seiner Entscheidungen. Ziel aus Sicht von Spieler 1 ist dabei, die Wahrscheinlichkeiten so zu wählen, dass unabhängig von der Strategie von Spieler 2 sein maximal möglicher Verlust geringer ist, als bei einer reinen Strategie.

In Bild 2.2 ist für die selbst gewählte Kostenmatrix

$$\mathbf{L} = \begin{pmatrix} 0 & 5 \\ 2 & 1 \end{pmatrix}$$

die Bestimmung der optimalen Wahrscheinlichkeiten $P(e_k)$ und $P(z_j)$ für die beiden Spieler dargestellt. Links ist die Situation aus der Sicht des Entscheidungssystems und rechts aus der Sicht der Umwelt dargestellt. Auf der x-Achse ist die Wahrscheinlichkeit aufgetragen, dass das Entscheidungssystem die Entscheidung e_1 trifft beziehungsweise die Umwelt den Zustand z_1 einnimmt. Die Kosten beziehungsweise der Gewinn für $P(e_1) = P(z_1) = \{0, 1\}$ werden aus der Kostenmatrix übernommen. Die Verbindungslinien entsprechen jeweils den beiden Optionen des anderen Spielers. Für das Entscheidungssystem kennzeichnet die Linie von 1 nach 5 die Kosten der Strategie, wenn die Umwelt stets im Zustand z_2 ist. Wählt Spieler 1 stets die Entscheidung e_1 und Spieler 2 stets die Entscheidung z_2 , hat Spieler 1 Kosten von 5 und Spieler 2 einen Gewinn von 5. Die Bedeutung der weiteren Linien ergibt sich analog. Der Abstand zwischen den beiden Kurven entspricht der Ungewissheit über die tatsächlichen Konsequenzen, die von den tatsächlichen Entscheidungen der beiden Spieler abhängen. Der Verlauf der erwarteten maximalen Kosten für das Entscheidungssystem und der Verlauf des erwarteten minimalen Gewinns der Umwelt sind durch die dicken Linien gekennzeichnet.

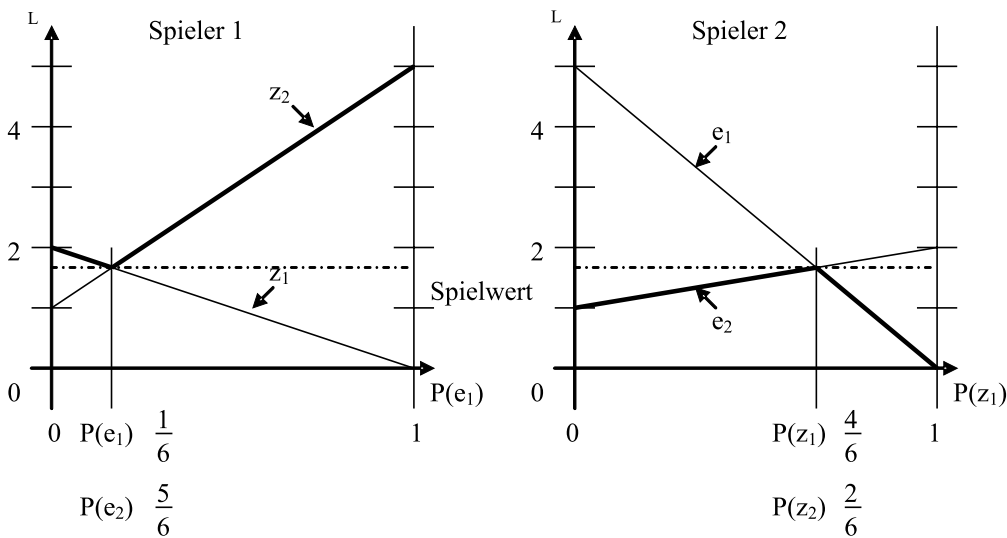


Bild 2.2: Gemischte Strategien an einem Beispiel. Links: Spieler 1 (Entscheidungssystem), rechts: Spieler 2 (Umwelt).

Durch die gemischte Strategie kann Spieler 1 seine erwarteten maximalen Kosten pro Durchgang des Spiels reduzieren (von 2 für seine beste Entscheidung nach (2.13) auf 1.66).

Spieler 2 kann seinen minimalen erwarteten Gewinn je Spiel steigern (von 1, sofern Spieler 2 versucht, seinen minimalen erwarteten Gewinn zu maximieren, auf ebenfalls 1.66). Die Erwartungen der beiden Spieler durch die gemischte Strategie sind somit identisch. Sobald ein Spieler die optimale gemischte Strategie wählt, kann der andere seine Situation durch Ändern seiner Strategie nicht beeinflussen. Dieser Zusammenhang wurde 1928 von v. NEUMANN für diskrete Zustände und Entscheidungen bewiesen [109]. Eine weitere Strategie wurde von STACKELBERG untersucht. Dabei reagiert ein Spieler auf die Entscheidung seines Gegenspielers mit seiner optimalen Entscheidung. Er geht davon aus, dass sein Gegenspieler seine Entscheidung daraufhin nicht ändern wird. Dieses Konzept wird in der Preistheorie zur Untersuchung von Oligopolen verwendet [84].

Bei nicht-spieltheoretischen Entscheidungsproblemen mit konstanten Wahrscheinlichkeiten von Umweltzuständen kann als zusätzliche Information zur Kostenmatrix die Wahrscheinlichkeit der Umweltzustände als relative Häufigkeit aus den Lerndaten geschätzt werden:

$$\hat{P}(z_j) = \frac{N_{z_j}}{N}. \quad (2.14)$$

Dabei ist N_{z_j} die Anzahl der Beispiele, bei denen der Umweltzustand z_j vorliegt, und N die Gesamtzahl der Beispiele.

Wird diese Information aus den Lerndaten für das Entscheidungssystem verwendet, ergibt sich als beste Entscheidung e_{opt} für alle Beispiele nach der μ -Regel:

$$e_{opt} = \arg \min_{e_k} \sum_{j=1}^{N_j} L(e_k, z_j) \cdot \hat{P}(z_j). \quad (2.15)$$

Eine Schätzung der Kosten pro Entscheidung ist ebenfalls möglich:

$$\hat{L}_D = \sum_{j=1}^{N_j} L(e_{opt}, z_j) \cdot \hat{P}(z_j). \quad (2.16)$$

Die Matrix der Verbundwahrscheinlichkeiten ist dann allerdings auf einen Zeilenvektor $\hat{P}(e_{opt} \wedge z_j)$ reduziert, da stets dieselbe Entscheidung e_{opt} getroffen wird.

Zu untersuchen ist nun, ob durch eine gemischte Strategie die erwarteten Kosten einer festen Entscheidung reduziert werden können. Die Strategie der Umwelt (Spieler 2) ist über die Schätzung der Wahrscheinlichkeiten der Umweltzustände bekannt. Daraus folgt, dass die Kostenmatrix aus Sicht des Entscheidungssystems (Spielers 1) in einen Vektor übergeht:

$$\hat{L}(\hat{P}(z_j)) = \begin{pmatrix} \sum_{j=1}^{N_z} L(e_1, z_j) \cdot \hat{P}(z_j) \\ \vdots \\ \sum_{j=1}^{N_z} L(e_k, z_j) \cdot \hat{P}(z_j) \\ \vdots \\ \sum_{j=1}^{N_z} L(e_{N_e}, z_j) \cdot \hat{P}(z_j) \end{pmatrix} \quad (2.17)$$

Das Minimum der erwarteten Kosten ist dann stets eine reine Strategie. Ist die Strategie der Umwelt in Form der Wahrscheinlichkeiten $\hat{P}(z_j)$ bekannt, bringen gemischte Strategien keine Verbesserung für Spieler 1.

Nun stehen zur Bestimmung der besten Entscheidung zusätzlich die Beobachtungen b_i zur Verfügung. Die Entscheidungsfunktion $d(b_i)$ weist dann jeder Beobachtung b_i die beste Entscheidung $e_{opt}(b_i)$ zu:

$$d : b_i \mapsto e_{opt}(b_i). \quad (2.18)$$

Die beste Entscheidung $e_{opt}(b_i)$ für die Beobachtung b_i ergibt sich dann wiederum nach der μ -Regel zu:

$$e_{opt}(b_i) = \arg \min_{e_k} \underbrace{\sum_{j=1}^{N_j} L(e_k, z_j) \cdot \hat{P}(z_j|b_i)}_{\hat{L}_D(e_k|b_i)}. \quad (2.19)$$

Die Bestimmung der Wahrscheinlichkeiten der Zustände bei gegebener Beobachtung $\hat{P}(z_j|b_i)$ ist abhängig vom Entwurfsverfahren und wird an dieser Stelle noch nicht betrachtet.

Soll die Entscheidung auf dem μ - σ -Prinzip basieren, ist zusätzlich ein Maß für die Schwankung der Kosten der einzelnen Entscheidungen zu berücksichtigen. Es wird bewusst nicht von einer Standardabweichung gesprochen, da diese eine Normalverteilung der Kosten voraussetzt und dies in den meisten Fällen nicht gegeben ist. Die Berechnung des in der Literatur verwendeten Maßes für die Schwankung beruht allerdings auf der Berechnung der Standardabweichung [88, 95]:

$$e_{opt,\sigma}(b_i, \alpha) = \arg \min_{e_k} \left[\underbrace{\hat{L}_D(e_k|b_i)}_{L_{D,\sigma}(e_k|b_i)} + \alpha \cdot \overbrace{\sqrt{\sum_{j=1}^{N_z} (L(e_k, z_j) - \hat{L}_D(e_k|b_i))^2 \cdot \hat{P}(z_j|b_i)}}^{\sigma} \right]. \quad (2.20)$$

Dabei ist α ein Maß für die Berücksichtigung der Risikobereitschaft des Entscheidungsträgers. Für Werte von $\alpha > 0$ liegt Risikoaversion vor. Bei identischen Kosten pro Entscheidung $\hat{L}_D(e_k|b_i)$ wird dann die Entscheidung vorgezogen, deren Kosten weniger schwanken. Wird dagegen $\alpha < 0$ gewählt, liegt Risikofreude vor und es wird bei den gleichen erwarteten Kosten die Entscheidung bevorzugt, deren Kosten stärker schwanken.

Ist $P(z_r|b_i) = 1$ für einen Zustand z_r , gehen alle Entscheidungen, für die die Information b_i vorliegt, in ein Entscheidungsproblem bei vollständiger Information über (Entscheidung bei Sicherheit in Bild 1.2). Es liegt dann der Zustand z_r vor und die für diesen Zustand beste Entscheidung ist ein Optimierungsproblem. Für die Schwankungsgröße σ aus (2.20) gilt dann $\sigma = 0$. Für diesen Fall gilt

$$e_{opt,\sigma}(b_i) = e_{opt}(b_i) \quad (2.21)$$

für alle α .

Die Entscheidungsfunktion $d(b_i)$ kann bei endlicher Anzahl Beobachtungen als Matrix \mathbf{D} der Dimension $N_e \times N_b$ geschrieben werden. Die Zuordnung von Beobachtungen zu Entscheidungen erfolgt durch eine Eins an der entsprechenden Stelle in der Matrix. Alle anderen Elemente der Matrix sind gleich Null:

$$D(k, i) = 1 \vee \{i|b_i \mapsto e_k\}. \quad (2.22)$$

	Information	Entscheidungs- kriterium	Entscheidungs- funktion
Ungewissheit	nur \mathbf{L}	MinMax-Regel (reine Strategie) MinMax-Regel (gemischte Strategie)	$d = e_{opt}$ $d(P(e_k))$
Risiko	$\mathbf{L}, \hat{P}(z_j)$	μ -Regel	$d = e_{opt}$
	$\mathbf{L}, \hat{P}(z_j b_i)$	μ -Regel $\mu - \sigma$ -Prinzip	$d(b_i)$ $d(b_i)$
Sicherheit	$\mathbf{L}, \hat{P}(z_r b_i) = 1$	Optimierung	$d(b_i)$

Tabelle 2.3: Entscheidungskriterien und Entscheidungsfunktionen bei zunehmender Information.

Die Elemente der Matrix der Verbundwahrscheinlichkeiten ergeben sich dann zu (2.11):

$$\hat{P}(e_k \wedge z_j) = \sum_{i=1}^{N_b} D(k, i) \cdot \hat{P}(b_i|z_j) \cdot \hat{P}(z_j). \quad (2.23)$$

Für gegebene Zustände z_j lassen sich die Erkennungsraten der Beobachtungen $\hat{P}(b_i|z_j)$ über das Bayes-Theorem berechnen:

$$\hat{P}(b_i|z_j) = \frac{\hat{P}(z_j|b_i) \cdot \hat{P}(b_i)}{\sum_{i=1}^{N_b} \hat{P}(z_j|b_i) \cdot \hat{P}(b_i)}. \quad (2.24)$$

Die Erkennungsrate entspricht dem Anteil Beispiele von z_j , für die die Beobachtung b_i vorliegt. Sind mehr als abzählbar viele Beobachtungen möglich, so ist d über \mathbb{R}_{N_x} auszuwerten:

$$\hat{P}(e_k \wedge z_j) = \int_B d(b) \cdot \hat{P}(b|z_j) \cdot \hat{P}(z_j) db. \quad (2.25)$$

Die Entscheidungsfunktion d beschreibt das Entscheidungssystem für gegebene Beobachtungen b_i eindeutig.

Einfluss variabler Kostenmatrizen

Ist die Kostenmatrix linear abhängig von einer Eingangsgröße c , so kann diese vor die Matrix gezogen werden:

$$L(e_k, z_j) = c \cdot L^*(e_k, z_j). \quad (2.26)$$

Werden die Wahrscheinlichkeiten der Umweltzustände für die Beobachtungen ohne die Kosteninformation geschätzt, dann ist die Entscheidung unabhängig von einer positiven Eingangsgröße c (2.19) [7]:

$$e_{opt}(b_i) = \arg \min_{e_k} \sum_{j=1}^{N_j} c \cdot L(e_k, z_j) \cdot \hat{P}(z_j|b_i) = \arg \min_{e_k} \sum_{j=1}^{N_j} L(e_k, z_j) \cdot \hat{P}(z_j|b_i), \quad c > 0. \quad (2.27)$$

Damit ändern sich die Verbundwahrscheinlichkeiten $\hat{P}(e_{opt} \wedge z_j)$ nicht und eine Verbesserung des Entscheidungssystems durch die variable Kostenmatrix ist nicht möglich. Entscheidungsprobleme mit variablen Kosten wie in Tabelle 2.2, die die Bedingung (2.26) erfüllen, können daher in äquivalente Entscheidungsprobleme mit konstanten Kosten umgewandelt werden.

Wird c als zusätzliches Merkmal dem Lerndatensatz hinzugefügt, können sich die Beobachtungen und damit auch die Schätzung von $\hat{P}(z_j|b_i)$ ändern. Dies wiederum kann zu einer geänderten Entscheidungsfunktion d und zu anderen Verbundwahrscheinlichkeiten $\hat{P}(e_{opt} \wedge z_j)$ führen.

2.2.2 Merkmalskosten

Überblick

Neben den Kosten für Fehlentscheidungen existieren noch weitere Kostenarten, die beim Entwurf optional berücksichtigt werden können. Die Kosten für Informationen beziehungsweise Merkmale, die das Entscheidungssystem verwendet, sind ein Beispiel. Eine detaillierte Auflistung der Kostenarten bei Entscheidungssystemen ist in [146] veröffentlicht. Neben den Kosten für Fehlentscheidungen und Merkmale werden unter anderem Kosten für ein übergeordnetes Entscheidungssystem, die Beeinflussung des betrachteten Prozesses, Berechnungen der Merkmale, Bereitstellung der Lerndaten oder für die Interaktion zwischen Mensch und Computer unterschieden. Es wird angenommen, dass das übergeordnete Entscheidungssystem, zum Beispiel ein Lehrer, immer die richtige Entscheidung trifft. Diese Unterstützung ist jedoch zu bezahlen. Wenn die vorhandene Rechenzeit knapp ist, ist es sinnvoll, den Merkmalen je nach Komplexität ihrer Berechnung Kosten zuzuweisen. Die Erfassung der Lerndaten erfordert ebenfalls einen finanziellen Aufwand. Je mehr Lerndaten verwendet werden, desto besser wird tendenziell das Entscheidungssystem sein, jedoch mit umso höheren Kosten für die Bereitstellung der Lerndaten. Wichtiger als die absolute Menge der Daten ist allerdings die wirklichkeitsgetreue Abbildung des Prozesses. Dies ist insbesondere dann schwierig, wenn seltene Ereignisse nicht in den Daten enthalten sind. Messdaten für den größten anzunehmenden Unfall (GAU) eines Kernkraftwerkes stehen für den Entwurfsprozess eines Entscheidungssystems nicht zur Verfügung und müssen durch Simulationen ersetzt werden. Die Kosten für die Interaktion zwischen Mensch und Computer beinhalten zum Beispiel die Zeit, die der Mensch für die Konfigurierung des Computers vor dem Entwurf des Entscheidungssystems benötigt. Darüber hinaus werden in dieser Kategorie auch Aspekte der Interpretierbarkeit des Entscheidungssystems selbst und seiner Entscheidungen betrachtet [16, 39, 105].

Im Folgenden werden die Merkmalskosten als wichtigste Kostenart neben den Entscheidungskosten weiter untersucht. Verschiedene Autoren haben sich mit Merkmalskosten bei der Entscheidungsfindung beschäftigt. Das Verhältnis zwischen den Kosten für medizinische Untersuchungen und Erwartungen über die zusätzliche Genauigkeit der Diagnose beziehungsweise dem therapeutischen Erfolg wird in [143] diskutiert. Ein konkretes Entscheidungssystem wird allerdings nicht betrachtet. In [114, 145] werden die Merkmalskosten des Entscheidungssystems erfasst und in den Entwurfsprozess rückgekoppelt. Die Merkmale können in beiden Arbeiten zu Gruppen zusammengefasst werden. Bei der Verwendung eines Merkmales aus einer Gruppe stehen alle weiteren dem Entscheidungssys-

tem ohne zusätzliche Kosten [114] oder mit reduzierten Kosten [145] zur Verfügung.

In [2] wird bei sequentieller Generierung der Merkmale für die Entscheidungsfindung die kostenminimale Reihenfolge über eine Baumsuche bestimmt. Ein ähnliche Aufgabe wird in [121] bearbeitet. Dort ist die beste Testreihenfolge zur Bestimmung des Umweltzustandes gesucht. Gegeben sind die Kosten der einzelnen Tests und die Zustände, die abhängig vom Testergebnis nach der Durchführung des Tests zur Auswahl stehen. In [63] hat das Entscheidungssystem nach jedem ausgewerteten Merkmal die Möglichkeit, weitere Merkmale zu verwenden oder eine Entscheidung zu treffen. Ein ähnlicher Fall wird in [48] betrachtet. Der Entscheidungsträger kann hier solange weitere detailliertere Informationen kaufen, bis er bereit ist, seine Entscheidung zu treffen. Für die Prüfung von Silizium-Wafern ist in [108] die Kostenstruktur für den gesamten Prüfprozess aufgegliedert.

Neben den Kosten der Merkmale kann auch das Wissen über deren Genauigkeit in den Entwurf eines Entscheidungssystems einfließen. Mit Ungewissheit bei der Erfassung von Merkmalen mittels Sensoren beschäftigt sich [24, 76]. Dabei werden als Einflussfaktoren zum Beispiel Drift durch Alterung des Sensors, Temperaturänderungen, fehlerhafte Kalibrierung, Diskretisierungsfehler oder Interferenzen mit anderen elektrischen Geräten betrachtet. Nach diesen allgemeinen Bemerkungen zu Merkmalen und den dadurch verursachten Kosten folgt nun eine exakte Definition des Begriffes Merkmal. Anschließend wird eine Aufteilung der Merkmalskosten in variable und fixe Kosten vorgenommen.

Definition Merkmal

Ein Merkmal ist eine sichtbare oder messbare Eigenschaft von Objekten, Personen oder Prozessen und Systemen. Eine sichtbare Eigenschaft ist die Haarfarbe einer Person, eine messbare das Volumen einer Flasche. Merkmale können auf einer oder mehreren Messungen basieren. Die Durchschnittstemperatur des Kühlwassers eines Kraftwerks über mehrere Tage ist ein Merkmal des Prozesses 'Stromerzeugung im Kraftwerk'. Ein Merkmal entsteht allgemein durch Abbildung von sichtbaren oder messbaren Größen auf den Wertebereich des Merkmals. Die einfachste Abbildung $f(x) = x$ bedeutet, dass der Messwert selbst das Merkmal ist. Für die verwendeten Sensoren, Tests oder die beanspruchte menschliche Arbeit entstehen Kosten.

Eine Einteilung der Merkmale in Kategorien ist in Bild 2.3 dargestellt [50]. Zunächst ist zu unterscheiden, ob der Wertebereich des Merkmal auf einer kardinalen, ordinalen oder nominalen Skala liegt. Auf kardinalen beziehungsweise metrischen Skalen sind Addition und Multiplikation definiert. Solche Merkmale heißen quantitativ. Ein quantitatives Merkmal ist diskret, wenn höchstens abzählbar viele Ausprägungen des Merkmals vorliegen (z.B. Anzahl Fehler). Kontinuierliche Merkmale können dagegen unendlich viele Ausprägungen annehmen. Ordinale und nominale Merkmale sind qualitative Merkmale, für deren Skalen Addition und Multiplikation nicht zulässig ist. Ordinale Merkmale erlauben eine Anordnung beziehungsweise Sortierung ihrer Ausprägungen. Zum Beispiel der Grad einer Verbrennung entspricht einem ordinalen Merkmal mit den Ausprägungen Verbrennung ersten, zweiten oder dritten Grades.

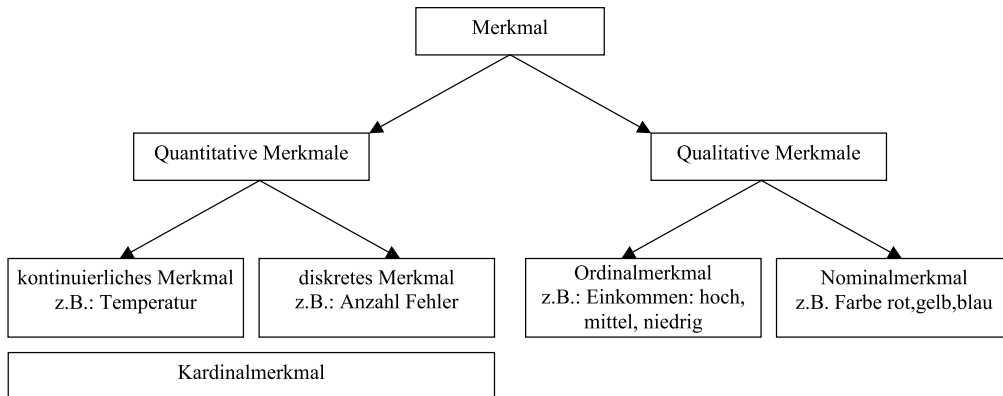


Bild 2.3: Einteilung von Merkmalen vereinfacht nach [50].

Variable und fixe Kosten bei Merkmalen

Die Kostentheorie teilt die Kosten von Aktivitäten in der Produktion in fixe und variable Kosten [56]. Produktion bedeutet dabei die Kombination von Produktionsfaktoren (Arbeit, Maschinen, Werkstoffe). Das Ergebnis sind Produkte (Autos, Kleidung, Bücher). Ein Merkmal ist ebenfalls ein Produkt. Die Kombination der Faktoren Sensor, Abbildung und Wertebereich führt zu dem Produkt Merkmal.

Fixe Kosten K_{fix} sind unabhängig davon, ob produziert wird oder nicht. Es handelt sich um Kosten für die Bereitstellung der Produktionsfaktoren. Für einen Sensor sind folgende Kosten fix, unabhängig von der Nutzung des Sensors:

- Sensoranschaffung,
- Integration in die Anlage (Arbeit und Material),
- Entwicklungsarbeiten.

Bezogen auf die einzelne Auswertung des Merkmals sinken die fixen Kosten mit steigender Anzahl Auswertungen. Für die fixen Kosten pro Entscheidung eines Merkmals x_l ergibt sich:

$$L_{C,l,fix} = \frac{K_{fix}}{N_l}. \quad (2.28)$$

Dabei steht der Index C für Merkmalskosten und N_l für die Anzahl Auswertungen des Merkmals x_l .

Variable Kosten $L_{C,l,var}$ entstehen zusätzlich durch die Produktion. Für einen Sensor resultieren variable Kosten zum Beispiel durch:

- Stromverbrauch,
- Verbrauchsmaterial,
- nutzungsabhängige Wartung, Kalibrierung.

Die variablen Kosten sind unabhängig von der Anzahl Auswertungen. Für den Entwurf eines Entscheidungssystems sind die Gesamtkosten eines Merkmals x_l pro Auswertung $L_{C,l}(N_l)$ zu ermitteln. Die Gesamtkosten pro Auswertung des Merkmals ergeben sich zu:

$$L_{C,l}(N_l) = \frac{K_{fix}}{N_l} + L_{C,l,var}. \quad (2.29)$$

In Bild 2.4 ist der Verlauf der Kosten für $K_{fix} = 10$ und $L_{C,l,var} = 1$ dargestellt. Stellt L_{fv} das Verhältnis der fixen Kosten K_{fix} zu den variablen Kosten pro Entscheidung $L_{C,l,var}$ dar, ergibt sich das Verhältnis aus fixen Kosten pro Entscheidung zu variablen Kosten pro Entscheidung:

$$\frac{L_{C,l,fix}(N_l)}{L_{C,l,var}} = \frac{\frac{K_{fix}}{N_l}}{L_{C,l,var}} = \frac{L_{C,l,var} \cdot L_{fv}}{N_l} = \frac{L_{fv}}{N_l}. \quad (2.30)$$

Unabhängig von der Höhe der fixen Kosten K_{fix} geht deren Anteil für $N_l \gg L_{fv}$ gegen Null. Die Kosten des Merkmals pro Entscheidung entsprechen dann den variablen Kosten. In Bild 2.4 ist dies an der Annäherung der gepunkteten Kurve von $L_{C,l}(N_l)$ an die strichpunktierte Linie von $L_{C,l,var}(N_l)$ zu erkennen. Die durchgezogene Linie stellt die fixen Kosten des Merkmals pro Entscheidung $L_{C,l,fix}(N_l)$ dar. Die fixen Kosten pro Entscheidung gehen für große N_l gegen Null. Im weiteren Verlauf der Arbeit werden daher die fixen Kosten vernachlässigt beziehungsweise über eine Konstante in $L_{C,l,var}$ integriert, indem eine Abschätzung für N_l vorgenommen wird.

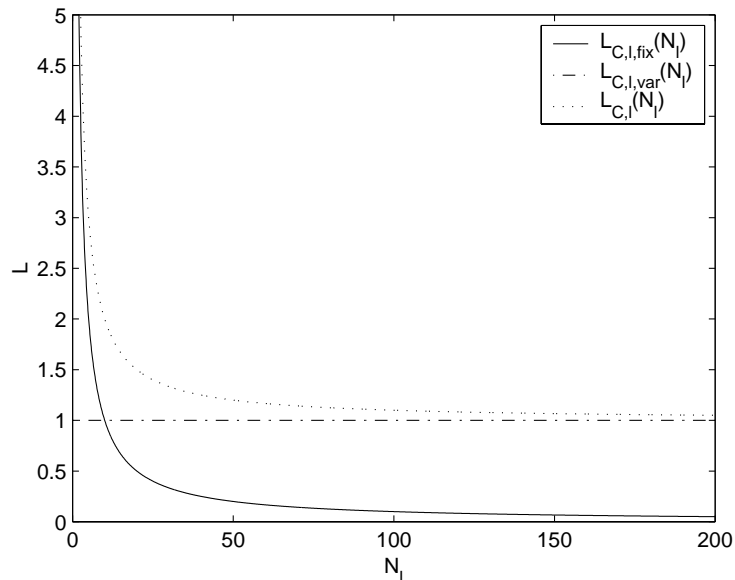


Bild 2.4: Verlauf der variablen und fixen Merkmalskosten in Abhängigkeit von der Anzahl Auswertungen: $K_{fix} = 10$, $L_{C,l,var} = 1$.

2.2.3 Neuartige zweidimensionale Visualisierung

In diesem Abschnitt wird eine neuartige grafische zweidimensionale Visualisierung von Entscheidungssystemen mit zwei Umweltzuständen z_1 und z_2 vorgestellt.

Aus dem Lerndatensatz werden die a-priori Wahrscheinlichkeiten $\hat{P}(z_1)$ und $\hat{P}(z_2)$ sowie zwei Beobachtungen b_1 und b_2 ermittelt. Es stehen zwei Entscheidungen e_1 und e_2 zur Verfügung. Eine 2×2 Kostenmatrix \mathbf{L} , die keine Dominanzbeziehungen aufweist, ist ebenfalls gegeben. Für den Umweltzustand z_1 sei e_1 die beste Entscheidung und für z_2 sei e_2 die beste. Für \mathbf{L} gilt dann: $L(e_1|z_2) > L(e_2|z_2)$ und $L(e_2|z_1) > L(e_1|z_1)$.

Für die Visualisierung eines Entscheidungssystems werden die bedingten Wahrscheinlichkeiten der Beobachtungen bei gegebenen Zuständen $\hat{P}(b_1|z_1)$ und $\hat{P}(b_2|z_2)$ herangezogen. Die bedingten Wahrscheinlichkeiten $\hat{P}(b_1|z_1)$ und $\hat{P}(b_2|z_2)$ entsprechen den Erkennungsraten des Entscheidungssystems für die Zustände. Diese Erkennungsraten entsprechen den Anteilen korrekt entschiedener Beispiele, wenn die Entscheidungsfunktion b_1 die Entscheidung e_1 und b_2 die Entscheidung e_2 zuordnet. Ein optimales Entscheidungssystem erkennt alle Beispiele korrekt: $\hat{P}(b_1|z_1) = \hat{P}(b_2|z_2) = 1$. Ein Entscheidungssystem mit den Erkennungsraten $\hat{P}(b_1|z_1) = \hat{P}(b_2|z_2) = 0$ und obiger Entscheidungsfunktion trifft systematisch falsche Entscheidungen. Durch Veränderung der Zuordnung zwischen Beobachtungen und Entscheidungen kann für den letzten Fall das optimale Entscheidungssystem erzeugt werden.

Für Werte von $\hat{P}(b_1|z_1)$ und $\hat{P}(b_2|z_2)$ zwischen 0 und 1 ist nicht unmittelbar erkennbar, welche Entscheidungen für die beiden Beobachtungen optimal sind. Die Entscheidung e_1 ist bei Beobachtung b_1 die günstigste, wenn gilt (2.19):

$$L(e_1, z_1) \cdot \hat{P}(z_1|b_1) + L(e_1, z_2) \cdot \hat{P}(z_2|b_1) < L(e_2, z_1) \cdot \hat{P}(z_1|b_1) + L(e_2, z_2) \cdot \hat{P}(z_2|b_1). \quad (2.31)$$

Umformen führt zu:

$$[L(e_1, z_1) - L(e_2, z_1)] \cdot \hat{P}(z_1|b_1) + [L(e_1, z_2) - L(e_2, z_2)] \cdot \hat{P}(z_2|b_1) < 0. \quad (2.32)$$

Mit (2.24) und $\hat{P}(b_1|z_1) \cdot \hat{P}(z_1) + \hat{P}(b_1|z_2) \cdot \hat{P}(z_2) = \hat{P}(b_1)$ ergibt sich:

$$[L(e_1, z_1) - L(e_2, z_1)] \cdot \frac{\hat{P}(b_1|z_1) \cdot \hat{P}(z_1)}{\hat{P}(b_1)} < -[L(e_1, z_2) - L(e_2, z_2)] \cdot \frac{\hat{P}(b_1|z_2) \cdot \hat{P}(z_2)}{\hat{P}(b_1)}. \quad (2.33)$$

Umstellen und $\hat{P}(b_1|z_2) = 1 - \hat{P}(b_2|z_2)$ sowie $L(e_1, z_2) - L(e_2, z_2) > 0$ führt auf die Bedingung für e_1 als günstigste Entscheidung für die Beobachtung b_1 :

$$\frac{1 - \hat{P}(b_2|z_2)}{\hat{P}(b_1|z_1)} < \frac{L(e_2, z_1) - L(e_1, z_1)}{L(e_1, z_2) - L(e_2, z_2)} \cdot \frac{\hat{P}(z_1)}{\hat{P}(z_2)}. \quad (2.34)$$

Umformen führt zur Struktur einer Geradengleichung:

$$\hat{P}(b_2|z_2) > 1 - \frac{L(e_2, z_1) - L(e_1, z_1)}{L(e_1, z_2) - L(e_2, z_2)} \cdot \frac{\hat{P}(z_1)}{\hat{P}(z_2)} \cdot \hat{P}(b_1|z_1). \quad (2.35)$$

In dem Gebiet innerhalb der von $\hat{P}(b_1|z_1), \hat{P}(b_2|z_2) \in [0 \dots 1]$ aufgespannten Ebene, in dem Gleichung (2.35) erfüllt ist, verursacht die Entscheidung e_1 bei b_1 die niedrigeren Kosten. Im Rest des Gebietes ist für b_1 die Entscheidung e_2 die günstigere. Die beiden Gebiete

werden durch die Koordinatenachsen, die Geraden $\hat{P}(b_1|z_1) = 1$ sowie $\hat{P}(b_2|z_2) = 1$ und die Gerade aus (2.35) begrenzt.

Eine analoge Rechnung führt auf die Bedingung für die günstigere Entscheidung bei der Beobachtung b_2 :

$$\hat{P}(b_1|z_1) > 1 - \frac{L(e_1, z_2) - L(e_2, z_2)}{L(e_2, z_1) - L(e_1, z_1)} \cdot \frac{\hat{P}(z_2)}{\hat{P}(z_1)} \cdot \hat{P}(b_2|z_2). \quad (2.36)$$

Ist diese Bedingung erfüllt, ist e_2 die günstigere Entscheidung für die Beobachtung b_2 , andernfalls ist e_1 die bessere Entscheidung.

Zwei Vereinfachungen werden kurz erläutert. Sind die Kosten für die korrekten Entscheidungen und die Fehlentscheidungen jeweils identisch:

$$L(e_1|z_2) = L(e_2|z_1) \quad \wedge \quad L(e_1|z_1) = L(e_2|z_2) = 0, \quad (2.37)$$

wird (2.35) zu:

$$\hat{P}(b_1|z_1) > 1 - \frac{\hat{P}(z_2)}{\hat{P}(z_1)} \cdot \hat{P}(b_2|z_2). \quad (2.38)$$

In diesem Fall ist die Gerade, an der die Entscheidung kippt, von den Kosten unabhängig. Sind zusätzlich zu (2.37) die Wahrscheinlichkeiten der beiden Umweltzustände identisch

$$\hat{P}(z_1) = \hat{P}(z_2), \quad (2.39)$$

wird (2.35) zu:

$$\hat{P}(b_1|z_1) > 1 - \hat{P}(b_2|z_2). \quad (2.40)$$

Die Gerade, an der die Entscheidung kippt, ist nun stets die gleiche. Für die Verlustmatrix:

$$\mathbf{L} = \begin{pmatrix} 0 & 1 \\ L_R & 0 \end{pmatrix} \quad (2.41)$$

und die Wahrscheinlichkeiten der Umweltzustände $P(z_1) = 1/6$ und $P(z_2) = 5/6$ sind in Bild 2.5 die Geraden eingezeichnet, an denen sich die Entscheidungen für die Beobachtungen ändern. L_R entspricht dem Kostenverhältnis der beiden Fehlentscheidungen zueinander. Im linken Bild ist $L_R = 10$ und für das rechte Bild gilt $L_R = 2.5$.

Die strichpunktierte Linie entspricht dabei der Bedingung aus (2.35). An dieser Gerade ändert sich die Entscheidung für b_1 . Die punktierte Linie entspricht der Bedingung aus (2.36). An dieser Gerade ändert sich die optimale Entscheidung für b_2 . Die Endpunkte der punktierten Linie lassen sich durch einsetzen von $P(b_2|z_2) = 0$ und $P(b_2|z_2) = 1$ in (2.36) bestimmen. Für den linken Teil von Bild 2.5 ergibt sich:

$$\begin{aligned} P(b_2|z_2) = 0 & \rightarrow P(b_1|z_1) > 1 - \frac{1-0}{10-0} \cdot \frac{5}{6} \cdot 0 = 1 \rightarrow A(0, 1) \\ P(b_2|z_2) = 1 & \rightarrow P(b_1|z_1) > 1 - \frac{1-0}{10-0} \cdot \frac{5}{6} \cdot 1 = \frac{1}{2} \rightarrow B(1, 0.5) \end{aligned}$$

Die anderen Linien können analog bestimmt werden. Zu bemerken ist, dass im Bereich zwischen den beiden Geraden für beide Beobachtungen die selbe Entscheidung kostenminimal wird. Ein Entscheidungssystem, das allein auf einer Entscheidungsregel basiert, die

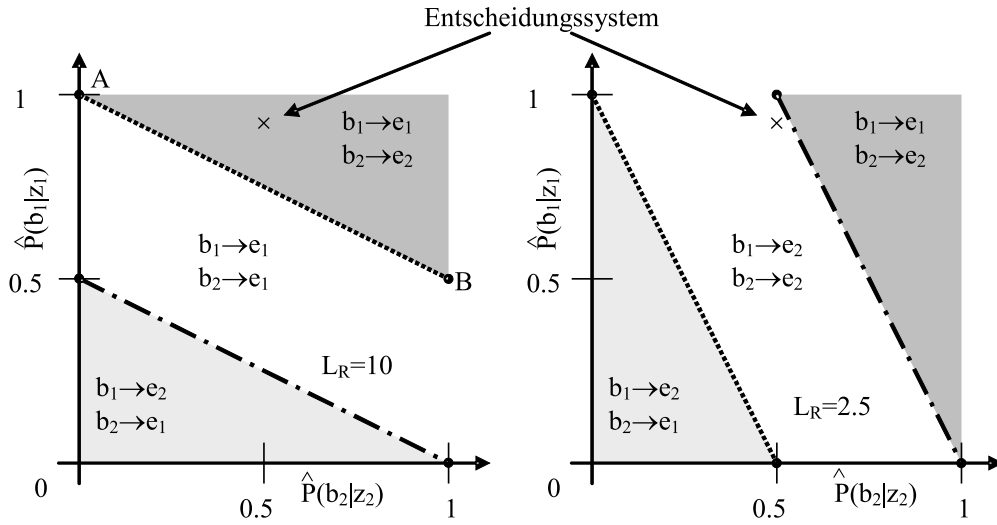


Bild 2.5: Darstellung der Linien, an denen die Entscheidungen kippen. Strichpunktirt: Entscheidung für b_1 kippt, punktiert: Entscheidung für b_2 kippt. ($P(z_1) = 1/6, P(z_2) = 5/6$; links $L_R = 10$, rechts: $L_R = 2.5$). \times : Visualisierung eines Entscheidungssystems.

in diesem Gebiet liegt, entspricht einer festen Entscheidung. Ein solches Entscheidungssystem ist damit wertlos. Das mit \times gekennzeichnete Entscheidungssystem in Bild 2.5 ist für $L_R = 10$ sinnvoll, für $L_R = 2.5$ dagegen nicht.

Der Bereich, in dem für beide Beobachtungen die selbe Entscheidung zu treffen ist, wird dann minimal, wenn für die Steigung der Trenngeraden gilt:

$$m = -\frac{L(e_1, z_2) - L(e_2, z_2)}{L(e_2, z_1) - L(e_1, z_1)} \cdot \frac{\hat{P}(z_2)}{\hat{P}(z_1)} = -1. \quad (2.42)$$

Dieser Fall ist genau dann möglich, wenn das Verhältnis der Wahrscheinlichkeiten der Umweltzustände identisch zum inversen Verhältnis der absoluten Spaltendifferenzen der Kostenmatrix ist:

$$\frac{L(e_1, z_2) - L(e_2, z_2)}{L(e_2, z_1) - L(e_1, z_1)} = \frac{\hat{P}(z_1)}{\hat{P}(z_2)}. \quad (2.43)$$

Zu bestimmen ist noch die Steigung der Isokostenlinien. Am Beispiel der Entscheidungsfunktion $b_1 \mapsto e_1, b_2 \mapsto e_2$ wird diese hergeleitet. Die Verbundwahrscheinlichkeiten $\hat{P}(e_k \wedge z_j)$ ergeben sich dann in Abhängigkeit von den Erkennungsraten zu:

$$\hat{P}(e_k \wedge z_j) = \begin{pmatrix} \hat{P}(b_1|z_1) \cdot \hat{P}(z_1) & [1 - \hat{P}(b_2|z_2)] \cdot \hat{P}(z_2) \\ [1 - \hat{P}(b_1|z_1)] \cdot \hat{P}(z_1) & \hat{P}(b_2|z_2) \cdot \hat{P}(z_2) \end{pmatrix} \quad (2.44)$$

Da hier von einer konstanten Kostenmatrix ausgegangen wird, ist zur Bestimmung des

totalen Differentials von Gleichung (2.12) lediglich $\hat{P}(e_k \wedge z_j)$ total zu differenzieren:

$$\frac{\partial \hat{P}(e_k \wedge z_j)}{\partial \hat{P}(b_1|z_1)} \cdot \Delta P(b_1|z_1) + \frac{\partial \hat{P}(e_k \wedge z_j)}{\partial \hat{P}(b_2|z_2)} \cdot \Delta P(b_2|z_2) = \quad (2.45)$$

$$\begin{pmatrix} \Delta \hat{P}(b_1|z_1) \cdot \hat{P}(z_1) & -\Delta \hat{P}(b_2|z_2) \cdot \hat{P}(z_2) \\ -\Delta \hat{P}(b_1|z_1) \cdot \hat{P}(z_1) & \Delta \hat{P}(b_2|z_2) \cdot \hat{P}(z_2) \end{pmatrix}.$$

Ausmultiplizieren mit **L** und Summieren ergibt:

$$\Delta \hat{P}(b_2|z_2) = -\frac{L(e_1, z_2) - L(e_2, z_2)}{L(e_2, z_1) - L(e_1, z_1)} \cdot \frac{\hat{P}(z_2)}{\hat{P}(z_1)} \cdot \Delta \hat{P}(b_1|z_1). \quad (2.46)$$

Die Steigung der Isokostenlinie ist identisch zur Steigung der Trenngeraden der Entscheidungsgebiete (2.42).

Das vorgestellte Verfahren zur Visualisierung von Entscheidungssystemen mit je zwei Zuständen, Entscheidungen und Beobachtungen bietet insbesondere für die Visualisierung von einzelnen WENN-DANN-Regeln Vorteile. Die Prämisse der Regel und deren Negation entsprechen den zwei Beobachtungen. Die Konklusionen für die Prämisse und ihre Negation entsprechen den beiden Entscheidungen. Anhand der Erkennungsraten der Regel für die beiden Umweltzustände ist unmittelbar ersichtlich, ob die Regel die Kosten für Entscheidungen im Vergleich zu einer Standard-Entscheidung senkt oder nicht.

Das Verfahren ist vergleichbar mit Visualisierungen auf der Basis der ROC³-Analyse [52, 119]. Die ROC-Analyse ist ebenfalls auf zwei Zustände (in der Regel 'Positiv' und 'Negativ' genannt) beschränkt. Ein Entscheidungssystem wird dabei häufig über die Erkennungsrate der Beispiele für den Zustand 'Positiv' auf der y-Achse und die Fehlerrate für den Zustand 'Negativ' auf der x-Achse visualisiert. Die Fehlerrate entspricht analog zur Erkennungsrate dem Anteil Beispiele eines Zustandes, für den das Entscheidungssystem die falsche Entscheidung trifft. Für jede Kombination aus Erkennungsrate und Fehlerrate ergibt sich auch hier ein Punkt im ROC-Graphen. Bei dem in dieser Arbeit vorgestellten Verfahren werden zusätzlich noch mittels der Kostenmatrix Bereiche hervorgehoben, in denen Entscheidungsfunktionen zu Kostenreduktionen führen beziehungsweise überflüssig sind.

Mit der Darstellung der entscheidungstheoretischen Grundlagen und dem neuartigen Visualisierungskonzept für Entscheidungssysteme sind alle Vorarbeiten zur Entwicklung von neuen entscheidungstheoretischen Bewertungsmaßen für Fuzzy-Systeme erfolgt.

2.3 Fuzzy-System

In diesem Abschnitt wird zunächst die Struktur der betrachteten Fuzzy-Systeme erläutert. Anschließend werden die Wahrscheinlichkeitstheorie und die Theorie der unscharfen Mengen (Fuzzy-Set Theory) voneinander abgegrenzt. Schließlich wird ein bekanntes und ein neues Inferenzverfahren für Fuzzy-Regeln vorgestellt.

2.3.1 Struktur und Bezeichnungen

Die zu entwerfenden Fuzzy-Systeme bestehen aus Regeln R_r der Struktur:

³ROC: Receiver Operating Characteristic

$$R_r : \text{WENN } \underbrace{x_1 = A_{1,R_r}}_{\text{Teilprämisse } P_{r,1}} \text{ UND } \cdots \text{ UND } \underbrace{x_{N_x} = A_{N_x,R_r}}_{\text{Teilprämisse } P_{r,N_x}} \text{ DANN}$$

$$\underbrace{\hspace{10em}}_{\text{Prämisse } P_r}$$

$$e = C_r, \quad r = 1, \dots, N_r,$$

und einer SONST-Regel $R_{N_r+1} : \text{SONST } e = C_{N_r+1}$.

Dabei stehen Lerndaten mit N Beispielen, N_x Merkmalen $x_l[n]$ und einer bestimmten Ausgangsvariablen y mit N_z unterschiedlichen Umweltzuständen z_j zur Verfügung. Die Eingangsgrößen eines Beispiels sind im Vektor $\mathbf{x}[n]$ zusammengefasst. Die Prämisse P_r besteht aus 'UND'-Verknüpfungen von Teilprämissen $P_{r,1}, \dots, P_{r,N_x}$. Die linguistischen Terme A_{l,R_r} können einzelne linguistische Terme $A_{l,i}$ ($i = 1, \dots, m_l$) des Merkmals x_l oder 'ODER'-Verknüpfungen der benachbarten Terme $A_{l,i_{min}}$ bis $A_{l,i_{max}}$ oder von allen linguistischen Termen des Merkmals x_l sein. Dabei ist m_l die Anzahl linguistischer Terme des Merkmals x_l . Nicht alle Merkmale x_l müssen in den Teilprämissen einer Regel vertreten sein. Jede Regelkonklusion C_r entspricht genau einer Entscheidung $e(C_r) \in \{e_1, \dots, e_{N_e}\}$. Die Konklusion C_{N_r+1} der SONST-Regel entspricht einer Default-Entscheidung für alle Beispiele, für die keine Prämisse aktiviert ist. Da die Entscheidungen keiner Ordnung folgen müssen und in der Regel keine Mischungen aus Entscheidungen möglich sind⁴, wird eine Maximum-Defuzzifizierung verwendet, um die beste Entscheidung $e_{opt}[\mathbf{x}]$ für die gegebenen Werte der Merkmale und die verwendete Regelbasis zu bestimmen.

In [102] werden UND- und ODER-Operatoren für Fuzzy-Systeme der beschriebenen Struktur untersucht. Als UND-Operator zur Verknüpfung von Teilprämissen ist das Produkt geeignet. Als ODER-Operator zur Verknüpfung von benachbarten Termen ist die beschränkte Summe zu verwenden.

Die mittleren linguistischen Terme ($A_{l,2}, \dots, A_{l,m_l-1}$) haben dreieckige Zugehörigkeitsfunktionen und die Randterme ($A_{l,1}, A_{l,m_l}$) haben trapezförmige Zugehörigkeitsfunktionen. Die Summe der Werte der Zugehörigkeit zu den einzelnen Termen ist für alle Werte von x_l stets Eins. Bild 2.6 zeigt ein Beispiel. Die Zugehörigkeitsfunktionen für ein Merkmal x_l sind somit durch m_l Parameter $a_{l,i}$, die Knickpunkte der Randterme und die Spitzen der mittleren Terme, eindeutig bestimmt.

Für die Initialisierung der m_l Parameter $a_{l,i}$ sind verschiedene Verfahren bekannt. Neben der manuellen auf heuristischem Expertenwissen basierten Methode werden die Parameter auch äquidistant verteilt oder über Clusterverfahren [86]⁵ bestimmt. In der vorliegenden Arbeit wird das in [103] vorgeschlagene Verfahren verwendet. Die Parameter werden so initialisiert, dass die Intervalle $a_{l,j} < x_l < a_{l,j+1}$ stets ungefähr die selbe Anzahl an Beispielen von ca. $N/(m_l - 1)$ beinhalten und möglichst gerundete Werte aufweisen. Dieses Verfahren ist reproduzierbar und benötigt kein Expertenwissen.

Bei der Initialisierung der Zugehörigkeitsfunktionen werden die Kostenmatrix \mathbf{L} und die Umweltzustände z_j der Beispiele nicht berücksichtigt. Eine spätere Optimierung der Zugehörigkeitsfunktionen im Anschluss an den Entwurfsprozess kann zu einer Verbesserung des Entscheidungssystems führen. Ein entsprechendes Verfahren wird in Ab-

⁴Für ordinale Entscheidungen ist die Verwendung einer Defuzzifizierung nach der Schwerpunktmethod für Singletons möglich.

⁵Wird eine zufällige Startaufteilung verwendet, ist die Reproduzierbarkeit der Ergebnisse nicht gewährleistet.

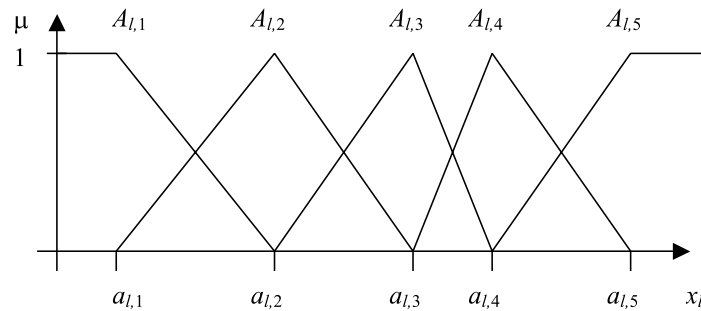


Bild 2.6: Beispiel: Zugehörigkeitsfunktionen des Merkmals x_l (fünf Terme).

schnitt 3.3.3 entworfen und diskutiert.

Für die Veranschaulichung der diskutierten Zusammenhänge wird ein selbst erstellter Beispieldatensatz mit den beiden diskreten Umweltzuständen z_1 und z_2 verwendet. Dieser Datensatz besteht aus 60 Beispielen des Umweltzustandes z_1 und 300 Beispielen des Umweltzustandes z_2 . Es existieren die beiden Merkmale x_1 und x_2 . Die Werte für die einzelnen Beispiele sind Zufallswerte. In Bild 2.7 ist der Beispieldatensatz über den Merkmalen x_1 und x_2 dargestellt. Die großen Punkte markieren die Beispiele des Umweltzustandes z_1 , die Rhomben markieren die Beispiele des Umweltzustandes z_2 . Zusätzlich ist in Bild 2.7 eine Regelbasis eingezeichnet. Die Regelbasis besteht im Einzelnen aus den in Tabelle 2.4 dargestellten Regeln. Die Regeln sind in Bild 2.7 durch Rahmen markiert. Es ist zu erkennen, dass sich die erste und die dritte Regel sowie die dritte und die vierte Regel überlappen. Die gepunktete Kurve kennzeichnet die Bereiche, in denen die selbe Entscheidung getroffen wird. Im Bereich ohne Regel wird die Entscheidung von der SONST-Regel getroffen. Es ist zu erkennen, dass nicht für alle Beispiele die richtige Entscheidung getroffen werden kann, da sich beispielsweise im unteren rechten Bereich von Regel 4 sowohl Beispiele des Umweltzustandes z_1 als auch Beispiele des Umweltzustandes z_2 befinden. Je nach Konsequenzen der Fehlentscheidungen der beiden Umweltzustände werden unterschiedliche Regelbasen und damit unterschiedliche Bereiche für die beiden Entscheidungen bei diesem Problem optimal sein.

2.3.2 Wahrscheinlichkeiten und Entscheidungen

Das Entwurfsverfahren des Fuzzy-Systems fasst Gebiete des Merkmalsraumes geeignet zusammen und repräsentiert diese durch Prämissen von Fuzzy-Regeln. Die Prämissen der Fuzzy-Regeln entsprechen den diskreten Beobachtungen bei der statistischen Entscheidungstheorie. Die Wahrscheinlichkeiten der einzelnen Umweltzustände innerhalb des Gebietes der Prämissen werden ausgewertet. Dazu werden die Verteilungsdichtefunktionen der kontinuierlichen Merkmale in Abhängigkeit von den Umweltzuständen für das Gebiet der Prämissen und den Rest des Merkmalsraumes diskretisiert. Auf der Basis der Wahrscheinlichkeiten wird bestimmt, welche Entscheidungen (Konklusionen der Regeln) für die Beispiele der Prämissen optimal sind. Es handelt sich daher um Entscheidungen bei Risiko. Die Prämissen sollen dabei so gewählt werden, dass sie möglichst viele Beispiele, für die

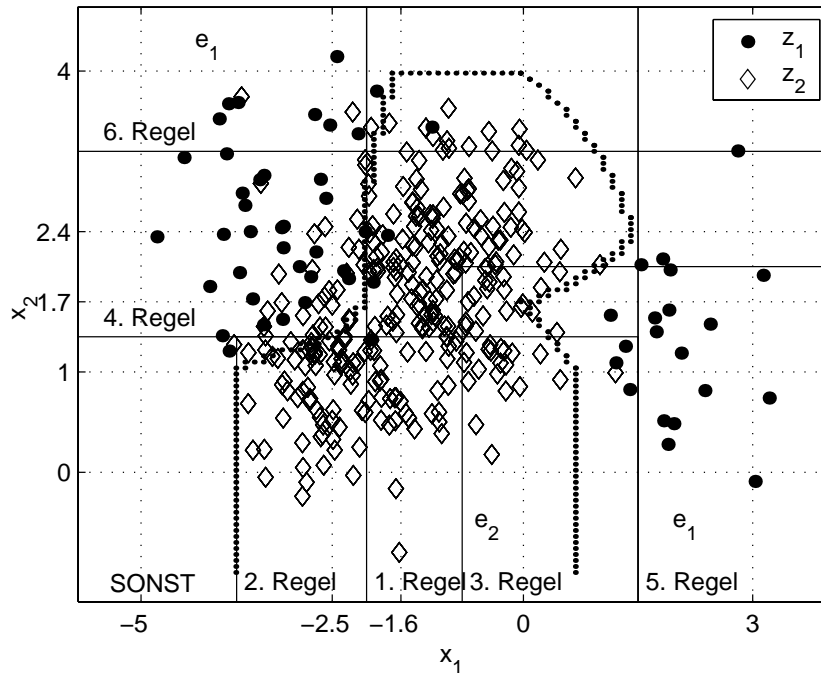


Bild 2.7: Regelbasis für den Beispieldatensatz über den Merkmalen x_1 und x_2 . Gepunktete Geraden: $\mu_{A_{l,i}}=1$, Durchgezogene Linien: Näherung für $\mu_{P_r} = 0.5$. Gepunktete Kurven: Begrenzungen der Gebiete der Entscheidungen.

eine Entscheidung optimal ist, erfassen. Die Verarbeitung der Informationen einer Regelbasis mit mehreren sich überlappenden Prämissen erfolgt dann über ein Inferenzverfahren, das auf der Theorie der unscharfen Mengen basiert.

Die meisten statistischen Verfahren treffen eine Annahme über die Verteilungsdichtefunktionen der kontinuierlichen Merkmale in Abhängigkeit von den Umweltzuständen bezogen auf den gesamten Raum der Eingangsgrößen. Die Prämisse einer Fuzzy-Regel trifft dagegen nur eine Annahme für einen Teil des Raumes der Eingangsgrößen. Das Ergebnis des Entwurfs eines Fuzzy-Entscheidungssystems hängt daher wesentlich von der Qualität der entworfenen Prämissen der Fuzzy-Regeln ab.

Die Wahrscheinlichkeit $P(A_{l,i})$ der unscharfen Menge $A_{l,i}$ ergibt sich bei bekannter Verteilungsdichtefunktion des Merkmals $f_{x_l}(x_l)$ zu [9, 160]:

$$P(A_{l,i}) = \int_{-\infty}^{\infty} \mu_{A_{l,i}}(x_l) f_{x_l}(x_l) dx_l. \quad (2.47)$$

Für den Lerndatensatz ist die tatsächliche Verteilungsdichtefunktion f_{x_l} allerdings nicht bekannt. Die Wahrscheinlichkeit $P(A_{l,i})$ und damit eine Diskretisierung der Verteilungsdichtefunktion des Merkmals $f_{x_l}(x_l)$ wird dann durch Aufsummieren der Zugehörigkeiten

r	WENN	DANN
1	$x_1 = A_{1,3}$ ODER $A_{1,4}$	$C_1 = e_2$
2	$(x_1 = A_{1,3}$ ODER $A_{1,4}$ ODER $A_{1,5})$ UND $(x_2 = A_{2,1}$ ODER $A_{2,2})$	$C_2 = e_2$
3	$(x_1 = A_{1,4}$ ODER $A_{1,5})$ UND $(x_2 = A_{2,1}$ ODER $A_{2,2}$ ODER $A_{2,3})$	$C_3 = e_1$
4	$(x_1 = A_{1,1}$ ODER $A_{1,2})$ UND $(x_2 = A_{2,3}$ ODER $A_{2,4}$ ODER $A_{2,5})$	$C_4 = e_1$
5	$x_1 = A_{1,5}$	$C_5 = e_1$
6	$x_2 = A_{2,5}$	$C_6 = e_1$
	SONST	$C_7 = e_1$
$a_1 =$	(-5, -2.5, -1.6, 0, 3)	
$a_2 =$	(0, 1, 1.7, 2.4, 4)	

Tabelle 2.4: Einzelregeln zu Bild 2.7.

aller Elemente des Lerndatensatzes zu $A_{l,i}$ bestimmt:

$$\hat{P}(A_{l,i}) = \frac{1}{N} \sum_{n=1}^N \mu_{A_{l,i}}(x_l[n]). \quad (2.48)$$

Die bedingte Wahrscheinlichkeit $P(z_j|A_{l,i})$ des Umweltzustandes z_j im Bereich $A_{l,i}$ wird bei bekannten Verteilungsdichtefunktionen der kontinuierlichen Merkmale $f_{x_l}(x_l|z_j)$ über das Bayes-Theorem bestimmt:

$$P(z_j|A_{l,i}) = \frac{\int_{-\infty}^{\infty} \mu_{A_{l,i}}(x_l) f_{x_l}(x_l|z_j) dx_l \cdot P(z_j)}{\int_{-\infty}^{\infty} \mu_{A_{l,i}}(x_l) f_{x_l}(x) dx_l}. \quad (2.49)$$

Analog zu (2.48) ergibt sich die Berechnung aus dem Lerndatensatz durch Summieren der Zugehörigkeiten der Beispiele. In Anlehnung an die Darstellung des Bayes-Theorems für diskrete Ereignisse wird die Notation für Fuzzy-Mengen und Ereignisse vorgenommen:

$$\hat{P}(z_j|A_{l,i}) = \frac{\hat{P}(z_j \wedge A_{l,i})}{\hat{P}(A_{l,i})} = \frac{\sum_{n=1}^N (\mu_{A_{l,i}}(x_l[n]), \mu_{z_j}[n])_t}{\sum_{n=1}^N \mu_{A_{l,i}}(x_l[n])}. \quad (2.50)$$

Im Zähler ist ein Fuzzy-UND-Operator für die Kombination der Zugehörigkeit $\mu_{A_{l,i}}[n]$ des Beispiels zum Term $A_{l,i}$ und der Zugehörigkeit $\mu_{z_j}[n]$ zum Umweltzustand z_j zu verwenden. Der Operator $(*,*)_t$ entspricht dabei einer t-Norm [30, 79] und steht für einen beliebigen UND-Operator. Im Nenner steht wiederum die Summe der Zugehörigkeiten aller Beispiele zum Term $A_{l,i}$.

Für kontinuierliche $\mu_{z_j}[n] \in [0; 1]$ ist das Ergebnis der Schätzung der bedingten Wahrscheinlichkeiten abhängig vom verwendeten UND-Operator. Für die in der vorliegenden Arbeit betrachteten diskreten Zugehörigkeiten $\mu_{z_j}[n] \in \{0; 1\}$ ist das Ergebnis wegen

der Definitionsgleichungen der t-Norm [30] dagegen unabhängig vom verwendeten UND-Operator und (2.50) wird in der exemplarischen Schreibweise mit dem Produktoperator zu:

$$\hat{P}(z_j|A_{l,i}) = \frac{\sum_{n=1}^N \mu_{A_{l,i}}(x_l[n]) \cdot \mu_{z_j}[n]}{\sum_{n=1}^N \mu_{A_{l,i}}(x_l[n])}, \mu_{z_j}[n] \in \{0; 1\}. \quad (2.51)$$

Zur Bestimmung der Wahrscheinlichkeiten $\hat{P}(z_j|A_{l,i})$ wird (2.51) über alle Beispiele des Lerndatensatzes ausgewertet. Bild 2.8 zeigt im mittleren Teil eine grobe Diskretisierung für eine Eingangsgröße x_l am Beispiel von zwei Umweltzuständen z_1 und z_2 . Andere Stützpunkte der Zugehörigkeitsfunktionen können zu anderen Diskretisierungen führen. Im oberen Teil sind die Zugehörigkeitsfunktionen und im unteren Teil die bedingten Wahrscheinlichkeiten $\hat{P}(z_j|A_{l,i})$ der unscharfen Mengen dargestellt.

Über ähnliche Beziehungen wie (2.51) können auch die bedingten Wahrscheinlichkeiten der Umweltzustände für mehrere Terme oder eine gesamte Prämisse P_r berechnet werden. Dazu sind die Zugehörigkeiten der Beispiele zu der Prämisse $\mu_{P_r}(x_1, \dots, x_{N_x})$ zu bestimmen.

Besteht die Prämisse P_r aus ODER-Verknüpfungen benachbarter Terme $i_{min} \dots i_{max}$ eines Merkmals x_l , wird als ODER-Operator hier die beschränkte Summe verwendet:

$$\mu_{P_r,l}(x_l) = \min\left(1, \sum_{i=i_{min}}^{i_{max}} \mu_{A_{l,i}}(x_l)\right). \quad (2.52)$$

Für die UND-Verknüpfung von einzelnen Teilprämissen $\mu_{P_r,l}$ verschiedener Merkmale wird als UND-Operator hier das Produkt verwendet:

$$\mu_{P_r}(\mathbf{x}) = \prod_{l=1}^{N_x} \mu_{P_r,l}(\mathbf{x}). \quad (2.53)$$

Über (2.52) und (2.53) können die Zugehörigkeiten der Beispiele zu beliebigen Prämissen bestimmt werden. Für die Berechnung von \hat{P}_{P_r} und $\hat{P}_{z_j|P_r}$ sind in (2.48) und (2.51) die Zugehörigkeiten zu einzelnen Termen $A_{l,i}$ durch die Zugehörigkeiten zu P_r zu ersetzen. Bild 2.9 zeigt die ODER-Verknüpfung für benachbarte Terme. An der Stelle x gilt:

$$\mu_{A_{l,2}} = 0.7 \quad \mu_{A_{l,3}} = 0.3$$

Für die ODER-Verknüpfung von $A_{l,2}$ und $A_{l,3}$ ergibt sich dann:

$$\mu_{A_{l,2} \cup A_{l,3}} = \mu_{A_{l,2}} + \mu_{A_{l,3}} = 0.7 + 0.3 = 1$$

Die dicken Linien markieren in Bild 2.9 die Zugehörigkeiten zu den ODER-Verknüpfungen der unscharfen Mengen $A_{l,2}$ und $A_{l,3}$ sowie $A_{l,4}$ und $A_{l,5}$.

Die Schätzung der Wahrscheinlichkeiten der Umweltzustände im Gebiet einer Prämisse $\hat{P}_{z_j|P_r}$ hängt von der Gestaltung der Prämisse ab. Somit kann die Struktur der Schätzung

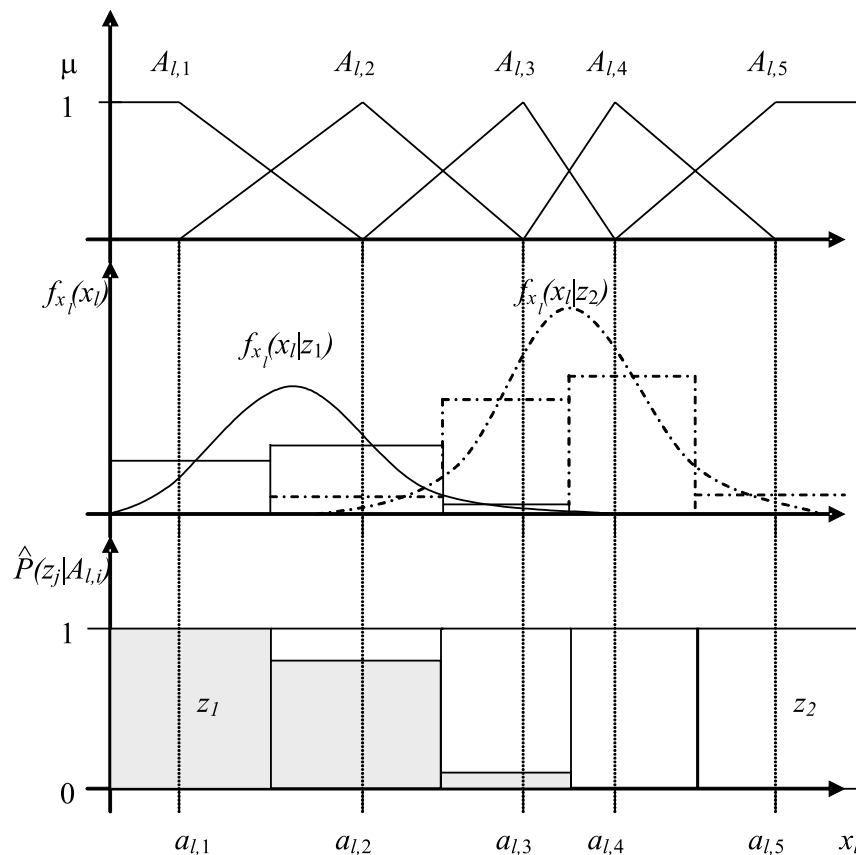


Bild 2.8: Oben: Zugehörigkeitsfunktionen zu den unscharfen Mengen $A_{l,i}$. Mitte: Verteilungsdichtefunktion $f_{x_l}(x_l|z_j)$ der Daten zweier Zustände und Diskretisierung bezüglich $A_{l,i}$. Unten: Bedingte Wahrscheinlichkeiten $\hat{P}(z_j|A_{l,i})$ für die unscharfen Mengen $A_{l,i}$.

der Wahrscheinlichkeiten für die Umweltzustände lokal beeinflusst werden. Die Bayes-Klassifikation mit Annahme einer Wahrscheinlichkeitsdichtefunktion erfordert dagegen die Annahme einer globalen Struktur für den gesamten Raum der Eingangsgrößen.

Bei Fuzzy-Systemen ergeben sich zwei mögliche Bedeutungen der Konklusionen der Regeln:

- **Schätzung der Umweltzustände:** Die Konklusion der Regel ist eine Schätzung der Wahrscheinlichkeiten der Umweltzustände. Aus den Regeln sind die Wahrscheinlichkeiten der Umweltzustände $\hat{P}(z_j|\mathbf{x})$ in Abhängigkeit von \mathbf{x} zu schätzen. Die beste Entscheidung wird über die erwarteten Kosten bestimmt (2.19). Dieser Weg entspricht dem Vorgehen bei der Bayes-Klassifikation.
- **Entscheidung mit maximaler Zugehörigkeit:** Die Konklusion ist die jeweils beste Entscheidung für die Beispiele im Gebiet der Prämisse nach (2.19). Über alle

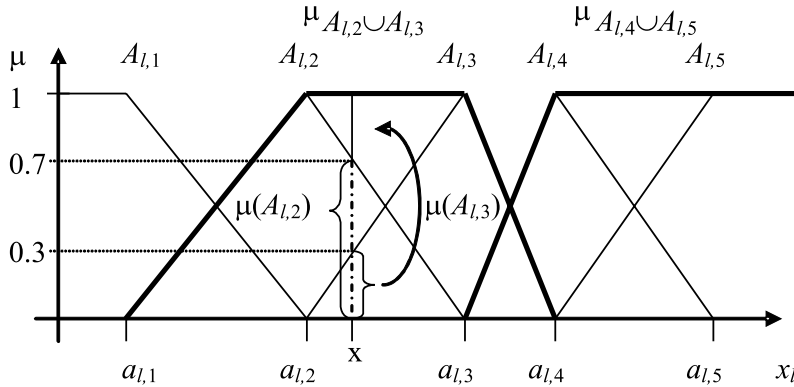


Bild 2.9: Beispiel für die ODER-Verknüpfung benachbarter Terme.

für \mathbf{x} aktivierten Regeln ($\mu_{P_r}(\mathbf{x}) > 0$) werden Zugehörigkeiten zu den Entscheidungen $\mu_{e_k}(\mathbf{x})$ bestimmt. Die Entscheidung mit der maximalen Zugehörigkeit wird getroffen.

Zur Veranschaulichung wird der in Bild 2.10 dargestellte Fall betrachtet. Es handelt sich um zufällig über zwei Gleichverteilungen erzeugte Daten für die Umweltzustände z_1 und z_2 , die durch die schwarzen Balken dargestellt sind. Für das Merkmal x sind die beiden unscharfen Mengen A_1 und A_2 definiert (unterer Teil von Bild 2.10). Diese repräsentieren auch gleichzeitig die beiden Prämissen P_1 und P_2 einer Regelbasis.

Es wird nun untersucht, welche Entscheidung in Abhängigkeit von x für die folgende Kostenmatrix getroffen wird:

$$\mathbf{L} = \begin{pmatrix} 0 & 1 \\ L_R & 0 \end{pmatrix}.$$

Dabei ist L_R eine Variable, die das Kostenverhältnis zwischen den beiden Fehlentscheidungen ausdrückt. Durch Auswertung von (2.51) über den Daten dieses Beispiels ergibt sich für die bedingten Wahrscheinlichkeiten $\hat{P}(z_j|A_i)$:

$$\hat{P}(z_j|A_i) = \begin{pmatrix} 0.85 & 0.15 \\ 0.15 & 0.85 \end{pmatrix}.$$

Zunächst wird die Entscheidung über die Schätzung der Wahrscheinlichkeit der Umweltzustände bestimmt. Aus dem Verlauf der Zugehörigkeitsfunktionen mit den Knickpunkten $[1, 4]$ ergibt sich über die beschränkte Summe als ODER-Operator die Schätzung der Wahrscheinlichkeiten der Zustände für ein gegebenes x :

$$\hat{P}(z_j|x) = \begin{pmatrix} 0.85 \cdot \mu_{A_1}(x) + 0.15 \cdot \mu_{A_2}(x) \\ 0.15 \cdot \mu_{A_1}(x) + 0.85 \cdot \mu_{A_2}(x) \end{pmatrix}.$$

Die Beschränkung der Summe kann hier weggelassen werden, da die Summe der Zugehörigkeitswerte für benachbarte Terme bei dem betrachteten Fuzzy-System nicht größer als 1 werden kann.

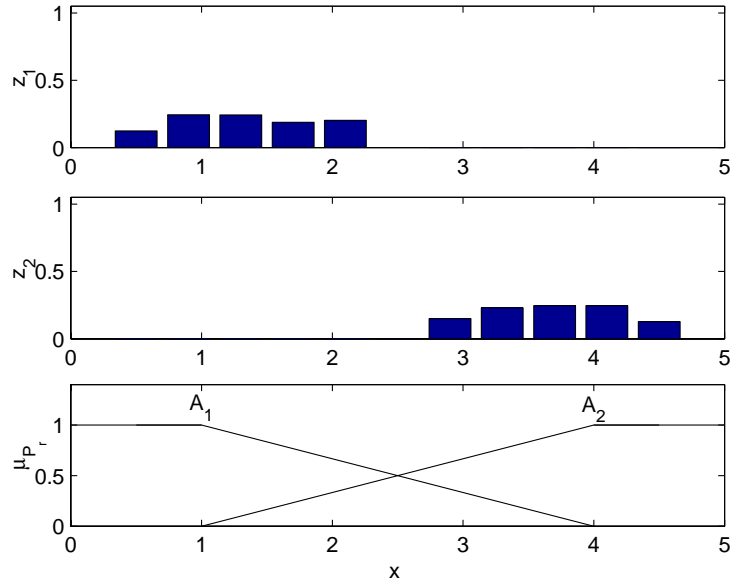


Bild 2.10: Beispielproblem mit zwei Umweltzuständen und zwei Prämissen.

Die beste Entscheidung kann nun mit der Kostenmatrix über (2.19) bestimmt werden:

$$e_{opt}(x) = \arg \min_{e_k} \sum_{j=1}^{N_j} L(e_k, z_j) \cdot \hat{P}(z_j|x). \quad (2.54)$$

Die durchschnittlichen Kosten pro Entscheidung können nach (2.12) bestimmt werden. In Bild 2.11 ist für variable Kostenverhältnisse L_R der Wert von x durch Punkte markiert, an dem die Entscheidung von e_1 nach e_2 kippt. Für alle Kostenverhältnisse oberhalb dieser Linie ist e_1 die beste Entscheidung und für alle Kostenverhältnisse unterhalb dieser Linie ist e_2 die beste Entscheidung.

Für ein Kostenverhältnis von $L_R > 5.67$ wird unabhängig von x stets die Entscheidung e_1 getroffen (gepunktete Linie in Bild 2.11). Für das gegebene Problem ist dies keine zufriedenstellende Lösung.

Die Schätzung von $\hat{P}(z_j|A_{l,i})$ wird nun modifiziert. Die Zugehörigkeit $\mu_{A_{l,i}}$ wird wie folgt bestimmt:

$$\mu_{A_{l,i}} = \begin{cases} 1 & \text{für } \mu_{A_{l,i}} \geq \mu_{A_{l,j}}, i < j \\ 0 & \text{für } \mu_{A_{l,i}} < \mu_{A_{l,j}}, i < j. \end{cases} \quad (2.55)$$

Dies entspricht einem α -Schnitt der unscharfen Mengen $A_{l,i}$ bei $\mu = 0.5$. Als α -Schnitt einer unscharfen Menge wird die scharfe Teilmenge, auf der die Zugehörigkeitsfunktion größer als α ist, bezeichnet. Für das Beispiel in Bild 2.10 gilt dann:

$$\begin{aligned} \mu_{A_1}(x) &= 1 \wedge \mu_{A_2}(x) = 0 & \text{für } x \leq 2.5 \\ \mu_{A_1}(x) &= 0 \wedge \mu_{A_2}(x) = 1 & \text{für } x > 2.5. \end{aligned} \quad (2.56)$$

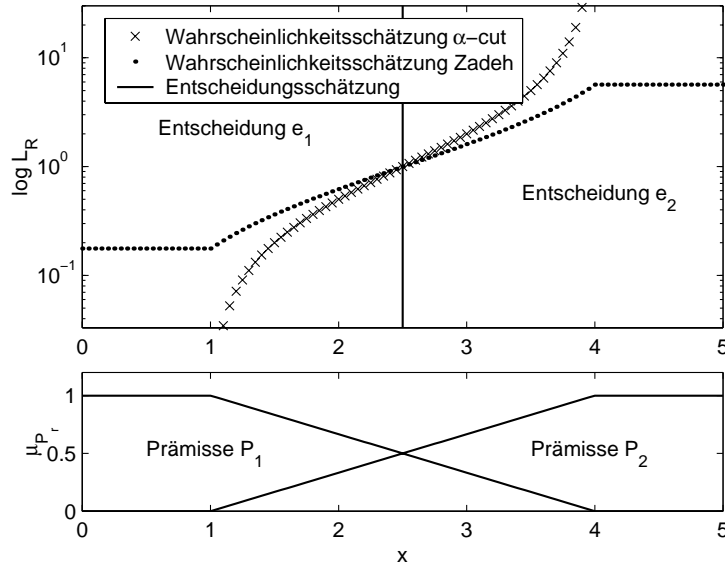


Bild 2.11: Entscheidungen in Abhängigkeit vom Kostenverhältnis L_R und dem Merkmal x .

r	WENN	DANN
1	$x = A_1$	$C_1 = e_1$
2	$x = A_2$	$C_2 = e_2$

Tabelle 2.5: Regeln zu Bild 2.10.

Für die bedingten Wahrscheinlichkeiten $\hat{P}(z_j|A_i)$ folgt:

$$\hat{P}(z_j|A_i) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Die Wahrscheinlichkeiten der Umweltzustände in Abhängigkeit von x sind dann:

$$\hat{P}(z_j|x) = \begin{pmatrix} 1 \cdot \mu_{A_1}(x) + 0 \cdot \mu_{A_2}(x) \\ 0 \cdot \mu_{A_1}(x) + 1 \cdot \mu_{A_2}(x) \end{pmatrix}.$$

Erneutes Auswerten von (2.54) führt zu den Werten von x , für die die Entscheidungen kippen. Diese sind für die Kostenverhältnisse durch Kreuze in Bild 2.11 markiert. Auch dieses Ergebnis ist nicht befriedigend. Für große Kostenverhältnisse wird für $x < 4$ stets e_1 entschieden. Dies führt zu einer großen Anzahl Fehlentscheidungen.

Nun werden die beiden Regeln aus Tabelle 2.5 betrachtet. Die Aktivierungen der Prämissen P_1 und P_2 entsprechen den Zugehörigkeiten zu den Entscheidungen e_1 und e_2 . Es wird die Entscheidung mit der maximalen Zugehörigkeit getroffen. Für obiges Beispiel bedeutet dies, dass unabhängig vom Kostenverhältnis L_R die Entscheidung stets an der selben Stelle $x = 2.5$ kippt. Die senkrechte Linie in Bild 2.11 drückt diesen Zusammenhang aus.

Der erstmals in dieser Arbeit vorgestellte Vergleich dieser Ansätze zeigt, dass die Nutzung der Prämissen von Fuzzy-Regeln als Wahrscheinlichkeitsschätzung der Umweltzustände mit darauf gestützter Entscheidung zu wenig plausiblen Ergebnissen führen kann. Daher wird im Folgenden die Konklusion einer Regel stets als Schätzung der besten Entscheidung für die Beispiele der Prämisse interpretiert.

Die Schätzung der Wahrscheinlichkeit der Umweltzustände im Gebiet einer Prämisse wird in Verbindung mit der Kostenmatrix zur Bestimmung der besten Entscheidungen für die Prämissen verwendet. Dabei kann sowohl die originale Aktivierung der Prämisse $\mu(P_r)$ oder ein α -Schnitt der Prämisse verwendet werden. Sofern nicht anders erwähnt, wird im Folgenden stets der α -Schnitt verwendet.

Die Entscheidung für einen vorliegenden Eingangsgrößenvektor \mathbf{x} wird über ein Inferenzverfahren bestimmt. Das Inferenzverfahren leitet die Entscheidung aus den Konklusionen der aktivierten Regeln ab. Die für \mathbf{x} aktivierten Regeln können Überlappungen und widersprüchliche Konklusionen haben (siehe Bild 2.7).

2.3.3 Fuzzy-Inferenz

Aufgabe des Inferenzverfahrens ist es, für eine gegebene Regelbasis und für gegebene Zugehörigkeitsfunktionen jedem \mathbf{x} eine Entscheidung zuzuweisen. Es ist zu unterscheiden, ob sich die Prämissen der Regeln in der Regelbasis überlappen oder nicht. Überlappen sich die Prämissen in der Regelbasis nicht, können die Zugehörigkeiten von Prämissen mit identischer Konklusion direkt zu Zugehörigkeiten zu den Entscheidungen e_k aufsummiert werden:

$$\mu_{e_k}(\mathbf{x}[n]) = \sum_{e(C_r)=e_k} \mu_{P_r}(\mathbf{x}[n]). \quad (2.57)$$

Die Entscheidung mit der höchsten Zugehörigkeit wird getroffen:

$$e_k(\mathbf{x}[n]) = \arg \max_{e_k} \mu_{e_k}(\mathbf{x}[n]). \quad (2.58)$$

Eine Regelbasis nach der in Abschnitt 2.3.1 vorgestellten Struktur ist dann frei von Überlappungen einzelner Prämissen, wenn für den gesamten Wertebereich der Eingangsgrößen gilt:

$$\sum_{r=1}^{N_r} \mu_{P_r}(\mathbf{x}) \leq 1. \quad (2.59)$$

Überlappungen liegen in den Bereichen vor, in denen die Summe der Zugehörigkeiten zu den Regeln größer als 1 ist:

$$\sum_{r=1}^{N_r} \mu_{P_r}(\mathbf{x}) > 1. \quad (2.60)$$

Liegt mindestens eine Überlappung vor, kann (2.57) nicht direkt ausgewertet werden. Überlappungen können für Regeln gleicher Konklusion, aber auch für Regeln unterschiedlicher Konklusion auftreten. In Bild 2.12 oben sind zwei Prämissen eingezeichnet, die sich nicht überlappen. Die beiden Regeln im unteren Teil von Bild 2.12 überlappen sich, weil beide die Terme $A_{l,3}$ und $A_{l,4}$ umfassen.

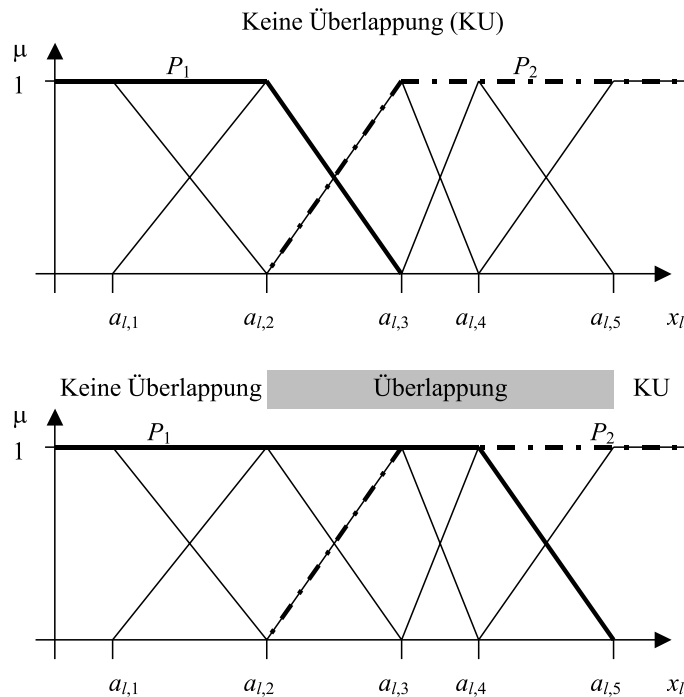


Bild 2.12: Zwei Regeln über ein Merkmal. Oben: ohne Überlappung, unten: mit Überlappung.

Um bei Regelbasen mit Überlappungen dennoch eindeutige Entscheidungen treffen zu können, ist ein Inferenzverfahren notwendig, das die Überlappungen erkennt und auflöst. Ein entsprechendes Inferenzverfahren für Fuzzy-Regeln wird in [104] vorgeschlagen. Dabei wird die Wichtung der aktivierten Regeln in den Überlappungsgebieten so modifiziert, dass (2.59) gilt. Die Zugehörigkeiten zu den Entscheidungen können dann gemäß (2.57) bestimmt werden. Das Verfahren weist jedem Eingangsgrößenvektor eine Entscheidung zu. Auch wenn die sich überlappenden Prämissen widersprüchliche Konklusionen haben, wird immer eine Entscheidung getroffen.

2.3.4 Modifizierte evidenzbasierte Inferenz bei Fuzzy-Systemen

Als Ergänzung zur Fuzzy-Inferenz wird in dieser Arbeit ein neues Inferenzverfahren für Fuzzy-Regeln auf der Basis der Evidenztheorie vorgeschlagen. Dieses Verfahren erlaubt es, für widersprüchliche Konklusionen keine Entscheidung zu treffen und somit eine Entscheidung außerhalb des Fuzzy-Systems anzufordern. Bei der Evidenztheorie treten an die Stelle von Wahrscheinlichkeiten Evidenzen zur Modellierung von Unsicherheit. Bei der

Wahrscheinlichkeitstheorie gilt für ein Experiment mit N verschiedenen Ausgängen ω_i :

$$P(\omega_i) \geq 0$$

$$\sum_{i=1}^N P(\omega_i) = 1.$$

Für ein Experiment mit zwei möglichen Ereignissen A und B bedeutet das, dass eine geringe Wahrscheinlichkeit für A eine starke Wahrscheinlichkeit für B zur Folge hat. Die Evidenztheorie sieht in einem schwachen Beweis für den Eintritt von Ereignis A dagegen keinen starken Beweis für den Eintritt von B . Dieses Konzept wird auf die Inferenz von Fuzzy-Regeln übertragen, indem die Konklusion einer Regel als Evidenz dafür gewertet wird, dass die Entscheidung e_{C_r} die beste ist.

Dempster's rule of combination dient dabei als Basis für die Inferenz von Fuzzy-Regeln. Über diese Regel können zwei Massefunktionen, die auf dem selben *frame of discernment* definiert sind, miteinander kombiniert werden [138]. Ein *frame of discernment* ist eine Menge elementarer, sich gegenseitig ausschließender, vollständiger Aussagen. Das Ergebnis der Kombination zweier Massefunktionen ist wiederum eine Massefunktion. Die Grundlagen der Evidenztheorie und Dempster's rule of combination sind in Anhang 6.1 zusammengestellt.

In [154] wird die Unsicherheit der Ausgangsgröße von Fuzzy-Modellen mittels der Evidenztheorie abgebildet. Einige weitere Beispiele für die Anwendung der Evidenztheorie sind: Fusion von Sensorinformationen [29, 116], Kombination der Ausgänge verschiedener Klassifikatoren zur Stimmerkennung [4] oder die Detektion von Schweißfehlern [75].

Der Gültigkeitsbereich von Dempster's rule of combination wird in [149, 162] diskutiert. Ein Verfahren zur Reduktion der Berechnungskomplexität für Dempster's rule of combination zeigt [12]. Beiträge zur Verbindung von Fuzzy-Systemen und der Evidenztheorie sind zum Beispiel [26, 27, 157]. In [27] werden ebenfalls regelbasierte Fuzzy-Systeme betrachtet. Die Konklusionen der Einzelregeln stellen dabei die Massefunktionen über der Menge der möglichen Zustände dar. Dies entspricht einer Interpretation der Fuzzy-Regeln als Schätzwert für die Umweltzustände. Im Gegensatz dazu wird hier ein Verfahren vorgeschlagen, das eine Massefunktion über den möglichen Entscheidungen nutzt. Damit wird die Erkenntnis der vorliegenden Arbeit, dass Konklusionen von Fuzzy-Regeln als Schätzungen für die besten Entscheidungen zu interpretieren sind, berücksichtigt. Aus der Aktivierung einer Regel werden die Werte der Massefunktion für die einzelnen Entscheidungen abgeleitet.

Liegt für ein Beispiel widersprüchliche oder unvollständige Information in der Regelbasis vor, kann es sinnvoll sein, keine Entscheidung zu treffen. Widersprüchliche Information ergibt sich durch benachbarte oder überlappende Regeln mit unterschiedlichen Konklusionen. Unvollständige Information liegt dann vor, wenn die SONST-Regel aktiviert ist. Steht zum Beispiel ein Experte zur Verfügung, der konstante Gesamtkosten pro Entscheidung $L_{T,E}$ verursacht, kann er für das Beispiel mit widersprüchlicher oder unvollständiger Information eine Entscheidung fällen. Die Befragung des Experten ist nach der μ -Regel sinnvoll, wenn dessen Kosten $L_{T,E}$ niedriger sind, als die erwarteten Kosten für eine Entscheidung des Beispiels.

Das in Abschnitt 2.3.3 vorgestellte Inferenzverfahren ist für die im letzten Absatz vorgestellte Aufgabe nur bedingt geeignet, da es einem Beispiel auch bei widersprüchlichen

Konklusionen der aktivierten Regeln eine Entscheidung zuweist.

Der *frame of discernment* Θ für die hier vorgeschlagene Inferenz von Fuzzy-Regeln ist die Menge aller möglichen Entscheidungen \mathcal{E} . Diese Menge erfüllt die Anforderungen an einen *frame of discernment*. Die Massefunktionen m_{R_r} der einzelnen Regeln sind über der Potenzmenge $2^{\mathcal{E}}$ zu definieren. Die Aktivierung der Prämisse P_r einer Regel μ_{R_r} wird dabei als Glaubensgrad $bel(e(P_r))$ aufgefasst, dass $e(P_r)$ die richtige Entscheidung ist:

$$bel(e(P_r)) = \mu_{P_r}. \quad (2.61)$$

Da $e(P_r)$ ein Element von Θ ist, folgt mit der Definition der Glaubensfunktion (siehe Anhang 6.1):

$$bel(e(P_r)) = \sum_{B \subseteq e(P_r)} m(B) = m(e(P_r)). \quad (2.62)$$

Dies folgt, da $e(P_r)$ keine Teilmengen B besitzt. Damit ist die Masse $m_{R_r}(e(P_r))$ identisch zum Glaubensgrad $bel(e(P_r))$:

$$m_{R_r}(e(P_r)) = bel(e(P_r)). \quad (2.63)$$

Über den Bereich, der nicht von der Prämisse P_r erfasst wird, enthält die Regel keine einschränkende Aussage. Daher wird die restliche Masse der ODER-Verknüpfung aller Entscheidungen zugewiesen:

$$m_{R_r}(e_1 \vee e_2 \vee \dots \vee e_{N_e}) = 1 - m_{R_r}(e(P_r)). \quad (2.64)$$

Für die Kombination der Massefunktionen zweier Regeln R_1 und R_2 ist Dempster's rule of combination wie folgt zu berechnen:

$$m_{R_{1,2}}(e^*) = m_{R_1} \oplus m_{R_2} = \frac{\sum_{\substack{i,j \\ \tilde{e}_i \cap \tilde{e}_j = e^*}} m_{R_1}(\tilde{e}_i) \cdot m_{R_2}(\tilde{e}_j)}{1 - \underbrace{\sum_{\substack{i,j \\ \tilde{e}_i \cap \tilde{e}_j = \emptyset}} m_{R_1}(\tilde{e}_i) \cdot m_{R_2}(\tilde{e}_j)}_W}. \quad (2.65)$$

Dabei gilt: $\tilde{e}, e^* \subseteq \Theta$. Bild 2.13 zeigt grafisch die Anwendung von Dempster's rule of combination am Beispiel von den Massefunktionen zweier Regeln mit dem *frame of discernment* $\{e_1; e_2\}$. Die Potenzmenge über die möglichen Entscheidungen ist dann: $2^{\mathcal{E}} = \{\emptyset, e_1, e_2, e_1 \vee e_2\}$. Die Zahlenwerte der kombinierten Massefunktionen m_1, m_2 und der resultierenden $m_1 \oplus m_2$ zeigt Tabelle 2.6. Für $m(e_1)$ ergibt sich für den linken Teil von Bild 2.13:

$$\begin{aligned} m_{1 \oplus 2}(e_1) &= \frac{m_1(e_1) \cdot m_2(e_1) + m_1(e_1 \vee e_2) \cdot m_2(e_1) + m_1(e_1) \cdot m_2(e_1 \vee e_2)}{1 - (m_1(e_1) \cdot m_2(e_2) + m_1(e_2) \cdot m_2(e_1))} \\ &= \frac{\frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot \frac{1}{3}}{1 - (\frac{1}{2} \cdot \frac{2}{3} + 0 \cdot 0)} = \frac{1}{4}. \end{aligned}$$

Die Flächen in Bild 2.13 kennzeichnen die Masse der einzelnen Entscheidungen nach der Kombination. Die Masse der Entscheidung $e_1 \vee e_2$ nimmt sowohl bei widersprüchlichen als

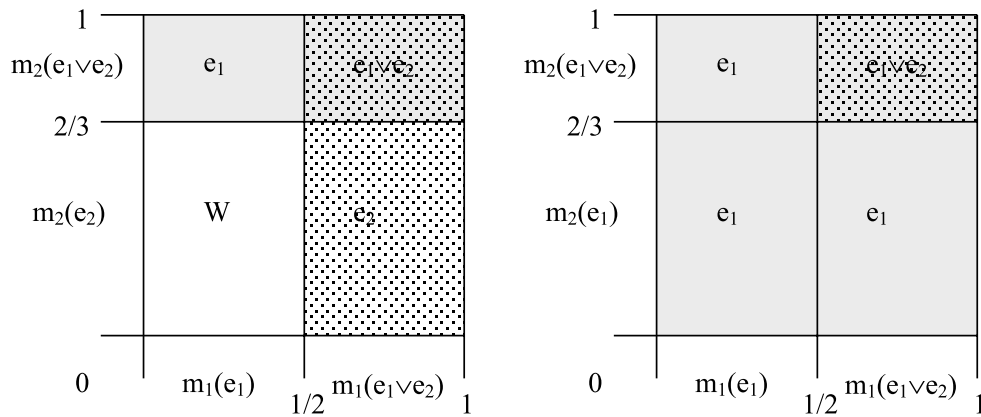


Bild 2.13: Visualisierung der Anwendung von Dempster's rule of combination für zwei Regeln. Links: Verschiedene Konklusionen, rechts: identische Konklusionen.

		e_1	e_2	$e_1 \vee e_2$	W
Links	m_1	$\frac{1}{2}$	0	$\frac{1}{2}$	—
	m_2	0	$\frac{2}{3}$	$\frac{1}{3}$	—
	$m_1 \oplus m_2$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{3}$
rechts	m_1	$\frac{1}{2}$	0	$\frac{1}{2}$	—
	m_2	$\frac{2}{3}$	0	$\frac{1}{3}$	—
	$m_1 \oplus m_2$	$\frac{5}{6}$	0	$\frac{1}{6}$	0

Tabelle 2.6: Zahlenwerte zu Bild 2.13.

auch bei identischen Konklusionen ab. Der widersprüchliche Bereich bei unterschiedlichen Konklusionen ist im linken Teil von Bild 2.13 mit W gekennzeichnet.

Betrachtet werden nun zwei Fuzzy-Regeln mit unterschiedlichen Konklusionen e_1 und e_2 . Die Regeln beziehen sich auf ein Merkmal x . Die Massefunktionen für die beiden Regeln sind:

$$m_{R_1}(e_1) = \{0, \mu_{R_1}(x), 0, 1 - \mu_{R_1}(x)\} \quad (2.66)$$

$$m_{R_2}(e_2) = \{0, 0, \mu_{R_2}(x), 1 - \mu_{R_2}(x)\} \quad (2.67)$$

Durch Kombination nach Dempster's rule of combination ergibt sich der Verlauf der Massefunktion für die einzelnen Entscheidungen e_1 , e_2 oder $e_1 \vee e_2$. Für die Masse der Entscheidung e_1 ergibt sich:

$$m_{R_{1,2}}(e_1) = \frac{\mu(P_1) \cdot (1 - \mu(P_2))}{1 - \mu(P_1) \cdot \mu(P_2)}, \quad m_{R_{1,2}}(e_2) = \frac{(1 - \mu(P_1)) \cdot \mu(P_2)}{1 - \mu(P_1) \cdot \mu(P_2)}. \quad (2.68)$$

Für Gebiete mit:

$$\mu(P_1) = \mu(P_2) = 1,$$

sind beide Regeln vollständig aktiviert. In diesem Fall sind $m_{R_{1,2}}(e_1)$ und $m_{R_{1,2}}(e_2)$ wegen einer Beziehung $\frac{0}{0}$ nicht definiert. Eine Grenzwertbetrachtung liefert jedoch:

$$\lim_{\mu(P_r) \rightarrow 1} \left(\frac{\mu(P_r) - \mu(P_r)^2}{1 - \mu(P_r)^2} \right) = \lim_{\mu(P_r) \rightarrow 1} \left(\frac{1 - 2 \cdot \mu(P_r)}{-2 \cdot \mu(P_r)} \right) = \frac{1}{2}. \quad (2.69)$$

In Bild 2.14 sind die Massefunktionen und der Widerspruch W für die in Tabelle 2.7 gelisteten Regeln dargestellt.

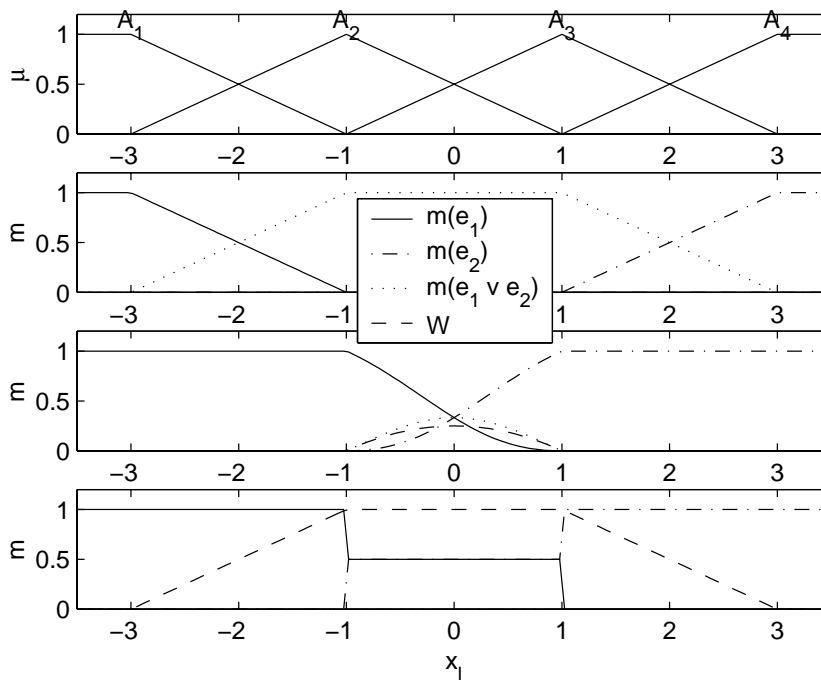


Bild 2.14: Verlauf der Massefunktionen für die Regeln in Tabelle 2.7. Oben: die Regeln überlappen sich nicht. Mitte: die Regeln sind benachbart. Unten: die Regeln überlappen sich. Durchgezogen: $m(e_1)$, strichpunktiert: $m(e_2)$, gepunktet: $m(e_1 \vee e_2)$, gestrichelt: W .

Der Wert der Massefunktion für $e_1 \vee e_2$ bedeutet dabei die Stärke des Glaubens daran, dass die Rückweisungsentscheidung die beste ist. Ob eine scharfe Entscheidung e_1 oder e_2 getroffen werden soll, kann vom Betrag der Ungewissheit $U = m(e_1 \vee e_2)$ oder vom Widerspruch W abhängig gemacht werden. Je nach Niveau der maximal zulässigen Ungewissheit U_{max} oder des maximalen Widerspruchs W_{max} ergibt sich ein Bereich, in dem die Rückweisungsentscheidung getroffen wird.

Bei mehr als zwei Regeln unterschiedlicher Konklusion ist die Reihenfolge der Kombination über Dempster's rule of combination wegen der Gültigkeit des Assoziativgesetzes

	r	WENN	DANN
Oben	1	$x = A_1$	$C_1 = e_1$
	2	$x = A_4$	$C_2 = e_2$
Mitte	1	$x = A_1$ ODER A_2	$C_1 = e_1$
	2	$x = A_3$ ODER A_4	$C_2 = e_2$
Unten	1	$x = A_1$ ODER A_2 ODER A_3	$C_1 = e_1$
	2	$x = A_2$ ODER A_3 ODER A_4	$C_2 = e_2$

Tabelle 2.7: Regeln zu Bild 2.14.

	Fuzzy-Inferenz		Evidenz-Inferenz	
	z_1	z_2	z_1	z_2
e_1	56	38	47	33
e_2	4	262	3	219
$e_1 \vee e_2$	0	0	10	48

Tabelle 2.8: Vergleich der Anzahl korrekter Entscheidungen und Fehlentscheidungen für die Regelbasis aus Tabelle 2.4 (Fuzzy-Inferenz und Evidenz-Inferenz). Zeilen: getroffene Entscheidung, Spalten: Umweltzustand.

unerheblich. Die Massefunktion bei mehr als zwei Regeln wird iterativ berechnet:

$$m_{RB} = m_{R_{N_R}} \oplus (m_{R_{N_R-1}} \oplus (\dots \oplus (m_{R_2} \oplus m_{R_1}))) \quad (2.70)$$

Aus der Massefunktion m_{RB} für die gesamte Regelbasis und dem bei der Kombination der Regeln maximal auftretenden Widerspruch $W_m[n]$ werden die Entscheidungen abgeleitet. Dabei wird allen Beispielen die Rückweisungsentscheidung zugewiesen, für die gilt:

$$1 - \sum_{i=1}^{N_e} m(e_i)[n] = U[n] > U_{max} \quad (2.71)$$

$$W_m[n] > W_{max} . \quad (2.72)$$

Für alle anderen Beispiele wird die Entscheidung mit der maximalen Masse getroffen:

$$e_{Ev}[n] = \max_{e_i} m(e_i)[n] . \quad (2.73)$$

Die Evidenz-Inferenz ignoriert die SONST-Regel, sofern diese in der Regelbasis enthalten ist. Die Auswertung der Regelbasis aus Tabelle 2.4 für den Beispieldatensatz mittels der Evidenz-Inferenz und den Parametern $U_{max} = W_{max} = 0.5$ ist in Bild 2.15 dargestellt.

Im Gegensatz zur Fuzzy-Inferenz existieren nun Gebiete in der x_1, x_2 -Ebene, für die keine Entscheidung ($e_1 \vee e_2$) getroffen wird (Bild 2.15). Diese Gebiete entstehen, wenn sich widersprüchliche Regeln überlappen. Dies ist bei den Regeln R_1 und R_6 sowie R_1 und R_3 der Fall. Im Gebiet, das nicht von einer Regel abgedeckt wird, wird ebenfalls keine Entscheidung getroffen. Beispiele im Gebiet der SONST-Regel werden somit nicht entschieden. Eine Betrachtung der getroffenen Entscheidungen in Tabelle 2.8 zeigt, dass die Zahl

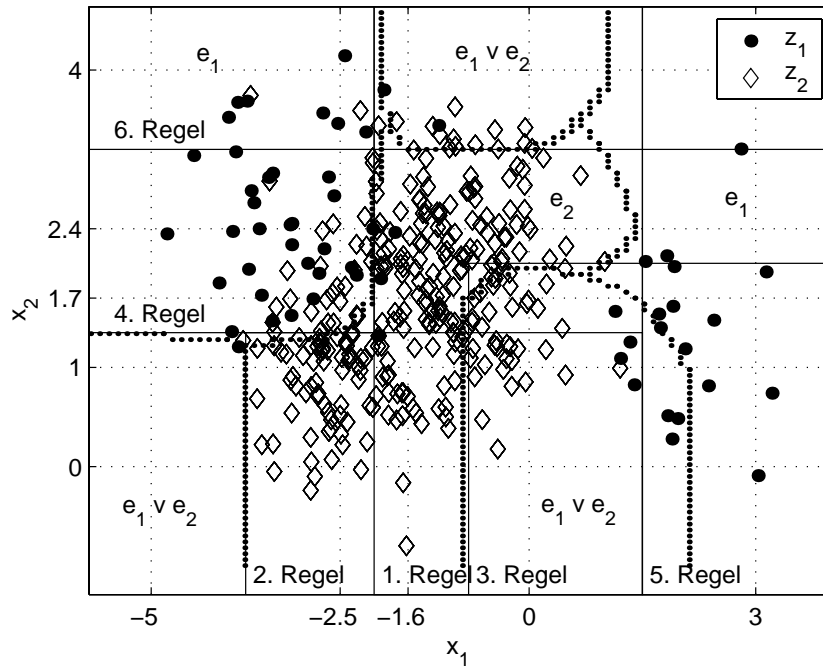


Bild 2.15: Regelbasis für den Beispieldatensatz über den Merkmalen x_1 und x_2 . Gepunktete Geraden: $\mu_{A_{l,i}} = 1$, Durchgezogene Linien: Näherung für $\mu_{P_r} = 0.5$. Gepunktete Kurven: Begrenzungen der Gebiete der Entscheidungen. Auswertung der Regeln aus Tabelle 2.4 mittels der Evidenz-Inferenz ($U_{max} = W_{max} = 0.5$).

der Fehlentscheidungen bei der Evidenz-Inferenz niedriger ist als bei der Fuzzy-Inferenz. Dafür wird für einen Teil der Beispiele keine Entscheidung getroffen.

Es stehen nun zwei Inferenzverfahren für regelbasierte Fuzzy-Systeme zur Verfügung: die Inferenz nach [104] und die in der vorliegenden Arbeit neu entwickelte Evidenz-Inferenz. Beide Inferenzverfahren können für die selben Regelbasen angewandt werden. In den nächsten beiden Abschnitten werden Bewertungsmaße für Fuzzy-Regeln und Fuzzy-Regelbasen entwickelt, die als Grundlage für die Entwurfsverfahren für Fuzzy-Systeme in Kapitel 3 dienen.

2.4 Neue Bewertungsmaße für Fuzzy-Einzelregeln

2.4.1 Optimale Fuzzy-Einzelregel

Eine Einzelregel besteht aus der Prämisse und ihrer Konklusion. Die Prämisse deckt in der Regel nicht den gesamten Wertebereich der Eingangsgrößen ab. Eine Ausnahme ist eine Regelbasis, die ausschließlich aus der SONST-Regel mit einer Default-Entscheidung besteht. Diese SONST-Regel deckt dann in dem besonderen Fall den gesamten Wertebereich ab.

Eine Einzelregel teilt den Lerndatensatz in zwei Teile mit graduellen Zugehörigkeiten: die Beispiele in der Prämisse und die restlichen Beispiele. Für die Beispiele in der Prämisse liegt eine Risikosituation vor, da die Wahrscheinlichkeiten der Umweltzustände nach (2.50) bestimmt sind. Für die restlichen Beispiele können die Wahrscheinlichkeiten der Umweltzustände ebenfalls über (2.50) bestimmt werden. Jedoch ist ungewiss, welche weiteren Prämissen die restlichen Beispiele teilweise oder vollständig erfassen.

Da eine Regelbasis aus möglichst wenig Regeln bestehen soll, sind die Prämissen der einzelnen Regeln möglichst weit zu generalisieren. Eine optimale Einzelregel erfasst alle Beispiele, für die die Entscheidung $e(P_r)$ die beste ist. Bei Problemen mit nicht kompakten Umweltzuständen kann es nötig sein, mehrere Regeln mit der selben Konklusion in die finale Regelbasis aufzunehmen. Bei einem nicht kompakten Umweltzustand sind dessen Beispiele nicht in einem zusammenhängenden Bereich des Merkmalsraumes konzentriert, sondern auf zwei oder mehrere Bereiche verteilt. Der Umweltzustand z_1 des Beispieldatensatzes ist nicht kompakt. Der Umweltzustand z_2 dagegen ist kompakt (siehe zum Beispiel Bild 2.15). Ob für einen Umweltzustand eine oder mehrere Regeln notwendig sind, ist aber bei der Bewertung der einzelnen Regel nicht bekannt. Es werden Regeln gesucht, die möglichst alle Beispiele einer Klasse abdecken, jedoch gleichzeitig so wenig Fehlentscheidungen wie möglich verursachen.

Für die entscheidungstheoretische Bewertung einer Fuzzy-Regel werden mehrere Ansätze vorgeschlagen und verglichen:

- Eine Einzelregel stellt eine Regelbasis dar. Die Restprämisse wird dabei ebenfalls als Risikogebiet betrachtet (Abschnitt 2.4.2).
- Nur die Beispiele in der Prämisse tragen zur Bewertung bei. Für die restlichen Beispiele sind andere Regeln zuständig (Abschnitt 2.4.3).
- Die Bewertung basiert auf den Beispielen in der Prämisse und auf den restlichen Beispielen von z_{C_r} , sofern für diese eine Fehlentscheidung getroffen werden kann (Abschnitt 2.4.4).

Tabelle 2.9 fasst die Varianten der Behandlung der Restprämisse zusammen. Die Beispiele in der Restprämisse werden dabei in zwei Gruppen aufgeteilt. Für die eine Gruppe ist $e(P_r)$ ebenfalls die beste Entscheidung, die Beispiele werden mit $(z_j = z_{C_r})$ bezeichnet. Für die zweite Gruppe ist $e(P_r)$ nicht die beste Entscheidung, für diese Beispiele gilt: $(z_j \neq z_{C_r})$. Bei der Bewertung als Regelbasis wird für alle Beispiele der Restprämisse eine Risikosituation angenommen. Bei der Bewertung über die Prämisse werden die Beispiele der Restprämisse vernachlässigt und als von anderen Regeln korrekt entschieden angenommen. Für die dritte Strategie sind die Gruppe der Beispiele z_{C_r} weiter zu unterteilen. Für einen Teil der Beispiele können Fehlentscheidungen getroffen werden und für den Rest nicht. In die Bewertung geht nur der Teil der Beispiele ein, für die Fehlentscheidungen getroffen werden können.

Die drei Ansätze werden anhand der Bewertung der Regeln aus Tabelle 2.10 verglichen. Im Vergleich zur optimalen Einzelregel ist Regel R_1 zu speziell und Regel R_2 zu stark generalisiert. Von den vier Regeln stellt R_4 den besten Kompromiss dar. Zur Demonstration besonderer Effekte werden zusätzliche Beispiele eingeführt. Das beste Verfahren wird ausgewählt und für die Bewertung von Einzelregeln verwendet.

	Restprämisse	
	$z_j = z_{C_r}$	$z_j \neq z_{C_r}$
Regelbasis	Risiko	Risiko
Prämisse	0	0
	Fehler möglich	
	ja	nein
Prämisse + z_{C_r}	Risiko	0

Tabelle 2.9: Varianten für die Berücksichtigung der Beispiele in der Restprämisse.
0: Die Beispiele dieser Gruppe werden bei der Bewertung nicht berücksichtigt.

2.4.2 Bewertung als Regelbasis

Bei dieser Bewertungsmethode wird zur Bewertung aus einer einzelnen Regel R_r eine vollständige Regelbasis mit zwei Regeln erzeugt und bewertet. Die Prämissen der Regelbasis sind die Prämisse P_r der Regel R_r und deren Negation \bar{P}_r . Den Prämissen werden die Konklusionen $C_r = e(P_r)$ und $C_{\bar{r}} = e(\bar{P}_r)$ zugeordnet. Die Regelbasis deckt damit den gesamten Wertebereich der Eingangsgrößen überlappungsfrei ab. Die Konklusion der Restprämisse $C_{\bar{r}}$ ist dabei die beste Entscheidung für die nicht von der Prämisse erfassten Beispiele (2.54). Bei dieser Bewertung werden die Beispiele außerhalb der Prämisse P_r ebenfalls als Entscheidungen bei Risiko betrachtet. Dies entspricht der Erwartung, dass keine weiteren Regeln existieren, die die restlichen Beispiele besser entscheiden können. Damit wird (2.12) zu:

$$\hat{L}_{D, RB} = \sum_{j=1}^{N_z} \left(L(e(P_r), z_j) \cdot \underbrace{\hat{P}(P_r \wedge z_j)}_{\hat{P}(P_r|z_j) \cdot \hat{P}(z_j)} + L(e(\bar{P}_r), z_j) \cdot \underbrace{\hat{P}(\bar{P}_r \wedge z_j)}_{\hat{P}(\bar{P}_r|z_j) \cdot \hat{P}(z_j)} \right). \quad (2.74)$$

Da lediglich zwei Prämissen existieren, gilt:

$$\hat{P}(\bar{P}_r|z_j) = 1 - \hat{P}(P_r|z_j). \quad (2.75)$$

Damit kann (2.74) wie folgt geschrieben werden:

$$\hat{L}_{D, RB} = \sum_{j=1}^{N_z} \left(L(e(P_r), z_j) \cdot \hat{P}(P_r|z_j) + L(e(\bar{P}_r), z_j) \cdot \hat{P}(\bar{P}_r|z_j) \right) \cdot \hat{P}(z_j). \quad (2.76)$$

Diese Funktion entspricht für $N_z = 2$ einer Ebene im $\hat{P}(\bar{P}_r|z_{C_{\bar{r}}}), (\hat{P}(P_r|z_{C_r}), \hat{L}_D(R_r))$ -Raum. Dabei steht z_{C_r} für die Umweltzustände, für die $e(P_r)$ die beste Entscheidung ist. Die Höhenlinien haben in der $(\hat{P}(P_r|z_{C_r}), \hat{P}(\bar{P}_r|z_{C_{\bar{r}}}))$ -Ebene die selbe Steigung m wie die Geraden, an denen die Entscheidung kippt (siehe Abschnitt 2.2.3).

Die in Bild 2.16 links eingezeichneten vier Regeln sollen bewertet werden. Tabelle 2.10 enthält eine Darstellung der Regeln in Textform. Bild 2.16 rechts zeigt die selben Regeln als Punkte im Bewertungsgraph (siehe Abschnitt 2.2.3). Zusätzlich sind die Höhenlinien gleicher Kosten für die Bewertung eingezeichnet. Der Punkt für Regel 1 ist bei (0.99,0.083)

r	WENN	DANN
1	$x_1 = A_{12}$ UND $x_2 = A_{25}$	$C_1 = e_1$
2	$x_1 = A_{12}$	$C_2 = e_1$
3	$x_2 = A_{25}$	$C_3 = e_1$
4	$x_1 = A_{12}$ UND $x_2 = A_{24}$ ODER A_{25}	$C_4 = e_1$

Tabelle 2.10: Einzelregeln zu Bild 2.16.

ingezeichnet. Über die Prämisse werden ca. 8% der Beispiele von z_1 und ca. 1% der Beispiele von z_2 erfasst. In der negierten Prämisse liegen ca. 99% der Beispiele von z_2 und 92% der Beispiele von z_1 . Die Beschriftung der Abszisse $\hat{P}(\bar{P}_r|z \neq z_{C_r})$ entspricht dabei der Erkennungsrate der negierten Prämisse (\bar{P}_r) für die Beispiele, deren Umweltzustand verschieden ist zu dem Zustand der Beispiele, für die die Prämisse die beste Entscheidung trifft ($z \neq z_{C_r}$).

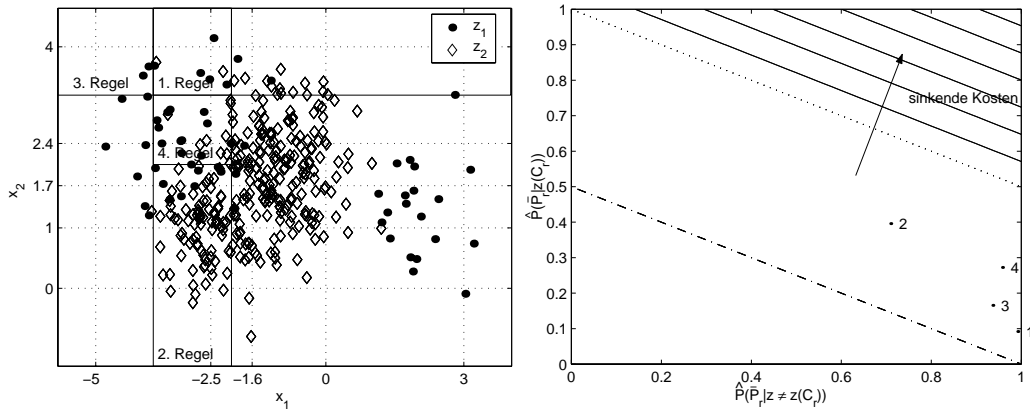


Bild 2.16: Links: Darstellung der Regeln in x_1 und x_2 . Rechts: Darstellung der Regeln aus Tabelle 2.10 im Bewertungsgraph für $L_R = 10$, $\hat{P}(z_1) = \frac{1}{6}$, $\hat{P}(z_2) = \frac{5}{6}$.

Die Zahlen können durch Abzählen der Punkte und Rhomben für Regel 1 in Bild 2.16 links verifiziert werden. In der Prämisse von Regel 1 befinden sich fünf Beispiele von z_1 und zwei Beispiele von z_2 . Damit ergibt sich:

$$\hat{P}(P_r|z_{C_r}) = \frac{5}{60} = 0.833 \quad , \quad \hat{P}(\bar{P}_r|z \neq z_{C_r}) = \frac{298}{300} = 0.993 \quad (2.77)$$

Für das gewählte Kostenverhältnis von $L_R = 10$ liegen die vier Regeln im Bewertungsgraph zwischen den Geraden, an denen die Entscheidungen kippen. Daraus folgt, dass für die Prämisse und die negierte Prämisse die selbe Entscheidung e_1 zu treffen ist. Die vier Regeln erhalten daher bei der Bewertung als Regelbasis alle die selben erwarteten Kosten pro Entscheidung. Diese Kosten sind zusätzlich mit denen der Default-Entscheidung identisch.

Zur Veranschaulichung eines weiteren Nachteils dieser Bewertung wird ein Problem mit drei Umweltzuständen $\{z_1, z_2, z_3\}$ betrachtet. Für die Entscheidungen e_1, e_2 und e_3 für ein Beispiel des Umweltzustandes z_3 gelte: $(L(e_3, z_3) \leq L(e_1, z_3) \leq L(e_2, z_3))$. Es

existiere eine Prämisse P , für die e_1 und für deren Negation \bar{P} e_2 die beste Entscheidung ist. Wird die Bewertung über (2.74) vorgenommen, ist das Ergebnis umso besser, je mehr Beispiele des Umweltzustandes z_3 im Gebiet der Prämisse liegen. Das Ziel, Regeln zu erzeugen, deren Prämissen möglichst frei von Fehlentscheidungen sind, kann somit nicht erreicht werden.

Die Bewertung von Einzelregeln als Regelbasis führt zu nicht plausiblen Ergebnissen und ist somit nicht geeignet.

2.4.3 Bewertung mit den Beispielen der Prämisse

Alternativ wird die Bewertung von Einzelregeln über die Beispiele in der Prämisse mit den Kosten pro Entscheidung L_{D,P_r} untersucht. Es wird von der optimistischen Erwartung ausgegangen, dass alle Beispiele außerhalb der Prämisse von anderen Regeln korrekt entschieden werden. Für die Schätzung der Wahrscheinlichkeiten der Umweltzustände in der Prämisse werden die Beispiele berücksichtigt, für die gilt:

$$\mu_{P_r}[n] > 0.$$

Die Zugehörigkeiten der Beispiele zur Prämisse können optional über einen α -Schnitt nach (2.55) auf die Menge:

$$\mu_{P_r}[n] \in \{0, 1\}, \quad (2.78)$$

abgebildet werden. Die Kosten ergeben sich zu:

$$L_{D,P_r} = \sum_{j=1}^{N_z} L(e(P_r), z_j) \cdot \hat{P}(z_j|P_r) \quad (2.79)$$

Über das Bayes-Theorem lässt sich (2.79) umschreiben zu:

$$L_{D,P_r} = \left[\sum_{j=1}^{N_z} L(e(P_r), z_j) \cdot \hat{P}(P_r|z_j) \cdot \hat{P}(z_j) \right] \cdot \frac{1}{\hat{P}(P_r)} \quad (2.80)$$

Für zwei Umweltzustände können die Höhenlinien von (2.80) mit (2.75) in der $(\hat{P}(P_r|z_{C_r}), \hat{P}(\bar{P}_r|z_{C_r}))$ -Ebene dargestellt werden. Die Regeln aus Tabelle 2.10 und die Linien konstanter Kosten sind in Bild 2.17 dargestellt.

Der wesentliche Nachteil dieser Bewertung ist, dass alle Regeln derselben Konklusion unabhängig von deren Größe identisch bewertet werden, wenn für die Beispiele der Prämisse keine Fehlentscheidung getroffen wird. Keine Fehlentscheidung in der Prämisse bedeutet:

$$\begin{aligned} \hat{P}(P_r \wedge z_j) &= 0, z_j \neq z_{C_r} \\ \hat{P}(\bar{P}_r|z_j) &= 1, z_j \neq z_{C_r}. \end{aligned} \quad (2.81)$$

Nun wird (2.80) unter der Annahme, dass die der Prämisse zugeordnete Entscheidung nur für einen Umweltzustand optimal ist, umgeformt zu:

$$\begin{aligned} L_{D,P_r} &= \left[L(e(P_r), z_{C_r}) \cdot \hat{P}(P_r|z_{C_r}) \cdot \hat{P}(z_{C_r}) + \right. \\ &\quad \left. \sum_{z_j \neq z_{C_r}} L(e(P_r), z_j) \cdot \hat{P}(P_r|z_j) \cdot \hat{P}(z_j) \right] \cdot \frac{1}{\hat{P}(P_r)} \end{aligned} \quad (2.82)$$

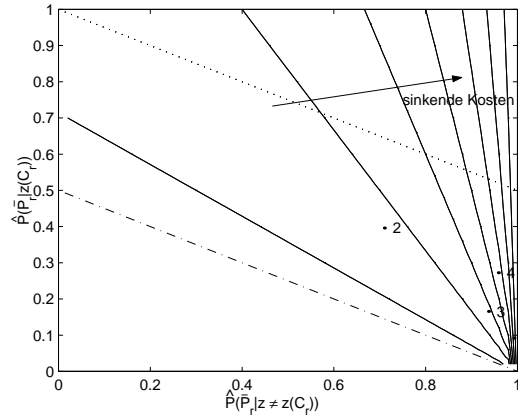


Bild 2.17: Darstellung der Regelbewertung mit Höhenlinien (nur Prämisse).

Mit (2.75) und (2.81) ergibt sich:

$$L_{D,P_r} = \left[L(e(P_r), z_{C_r}) \cdot \hat{P}(P_r|z_{C_r}) \cdot \hat{P}(z_{C_r}) \right] \cdot \frac{1}{\hat{P}(P_r)} \quad (2.83)$$

Für $\hat{P}(P_r)$ kann nach dem Satz der totalen Wahrscheinlichkeit auch geschrieben werden:

$$\hat{P}(P_r) = \sum_{j=1}^{N_z} \hat{P}(P_r \wedge z_j). \quad (2.84)$$

Da in der Prämisse jedoch kein Fehler auftreten soll, gilt:

$$\hat{P}(P_r) = \hat{P}(P_r \wedge z_{C_r}) = \hat{P}(P_r|z_{C_r}) \cdot \hat{P}(z_{C_r}). \quad (2.85)$$

Damit vereinfacht sich (2.83) zu:

$$L_{D,P_r} = L(e(P_r), z_{C_r}). \quad (2.86)$$

Die Kosten der Prämisse einer fehlerfreien Regel sind damit unabhängig von dem Anteil an den Beispielen des Umweltzustandes z_{C_r} , die von der Prämisse erfasst werden.

Bei dieser Bewertung verursacht die Regel 1 die geringsten Kosten. Für eine Regelbasis ist aber in den meisten Fällen die Regel 4 besser geeignet, da sie deutlich mehr Beispiele von z_1 erfasst.

Die Bewertung über die Beispiele in der Prämisse führt somit ebenfalls zu nicht zufriedenstellenden Ergebnissen.

2.4.4 Bewertung mit den Beispielen der Prämisse und den Fehlerkandidaten

Die dritte Variante zur Bewertung von Einzelregeln berücksichtigt die Beispiele in der Prämisse und die restlichen Beispiele des Umweltzustandes z_{C_r} , die nicht von der Prämisse erfasst werden. Fehlentscheidungen für Beispiele, für die $e(P_r)$ nicht die beste Entscheidung

ist, werden dadurch im Gebiet der negierten Prämisse nicht berücksichtigt. Dies entspricht der optimistischen Einstellung, dass eine andere Regel für diese Beispiele die richtige Entscheidung treffen wird. Die Kosten für nicht erfasste Beispiele von z_{C_r} drücken Ungewissheit darüber aus, ob eine weitere Regel diese Beispiele optimal entscheiden wird. Je mehr Beispiele von z_{C_r} eine Regel abdeckt, desto niedriger sind die Zusatzkosten. Damit erhalten größere Prämissen bei identischen Kosten der Prämisse L_{D,P_r} eine bessere Bewertung.

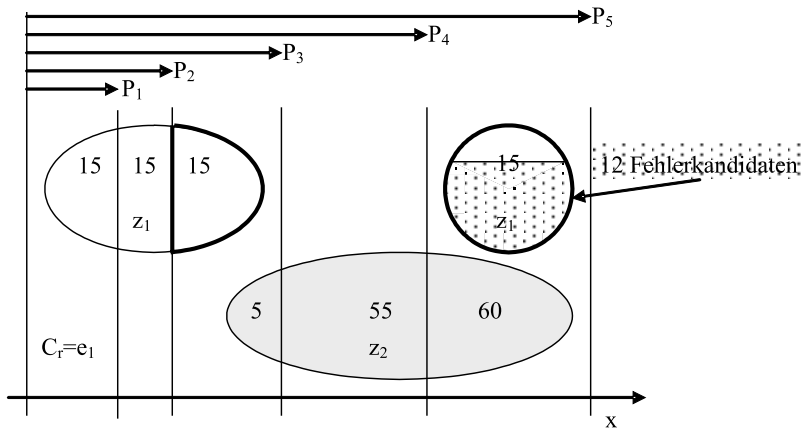


Bild 2.18: Veranschaulichung der Bewertung mit den Beispielen der Prämisse und den Fehlerkandidaten. Die Zahlen in den Bereichen entsprechen der jeweiligen Anzahl der Beispiele. Dicke Linie: Beispiele von z_{C_r} , die nicht von P_2 erfasst werden. Für alle Prämissen ist e_1 die beste Entscheidung.

Die Kosten pro Entscheidung im Gebiet der Prämisse werden nach (2.79) bestimmt. Für die nicht erfassten Beispiele von z_{C_r} müssen Kosten hinzuaddiert werden. Die fünf Prämissen P_1, \dots, P_5 in Bild 2.18 dienen der Visualisierung von Effekten, die bei der Bewertung auftreten. Die Zahlenwerte verdeutlichen die Anzahl der Beispiele der Umweltzustände, die von den einzelnen Prämissen erfasst werden. Vom Umweltzustand z_1 existieren 60 Beispiele und vom Umweltzustand z_2 120 Beispiele. Die dicken Linien in Bild 2.18 markieren exemplarisch die Beispiele von z_{C_r} , die nicht von der Prämisse P_2 erfasst werden. Gesucht ist zunächst der Anteil dieser Beispiele im Lerndatensatz:

$$\hat{P}(z_{C_r} \wedge \bar{P}_r).$$

Mit dem Bayes-Theorem folgt:

$$\hat{P}(z_{C_r} \wedge \bar{P}_r) = \hat{P}(z_{C_r} | \bar{P}_r) \cdot \hat{P}(\bar{P}_r) = \frac{\hat{P}(\bar{P}_r | z_{C_r}) \cdot \hat{P}(z_{C_r})}{\hat{P}(\bar{P}_r)} \cdot \hat{P}(\bar{P}_r). \quad (2.87)$$

Mit (2.75) und erneuter Anwendung des Bayes-Theorems folgt für $\hat{P}(\bar{P}_r|z_{C_r})$:

$$\hat{P}(\bar{P}_r|z_{C_r}) = 1 - \frac{\hat{P}(z_{C_r}|P_r) \cdot \hat{P}(P_r)}{\underbrace{\hat{P}(z_{C_r})}_{\hat{P}(P_r|z_{C_r})}}. \quad (2.88)$$

Einsetzen in (2.87) und Vereinfachen führt zu:

$$\hat{P}(z_{C_r} \wedge \bar{P}_r) = \hat{P}(z_{C_r}) - \hat{P}(z_{C_r}|P_r) \cdot \hat{P}(P_r). \quad (2.89)$$

Diese Terme können alle direkt aus dem Datensatz geschätzt werden. Für die nicht erfassten Beispiele von z_{C_r} werden zunächst die Kosten L_Z angenommen, die bei der Entscheidung der negierten Prämisse $e(\bar{P}_r)$ anfallen: $L_Z = L(e(\bar{P}_r), z_{C_r})$. Die Kosten pro Beispiel der Regel werden dann als gewichtete Summe berechnet:

$$\hat{L}_{D,P_r,Z} = \frac{\hat{L}_{D,P_r} \cdot \hat{P}(P_r) + L_Z \cdot \hat{P}(z_{C_r} \wedge \bar{P}_r)}{\hat{P}(P_r) + \hat{P}(z_{C_r} \wedge \bar{P}_r)}. \quad (2.90)$$

Sind die besten Entscheidungen für die Prämisse und ihre Negation identisch, ist eine andere Entscheidung für die Bestimmung von L_Z auszuwählen. Bei Zulässigkeit von $e(\bar{P}_r) = e(P_r)$ für die Bestimmung von L_Z tritt der gewünschte Effekt nicht ein. Um für die restlichen Beispiele von z_{C_r} die Kosten einer suboptimalen Entscheidung anzusetzen, wird die zweitbeste Entscheidung für das Gebiet der negierten Prämisse gewählt. Damit ergeben sich die Zusatzkosten L_Z pro Beispiel zu:

$$L_Z = \begin{cases} L(e(P_r), z_{C_r}) & \text{für } e(P_r) \neq e(\bar{P}_r) \\ L(e_{opt, \bar{P}_r} \neq e(\bar{P}_r), z_{C_r}) & \text{für } e(P_r) = e(\bar{P}_r) \end{cases}. \quad (2.91)$$

Auch diese Bewertung für Einzelregeln ist noch unzureichend. Die vier Prämissen in Bild 2.18 verdeutlichen die Effekte der Bewertung. Tabelle 2.11 zeigt die Bewertung für die Prämissen aus Bild 2.18 für ein Kostenverhältnis $L_R = 10$. Bei einer Bewertung nach (2.91) und (2.90) führt die vierte Prämisse zu den niedrigsten erwarteten Kosten. Dieses Ergebnis ist nicht plausibel. In (2.90) wird davon ausgegangen, dass für alle Beispiele von z_{C_r} , die nicht im Gebiet der Prämisse P_r liegen, durch eine andere Regel R_s eine Fehlentscheidung verursacht wird. Dies ist jedoch nur dann möglich, wenn es prinzipiell Regeln geben kann, die diese Fehler verursachen.

Zur Untersuchung dieses Falles wird geprüft, welche Bedingungen zur Existenz solcher Regeln aufzustellen sind. Dazu wird eine überlappungsfreie Aufteilung des Gebietes der negierten Prämisse in mindestens eine Prämisse P_s mit $e(P_s) \neq e(P_r)$ angenommen. Da die Konklusion für eine Prämisse nach (2.19) bestimmt wird, muss für alle P_s gelten:

$$\sum_{i=1}^{N_z} L(e(P_s), z_j) \cdot \hat{P}(z_j|P_s) < \sum_{i=1}^{N_z} L(e(P_r), z_j) \cdot \hat{P}(z_j|P_r). \quad (2.92)$$

Existiert nur eine Prämisse P_s , ergibt sich aus (2.92) nach Multiplikation mit $\hat{P}(\bar{P}_r)$ die Bedingung für die maximale Verbundwahrscheinlichkeit für Beispiele von z_{C_r} in der ne-

gierten Prämisse $\bar{P}_r = P_s$, so dass $e(P_s) \neq e(P_r)$ gilt:

$$\hat{P}(z_{C_r} \wedge \bar{P}_r) \leq \frac{\sum_{z_j=\{z_1, \dots, z_{C_r-1}, z_{C_r+1}, \dots, z_{N_z}\}} [L(e(\bar{P}_r), z_j) - L(e(P_r), z_j)] \cdot \hat{P}(z_j \wedge \bar{P}_r)}{\underbrace{L(e(P_r), z_{C_r}) - L(e(\bar{P}_r), z_{C_r})}_{\hat{P}_{\max}(z_{C_r} \wedge \bar{P}_r)}} \quad (2.93)$$

Diese Gleichung bedeutet aber, dass es eine obere Schranke $\hat{P}_{\max}(z_{C_r} \wedge \bar{P}_r)$ für den Anteil der Beispiele von z_{C_r} gibt, für die eine Fehlentscheidung getroffen werden kann. Nur wenn $\hat{P}(z_{C_r} \wedge \bar{P}_r)$ kleiner als die Schranke ist, kann \bar{P}_r aus genau einer Prämisse P_s mit $e(P_s) \neq e(P_r)$ bestehen. Übersteigt $\hat{P}(z_{C_r} \wedge \bar{P}_r)$ die Schranke, existiert mindestens eine Aufteilung des Gebiets der negierten Prämisse mit mindestens einer Regel, für deren Prämisse $e(P_s^*) = e(P_r)$ die beste Entscheidung ist. Für die Prämisse P_2 aus Bild 2.18 ist $p_{\max}(z_1 \wedge \bar{P}_1) = 12/180$. Bei $L_R = 10$ kann mit 120 Beispielen des Umweltzustandes z_2 höchstens für 12 Beispiele von z_1 eine Fehlentscheidung getroffen werden. In Bild 2.18 sind die 12 Fehlerkandidaten von z_1 durch die gepunktete Fläche und die 120 Beispiele von z_2 durch die graue Fläche markiert. Sobald mehr als 12 Beispiele von z_1 in einer Prämisse liegen, ist e_1 die beste Entscheidung unabhängig vom Anteil der Beispiele z_2 , die zusätzlich in der Prämisse sind. Bild 2.19 verdeutlicht nochmals die Möglichkeiten für die Aufteilung der Restprämisse. Dabei ist im oberen Teil der Anteil der Beispiele von z_{C_r} kleiner als $\hat{P}_{\max}(z_{C_r} \wedge \bar{P}_r)$. Damit ist eine andere Entscheidung für die gesamte Restprämisse möglich. Im unteren Teil ist der Anteil der Beispiele von z_{C_r} größer als $\hat{P}_{\max}(z_{C_r} \wedge \bar{P}_r)$ und damit existiert mindestens eine weitere Prämisse P_s mit der selben Konklusion wie P_r .

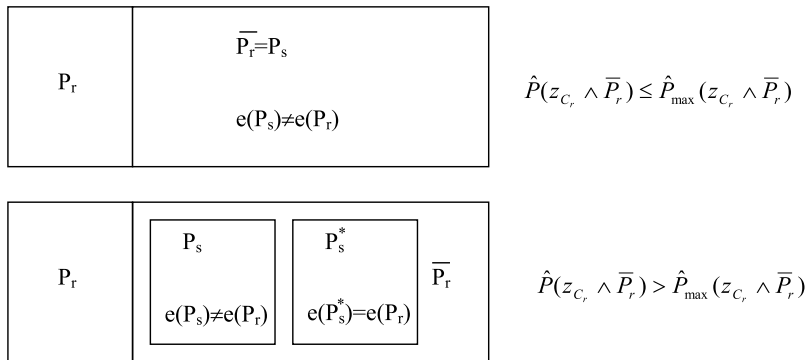


Bild 2.19: Aufteilungen der Restprämisse. Oben: negierte Prämisse kann unterschiedliche Konklusion haben. Unten: mindestens eine Prämisse P_s^* mit $e(P_s^*) = e(P_r)$ existiert.

Ist $\hat{P}(z_{C_r} \wedge \bar{P}_r)$ größer als die obere Schranke, wird die Regel nach (2.90) zu pessimistisch bewertet. Für einen Teil der nicht in der Prämisse erfassten Beispiele ist auch im ungünstigsten Fall wegen (2.93) keine Fehlentscheidung möglich. Diese Beispiele können also bei der Bewertung unberücksichtigt bleiben. Für die Gewichtung der Kosten ergibt

	Prämisse P_r				
	1	2	3	4	5
$N(z_{c_r} \wedge \bar{P}_r) / \hat{P}(z_{c_r} \wedge \bar{P}_r)$	45/0.25	30/0.167	15/0.083	15/0.083	0/0
$N_{max}(z_{c_r} \wedge \bar{P}_r) / \hat{P}_{max}(z_{c_r} \wedge \bar{P}_r)$	12/0.067	12/0.067	11.5/0.064	6/0.033	0/0
$\hat{L}_{D,P_r,Z}$ mit $L_Z = 10$, $\hat{P}(z_{c_r} \wedge \bar{P}_r)$, (2.91), (2.90)	7.5	5	2.38	1.75	0.67
$\hat{P}(z_{C_r} \wedge P_s)$, (2.91), (2.94)	4.44	2.86	1.95	1.08	0.67
$\hat{L}_{D,P_r,Z}$ mit $L_Z = 0.909$, $\hat{P}(z_{C_r} \wedge P_s)$ und (2.94),(2.100)	0.40	0.26	0.25	0.59	0.67
(2.102),(2.100), $\alpha_{ER} = 0.9$	0.38	0.240	0.238	0.59	0.67
(2.102),(2.100), $\alpha_{ER} = 0.5$	0.26	0.15	0.18	0.58	0.67
(2.102),(2.100), $\alpha_{ER} = 0$	0	0	0.1	0.57	0.67
$\hat{L}_{D,RB}$ kompletter Datensatz	0.67	0.67	0.67	0.67	0.67

Tabelle 2.11: Bewertungen $\hat{L}_{D,P_r,Z}$ für die Prämissen aus Bild 2.18. $L_R = 10$.

sich damit:

$$\hat{L}_{D,P_r,Z} = \frac{\hat{L}_{D,P_r} \cdot \hat{P}(P_r) + L_Z \cdot \min(\hat{P}(z_{c_r} \wedge \bar{P}_r), p_{\max}(z_{C_r} \wedge \bar{P}_r))}{\hat{P}(P_r) + \min(\hat{P}(z_{c_r} \wedge \bar{P}_r), p_{\max}(z_{C_r} \wedge \bar{P}_r))}. \quad (2.94)$$

Im oberen Teil von Tabelle 2.11 sind die Anteile der nicht erfassten Beispiele von z_{C_r} für die Prämissen aus Bild 2.18 und die Anzahl Fehlerkandidaten zusammengestellt. Der mittlere Teil von Tabelle 2.11 zeigt neben der Bewertung der Prämissen nach (2.91) und (2.90) auch die Bewertung nach (2.91) und (2.94). Weiterhin erhält P_5 die beste Bewertung. Der Grund dafür ist die Höhe der nach (2.91) bestimmten Kosten für die nicht erfassten Beispiele. Die Bewertungen der Prämissen P_3 und P_4 setzen sich wie folgt zusammen:

$$\hat{L}_{D,P_3,Z} = \underbrace{\frac{1 \cdot \frac{5}{50} \cdot \frac{50}{180}}{50+11.5}}_{\approx 0.08} + \underbrace{\frac{10 \cdot \frac{11.5}{180}}{50+11.5}}_{\approx 1.87} = 1.95, \quad \hat{L}_{D,P_4,Z} = \underbrace{\frac{1 \cdot \frac{60}{105} \cdot \frac{105}{180}}{105+6}}_{\approx 0.54} + \underbrace{\frac{10 \cdot \frac{6}{180}}{105+6}}_{\approx 0.54} = 1.08. \quad (2.95)$$

Die zusätzlichen 'günstigen' Fehler $L(e_1, z_2) = 1$ in der Prämisse P_4 in Kombination mit der Reduktion von $\max(\hat{P}(z_{c_r} \wedge \bar{P}_r))$ bewirken die Reduktion der Kosten $\hat{L}_{D,P_4,Z}$.

Es werden nun die maximalen erwarteten Kosten je Beispiel von z_{C_r} für den Fall bestimmt, dass eine andere Regel R_s für diese Beispiele eine Fehlentscheidung trifft. Die Kosten ergeben sich, wenn für alle Beispiele von z_{C_r} eine falsche Entscheidung getroffen wird, bei denen es überhaupt möglich ist, eine falsche Entscheidung zu treffen. Betrachtet wird die Prämisse P_s , die alle Beispiele des Umweltzustandes $z_{C_s} \neq z_{C_r}$ innerhalb \bar{P}_r erfasst. Der maximale Anteil an Beispielen von z_{C_r} für die die Fehlentscheidung $e(P_s)$

getroffen werden kann, ergibt sich aus (2.93):

$$\hat{P}_{\max}(z_{C_r} \wedge P_s) = \underbrace{\frac{L(e(P_s), z_{C_s}) - L(e(P_r), z_{C_s})}{L(e(P_r), z_{C_r}) - L(e(P_s), z_{C_r})}}_{V_{s,r}} \cdot \hat{P}(z_{C_s} \wedge \bar{P}_r). \quad (2.96)$$

Dabei entspricht $V_{s,r}$ dem Verhältnis der Beispiele von z_{C_s} zu z_{C_r} in P_s . Mit dem tatsächlichen Anteil von z_{C_r} in der negierten Prämisse folgt:

$$\hat{P}(z_{C_r} \wedge P_s) = \min(\hat{P}_{\max}(z_{C_r} \wedge P_s), \hat{P}(z_{C_r} \wedge \bar{P}_r)). \quad (2.97)$$

Damit ergibt sich die Wahrscheinlichkeit von P_s über den Satz der totalen Wahrscheinlichkeit zu:

$$\hat{P}(P_s) = \hat{P}(z_{C_r} \wedge P_s) + \underbrace{\frac{\hat{P}(z_{C_r} \wedge P_s)}{V_{s,r}}}_{\hat{P}(z_{C_s} \wedge P_s)}. \quad (2.98)$$

Die Kosten für die Prämisse P_s bezogen auf den gesamten Datensatz sind dann:

$$\hat{L}_{D,P_s,total} = \hat{P}(z_{C_r} \wedge P_s) \cdot L(e(P_s), z_{C_r}) + \hat{P}(z_{C_s} \wedge P_s) \cdot L(e(P_s), z_{C_s}). \quad (2.99)$$

Die Kosten für die nicht erfassten Beispiele L_Z werden dann als gewichtete Summe der Kosten der Prämissen P_s bestimmt. Die Kosten der Prämissen P_s werden dabei in der Reihenfolge absteigender Kosten der Fehlentscheidungen für z_{C_r} berechnet. Begonnen wird mit der Fehlentscheidung, die die größten Kosten verursacht. Dabei wird $\hat{P}(z_{C_r} \wedge \bar{P}_r)$ jeweils um den Teil verringert, der bereits durch ein berechnetes P_s entschieden wird. Für L_Z folgt:

$$L_Z = \frac{\sum_{P_s} \hat{L}_{D,P_s,total}}{\sum_{P_s} \hat{P}(P_s)}. \quad (2.100)$$

Für das Beispiel in Bild 2.18 ergibt sich mit $L_R = 10$ für P_3 mit $e(P_3) = e_1$:

$$V_{s,2} = \frac{L(e_2, z_2) - L(e_1, z_2)}{L(e_1, z_1) - L(e_2, z_1)} = \frac{0 - 1}{0 - 10} = \frac{1}{10}.$$

Und damit:

$$\hat{P}_{\max}(z_1 \wedge P_s) = \frac{1}{10} \cdot \underbrace{\frac{115}{180}}_{\hat{P}(z_2 \wedge \bar{P}_r)} = \frac{115}{1800} \approx 0.064.$$

Dieser Wert ist kleiner als der Anteil von z_1 in der negierten Prämisse. Für die Wahrscheinlichkeit von P_s folgt:

$$\hat{P}(P_s) = \frac{115}{1800} + 10 \cdot \frac{115}{1800} \approx 0.70,$$

und für L_Z

$$L_Z = \frac{\frac{115}{1800} \cdot 10 + \frac{115}{180} \cdot 0}{\frac{115}{1800} + 10 \cdot \frac{115}{1800}} = \frac{10}{11}. \quad (2.101)$$

Die erste Zeile des unteren Teils von Tabelle 2.11 zeigt die Bewertung der Prämissen aus Bild 2.18 mit der modifizierten Berechnung von L_Z . Nun stellt P_3 die beste Prämisse dar. Dies ist plausibel, da die restlichen Beispiele von z_1 über eine weitere Regel entschieden werden können. Eine mögliche Regel hat die Prämisse \bar{P}_4 .

Bild 2.20 zeigt die Bewertung für die Regeln des Beispieldatensatzes aus Tabelle 2.10 grafisch mit Höhenlinien.

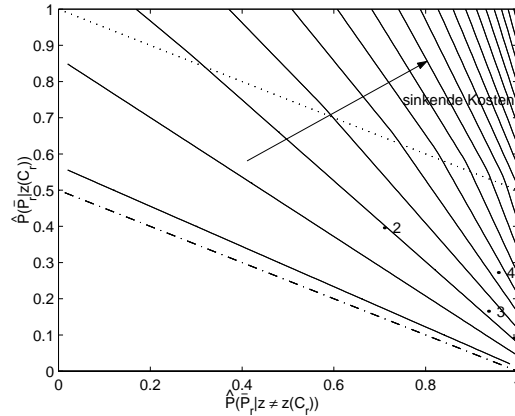


Bild 2.20: Darstellung der Bewertung der Regeln aus Tabelle 2.10 über die Kosten der Prämisse und den Beispielen z_{C_r} , für die eine Fehlentscheidung getroffen werden kann. L_Z wird über (2.100) berechnet.

Bei der Bewertung nach (2.94) mit (2.100) ist die vierte Regel die beste. Dieses Bewertungsverfahren ist noch immer pessimistisch, da für jede Regel Zusatzkosten in maximaler Höhe berücksichtigt werden. Ist eine optimistischere Bewertung gewünscht, können die Zusatzkosten durch einen Faktor α_{ER} reduziert werden:

$$\hat{L}_{D,P_r,Z} = \frac{\hat{L}_{D,P_r} \cdot \hat{P}(P_r) + \alpha_{ER} \cdot L_Z \cdot \min(\hat{P}(z_{C_r} \wedge \bar{P}_r), p_{\max}(z_{C_r} \wedge \bar{P}_r))}{\hat{P}(P_r) + \alpha_{ER} \cdot \min(\hat{P}(z_{C_r} \wedge \bar{P}_r), p_{\max}(z_{C_r} \wedge \bar{P}_r))}, \quad 0 \leq \alpha_{ER} \leq 1. \quad (2.102)$$

Für $\alpha_{ER} = 1$ geht (2.102) in die pessimistische Bewertung über. Für $\alpha_{ER} = 0$ geht (2.102) in die optimistische Bewertung (2.79) über, die nur die Beispiele in der Prämisse berücksichtigt. Je kleiner α_{ER} gewählt wird, desto höher ist das Gewicht der Kosten in der Prämisse. Damit werden fehlerfreie Prämissen gut bewertet. Solange $\alpha_{ER} > 0$ ist, werden Prämissen mit mehr Beispielen ($\hat{P}(P_r|z_{C_r})$ steigt bei identischer Fehlerzahl ($\hat{P}(\bar{P}_r|z_j \neq z_{C_r})$ ist konstant) besser bewertet (Bild 2.20). Im unteren Teil von Tabelle 2.11 sind für $\alpha_{ER} \in \{0.9; 0.5; 0\}$ die Bewertungen der Prämissen aus Bild 2.18 dargestellt. Je kleiner α_{ER} gewählt wird, desto stärker werden kleine fehlerfreie Prämissen bevorzugt. Es bestehen somit Parallelen zum $\mu - \sigma$ -Prinzip. Zum Erwartungswert der Kosten der Prämisse – erster Summand des Zählers von (2.102) – wird eine Schwankungsgröße hinzuaddiert. Die Schwankungsgröße ist hier nicht die Standardabweichung, sondern drückt die Ungewissheit über die Kosten der Entscheidungen für die von der Prämisse nicht erfassten Beispiele von z_{C_r} aus.

R_r		1	2	3	4	5	6	7
C_r		e_1	e_1	e_1	e_1	e_1	e_1	e_1
$N_{z_j P_r}$	z_1	5	24	10	16	0	60	50
	z_2	2	87	19	13	0	0	20
$N_{z_j \bar{P}_r}$	z_1	55	36	50	44	60	0	10
	z_2	298	213	281	287	300	300	280
$\hat{P}(z_j P_r)$	z_1	0.700	0.215	0.345	0.566	-	1	0.714
	z_2	0.300	0.875	0.655	0.434	-	0	0.286
$\hat{P}(z_j \bar{P}_r)$	z_1	0.155	0.145	0.151	0.132	0.167	0	0.035
	z_2	0.845	0.855	0.849	0.868	0.833	1	0.965
$\hat{P}(P_r)$		0.021	0.308	0.08	0.08	0	0.167	0.194
$\hat{P}(z_{C_r} \wedge P_s)$		0.083	0.059	0.078	0.079	0.083	0	0.028
$\hat{P}(P_s)$		0.909	0.650	0.859	0.878	0.917	0	0.306
L_Z		0.909	0.909	0.909	0.909	0.909	0	0.909
$\hat{L}_{D,RB}$		0.833	0.833	0.833	0.833	0.833	0	0.333
\hat{L}_{D,P_r}		0.297	0.785	0.655	0.434	0	0	0.286
$\hat{L}_{D,P_r,Z}$		0.774	0.804	0.776	0.665	0.909	0	0.363

Tabelle 2.12: Bewertung (R_1, \dots, R_4) der Einzelregeln aus Bild 2.16 links gemäß der drei Kriterien. Zusätzlich (theoretische Regeln): R_5 : leere Prämisse ($C_r = z_1$), R_6 : ideale Regel, R_7 : gute Regel.

Tabelle 2.12 fasst die verschiedenen Bewertungen und Zwischenergebnisse für die Regeln aus Tabelle 2.10 für den Beispieldatensatz zusammen. Zusätzlich sind die Bewertungen für drei theoretische Regeln angegeben. Eine Regel mit leerer Prämisse stellt R_5 dar. Eine ideale Regel, die fehlerfrei alle Beispiele eines Zustandes erfasst, wird durch R_6 repräsentiert. Eine weitere gute Regel ist durch R_7 gegeben. Nur die Bewertung der Regeln über $\hat{L}_{D,P_r,Z}$ erlaubt eine sinnvolle Ordnung der Güte der Regeln $R_1 \dots R_7$. Bei der Verwendung von $\hat{L}_{D,RB}$ werden die Regeln $R_1 \dots R_4$ identisch bewertet. Wird \hat{L}_{D,P_r} verwendet, sind beispielsweise die Regeln R_5 und R_6 nicht zu unterscheiden. Die Bewertung über $\hat{L}_{D,P_r,Z}$ ist auch in der Lage, die Probleme bei drei und mehr Umweltzuständen zu lösen. Bei dieser Bewertung werden Beispiele der Umweltzustände $z_j \neq z_{C_r}$ nur dann in der Bewertung berücksichtigt, wenn sie im Gebiet Prämisse liegen. Dies bedeutet aber, dass sie dann bei sonst identischen Beispielzahlen als zusätzliche Fehler die Bewertung der Prämisse verschlechtern.

Die Probleme der entscheidungstheoretischen Bewertung von Fuzzy-Einzelregeln sind somit gelöst. Für die Bewertung von Fuzzy-Einzelregeln ist (2.102) zu verwenden. Im folgenden Abschnitt werden die neuen Erkenntnisse auf die Bewertung von vollständigen und unvollständigen Regelbasen übertragen.

2.5 Neue Bewertungsmaße für Fuzzy-Regelbasen

2.5.1 Vollständige Regelbasis

Eine vollständige Regelbasis deckt den gesamten Wertebereich der Eingangsgrößen ab. Der Teil des Wertebereichs, der nicht von einer Prämisse abgedeckt wird, fällt in den Bereich der SONST-Regel. Eine vollständige Regelbasis ordnet jedem Eingangsgrößenvektor \mathbf{x} eine Entscheidung zu: $\mathbf{x} \mapsto e_k$. Für den gesamten Datensatz liegt somit eine Risikosituation vor.

Eine optimale Fuzzy-Regelbasis, häufig existiert diese wegen der Struktur des Problems nicht, weist jedem Beispiel die richtige Entscheidung zu und führt somit zu den niedrigsten möglichen Kosten pro Entscheidung:

$$\hat{L}_{D,min} = \sum_{j=1}^{N_z} L(e_{opt}(z_j), z_j) \cdot \hat{P}(z_j). \quad (2.103)$$

Die Kosten der Merkmale L_C werden zu den Entscheidungskosten hinzuaddiert:

$$\hat{L}_T = \hat{L}_D + L_C. \quad (2.104)$$

Liegen mehrere optimale Regelbasen im Sinne von (2.103) vor, verursacht die Regelbasis mit den niedrigsten Merkmalskosten $L_{C,min}$ auch die niedrigsten Gesamtkosten $\hat{L}_{T,min}$.

Zur Bewertung einer Regelbasis werden die Entscheidungen für die Beispiele des Lern Datensatzes betrachtet. Als Bewertungsmaß wird die Schätzung der erwarteten Kosten je Entscheidung \hat{L}_D nach (2.12) verwendet:

$$\hat{L}_{D,RB} = \sum_{k=1}^{N_e} \sum_{j=1}^{N_z} L(e_k, z_j) \cdot \hat{P}(e_k \wedge z_j). \quad (2.105)$$

Die Kosten der Merkmale werden zu den Entscheidungskosten hinzuaddiert:

$$\hat{L}_T = \hat{L}_{D,RB} + L_C. \quad (2.106)$$

Die Bestimmung von $\hat{P}(e_k \wedge z_j)$ erfolgt durch Auswerten eines der vorgestellten Inferenzverfahren.

Bei der Anwendung der Evidenz-Inferenz ist die Kostenmatrix um die Kosten des Experten und die zusätzliche Entscheidung zu erweitern.

2.5.2 Unvollständige Regelbasis

Bei der Generierung von Regelbasen sind bei einigen Verfahren unvollständige Regelbasen zu bewerten (siehe Abschnitt 3.3.1). Die unvollständige Regelbasis teilt den Lern Datensatz wie eine Einzelregel in zwei Teile: die Beispiele in den Prämissen und die Beispiele der SONST-Regel. Für die Beispiele in den Prämissen liegt wiederum eine Risikosituation vor. Die vorgestellten Inferenzverfahren bestimmen die Entscheidungen für diese Beispiele. Für die Beispiele der SONST-Regel wird eine Rückweisungsentscheidung mit Kosten \hat{L}_{RW} eingeführt. Die Wahrscheinlichkeit, dass ein Beispiel zu dem Teil der Beispiele gehört, für die die Rückweisungsentscheidung getroffen wird, sei \hat{P}_{RW} . Ungewissheit besteht darüber,

	Regelbasis Bild 2.21		Regelbasis Bild 2.7	
	z_1	z_2	z_1	z_2
e_1	16	0	55	38
e_2	0	0	4	262
RW	44	300	1	0
\hat{P}_{RW}	0.956		0.002	
$\hat{L}_{D,PRB}$ (2.105),(2.107)	0		0.217	
$\hat{L}_{D,RB}$ (2.105)	0.833		0.217	
$\hat{L}_{D,PRB,RW}$ (2.108), $\alpha_{RB} = 1$	0.796		0.219	
$\hat{L}_{D,PRB,RW}$ (2.108), $\alpha_{RB} = 0.5$	0.398		0.218	
$\hat{L}_{D,PRB,RW}$ (2.108), $\alpha_{RB} = 0.273$	0.217		0.217	
\hat{L}_{RW}	0.833		0.833	

Tabelle 2.13: Entscheidungen und Bewertungen für die Regelbasen aus Bild 2.21 und Bild 2.7.

ob eine weitere Regel der Regelbasis hinzugefügt und die Anzahl der Beispiele der SONST-Regel reduziert werden kann. Diese Ungewissheit ist in der Bewertung der Regelbasis zu berücksichtigen.

Eine Variante der Bewertung unvollständiger Regelbasen stellen analog zu den Einzelregeln die Kosten der Prämissen $\hat{L}_{D,PRB}$ dar. Es ergibt sich eine optimistische Bewertung, bei der davon ausgegangen wird, dass weitere Regeln existieren, die für die restlichen Beispiele die beste Entscheidung treffen. Zu dem Gebiet der Prämissen zählen alle Beispiele für die gilt:

$$\mu(P_{RB}) = \mu(\cup_{r=1}^{N_R} P_r) \geq 0.5. \quad (2.107)$$

Für Beispiele mit einer geringeren Aktivierung der ODER-Verknüpfung der Prämissen hat die SONST-Regel die maximale Aktivierung. Die Kosten $\hat{L}_{D,PRB}$ ergeben sich über die Inferenz und die Auswertung von (2.105) für die Beispiele der Prämissen. Wie bei den Einzelregeln ist diese Bewertung auch für Regelbasen nicht geeignet, da eine Regelbasis ohne Fehlentscheidung im Gebiet der ausgewählten Prämissen mit $\hat{P}_{RW} \rightarrow 1$ besser als eine Regelbasis mit wenigen Fehlentscheidungen aber $\hat{P}_{RW} \rightarrow 0$ bewertet wird.

Die unvollständige Regelbasis in Bild 2.21 besteht aus Regel 5 aus Tabelle 2.4 und der SONST-Regel. Das Gebiet der Prämissen umfasst lediglich 16 Beispiele von z_1 . Für den Rest des Datensatzes ist die SONST-Regel aktiviert. Bei Bewertung über die Prämissen erhält sie eine bessere Bewertung als die Regelbasis aus Tabelle 2.4 (ebenfalls als unvollständige Regelbasis bewertet). Bei der Betrachtung als vollständige Regelbasis schneidet letztere jedoch bei einem Kostenverhältnis $L_R = 10$ deutlich besser ab. Die Zahlenwerte zeigt der untere Teil von Tabelle 2.13. Zusätzlich sind die richtigen Entscheidungen und die Fehlentscheidungen sowie der Anteil nicht entschiedener Beispiele der beiden Regelbasen eingetragen.

Bei unvollständigen Regelbasen dürfen somit die Beispiele, die nicht von einer Prämisse erfasst sind, bei der Bewertung ebenfalls nicht vernachlässigt werden. Für die Bewertung von unvollständigen Regelbasen ergibt sich analog zur Bewertung von Einzelregeln

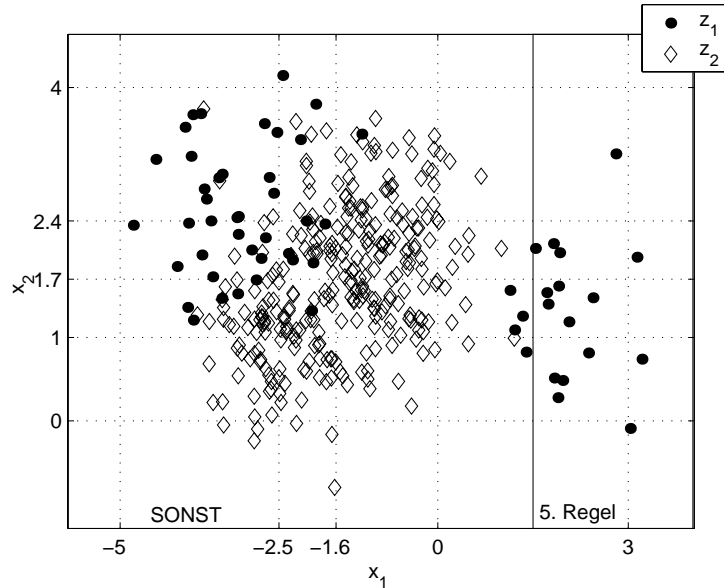


Bild 2.21: Unvollständige Regelbasis für den Beispieldatensatz. (Regel 5 aus Tabelle 2.4 und SONST-Regel.

in (2.102) folgende Struktur:

$$\hat{L}_{D,PRB,RW} = (1 - \hat{P}_{RW}) \cdot \hat{L}_{D,PRB} + \alpha_{RB} \cdot \hat{P}_{RW} \cdot \hat{L}_{RW}. \quad (2.108)$$

Dabei sind \hat{L}_{RW} die Kosten für die nicht von den Prämissen erfassten Beispiele. Der Faktor α_{RB} drückt wiederum die Erwartung über die Qualität der weiteren Regeln aus, die als Kandidaten für die Aufnahme in die Regelbasis bereit stehen. Optimismus spiegelt sich in $\alpha_{RB} \rightarrow 0$ und Pessimismus in $\alpha_{RB} \rightarrow 1$ wider.

Für \hat{L}_{RW} bietet sich zum Beispiel der Wert der Default-Kosten ohne Regeln an (2.16). Im unteren Teil von Tabelle 2.13 sind weiterhin die Werte für die Bewertung der beiden Regelbasen nach (2.108) zusammengestellt. Für $\alpha_{RB} > 0.273$ wird die Regelbasis aus Bild 2.7 bevorzugt. Eine Empfehlung für die Wahl von α_{RB} kann erst im Rahmen des Auswahlverfahrens für Regelbasen in Abschnitt 3.3.1 gegeben werden.

Die wesentlichen neuen Erkenntnisse aus Kapitel 2 werden im Folgenden nochmals kurz zusammengefasst. Für Entscheidungsprobleme mit zwei Zuständen wird in Abschnitt 2.2.3 eine neuartige Visualisierung vorgestellt. In Abschnitt 2.3 wird herausgearbeitet, dass Fuzzy-Regeln bei entscheidungstheoretischer Bewertung für die in der vorliegenden Arbeit verwendeten Kostenmatrizen als Schätzung für die beste Entscheidung und nicht als Schätzung für die Wahrscheinlichkeiten der Zustände zu verwenden sind. Im selben Abschnitt wird zusätzlich ein neues auf der Evidenztheorie basierendes Inferenzverfahren für Fuzzy-Regelbasen vorgeschlagen. In den Abschnitten 2.4 und 2.5 werden neue entscheidungstheoretische Bewertungsmaße für Fuzzy-Einzelregeln und Regelbasen entwickelt.

3 Modifizierte entscheidungstheoretische Entwurfsverfahren für Fuzzy-Systeme

3.1 Überblick

Bei der Generierung von Regelbasen können zwei Ansätze unterschieden werden [106]. Sind die erzeugten Hypothesen vollständige Regelbasen, wird vom Pittsburgh-Ansatz [142] gesprochen. Beim Michigan-Ansatz [68] werden zunächst einzelne Regeln als Hypothesen generiert und anschließend daraus eine Regelbasis zusammengestellt.

Der in [105] vorgeschlagene Algorithmus eignet sich sowohl zum Entwurf von Fuzzy-Einzelregeln als auch zum anschließenden Entwurf von Fuzzy-Regelbasen. Die Erweiterung dieses Algorithmus um die entscheidungstheoretische Bewertung erfolgt in diesem Abschnitt. Der Algorithmus besteht aus drei Teilen. Im ersten Teil werden aus Lerndaten Entscheidungsbäume aufgestellt und aus diesen Regelhypothesen extrahiert. Diese Regelhypothesen werden im zweiten Teil, dem so genannten Pruning (*to prune* - beschneiden), über eine Nachbarschaftssuche durch Weglassen von Teilprämissen oder Hinzufügen von benachbarten Termen generalisiert. Anschließend wird aus den generalisierten Regeln eine Regelbasis ausgewählt. In jeder der drei Stufen des Algorithmus wird zur Bewertung der Hypothesen für Einzelregeln und Regelbasen ein entscheidungstheoretisches Maß verwendet [16, 17].

Es existieren zahlreiche weitere Algorithmen zur datenbasierten Generierung von Fuzzy-Systemen, die auf verschiedenen Entwurfsverfahren basieren. Eine Einteilung kann in Cluster-Verfahren [46, 158], Neuro-Fuzzy-Verfahren [36, 80], Verfahren des Maschinellen Lernens (baumorientierte Verfahren) [125] und evolutionäre Ansätze [31, 45, 85] vorgenommen werden. Einen Überblick über Fuzzy-Entscheidungsbäume gibt [74].

Auf die Besonderheiten einiger Algorithmen wird im Folgenden kurz eingegangen. Der in [127] vorgeschlagene Algorithmus erstellt für jeden Umweltzustand über ein Cluster-Verfahren genau eine Regel über alle Merkmale. Diese Regeln werden anschließend generalisiert. In einem weiteren Schritt werden die Parameter der Regelbasis über genetische Algorithmen optimiert. Im Gegensatz dazu erstellt der in [81] vorgeschlagene Algorithmus zunächst für jedes Beispiel des Lerndatensatzes eine Regel. Die Anzahl der so entstandenen Regeln wird über logische Transformationen reduziert. Die Fuzzy-Regeln in [41] resultieren wiederum aus der Anwendung eines Cluster-Verfahrens. Anschließend werden hier über Ähnlichkeits-Maße redundante Bereiche der Regelbasis eliminiert. Dadurch soll die Interpretierbarkeit der Regelbasis gesteigert werden.

Ein Algorithmus zum Entwurf von Fuzzy-Entscheidungssystemen, die unterschiedliche Konsequenzen für Fehlentscheidungen berücksichtigen, wird erstmals in dieser Arbeit

diskutiert.

3.2 Hypothesengenerierung - Einzelregeln

Der Prozess zur Generierung von Fuzzy-Einzelregeln ist in zwei Stufen aufgeteilt. Im ersten Schritt werden in einem Eröffnungsverfahren Hypothesen aus einem Entscheidungsbaum generiert. Bei einem Eröffnungsverfahren steht das Finden von zulässigen Lösungen im Vordergrund [163]. Anschließend werden diese Einzelregeln in einem Verbesserungsverfahren über eine Nachbarschaftssuche generalisiert. Ein Verbesserungsverfahren sucht den Lösungsraum nach besseren zulässigen Lösungen ab [163]. Beim Pruning werden die Prämissen der Einzelregeln Schritt für Schritt generalisiert.

3.2.1 Eröffnungsverfahren (Entscheidungsbaum)

Bei der Entwicklung der Entscheidungsbäume wird der Lerndatensatz Schritt für Schritt mittels verschiedener Merkmale aufgetrennt. Ausgehend von einem Ursprungsknoten für alle Beispiele befinden sich in den nachfolgenden Knoten Teile des Datensatzes. Jedem Knoten wird auf der Basis der enthaltenen Beispiele die optimale Entscheidung zugewiesen. Die Auftrennung erfolgt für die m_l linguistischen Terme eines Merkmals. Die Zugehörigkeiten zu den einzelnen Termen werden dabei durch α -Schnitte auf $\{0, 1\}$ abgebildet:

$$\begin{aligned} \mu_{A_{l,i}}[n] &= 1 && \text{für} && \mu_{A_{l,i}}[n] \geq 0.5 \wedge \mu_{A_{l,i}}[n] \geq \mu_{A_{l,i+1}}[n] \\ \mu_{A_{l,i}}[n] &= 0 && \text{sonst.} \end{aligned} \quad (3.1)$$

In dieser Arbeit soll ein Knoten aufgetrennt werden, wenn dadurch die Entscheidungskosten für die Beispiele im Knoten gegenüber den Kosten ohne Auftrennen sinken. Das Verbesserungspotenzial dieser Spezialisierung besteht darin, dass durch die Auftrennung für die Beispiele im aktuellen Knoten unterschiedliche Entscheidungen getroffen werden können. Die Kosten pro Entscheidung ohne Auftrennen \hat{L}_D werden über (2.16) bestimmt. Aus den Termen eines Merkmals, das noch nicht zur Auftrennung verwendet wurde, lässt sich für die Beispiele im Knoten eine Regelbasis ohne Überlappung mit m_l Regeln generieren:

$$R_i: \text{WENN } A_{l,i} \text{ DANN } C_i, \quad i = 1, \dots, m_l. \quad (3.2)$$

Die Konklusionen $C_i = e(A_{l,i})$ werden über (2.19) bestimmt. Die Kosten pro Entscheidung \hat{L}_{D,x_l} nach einer Auftrennung mit dem Merkmal x_l ergeben sich durch Auswertung dieser Regelbasis für die Beispiele im Knoten:

$$\hat{L}_{D,x_l} = \sum_{i=1}^{m_l} \sum_{j=1}^{N_z} L(e(A_{l,i}), z_j) \cdot \hat{P}(z_j \wedge A_{l,i}) \quad (3.3)$$

Der aktuelle Knoten wird mit dem Merkmal x_a weiter aufgespalten, für das die Kostenreduktion maximal wird:

$$x_a = \arg \max_{x_l} [\hat{L}_D - \hat{L}_{D,x_l}] \quad (3.4)$$

Führt keine Aufspaltung zu einer Kostenreduktion, wird der aktuelle Knoten zu einem Endknoten des Baumes. Der Vorgang wird so lange fortgesetzt, bis alle Knoten Endknoten sind.

Die Entwicklung eines Entscheidungsbaumes für die Merkmale x_1 und x_2 des Beispieldatensatzes ist in Bild 3.1 dargestellt. Im Ursprungsknoten wird der Baum und damit der Datensatz auf der ersten Ebene über das Merkmal x_1 in fünf Teile aufgetrennt. Für jeden Teil wird die optimale Entscheidung bestimmt und untersucht, ob ein weiteres Auftrennen über das Merkmal x_2 zu einer weiteren Kostenreduktion führt. Exemplarisch für die zweite Ebene ist der Knoten $x_1 = A_{1,2}$ weiter aufgetrennt. Die Kosten einiger Knoten sind in Tabelle 3.1 zusammengestellt. Im oberen Teil der Tabelle sind die Kosten einer Auftrennung über die beiden Merkmale verglichen. Zusätzlich sind die besten Entscheidungen, die Anzahl Fehler und die Gesamtzahl an Beispielen der einzelnen Teile des Datensatzes eingetragen. Die Auftrennung des Ursprungsknotens nach x_1 führt zu Kosten von $\hat{L}_{D,x_1} = 0.51$. Die Auftrennung nach x_2 verursacht Kosten in Höhe von $\hat{L}_{D,x_2} = 0.82$. Die Auftrennung erfolgt deshalb über das Merkmal x_1 . Für die zweite Ebene ist die Auftrennung der Knoten $x_1 = A_{1,1}$ und $x_1 = A_{1,2}$ dargestellt.

Eine vergleichbare Strategie stellt die in [60] vorgeschlagene baumorientierte Regel-Induktionsstrategie dar. Allerdings unterscheiden sich die Bewertungskriterien für die Auftrennung der Äste des Baumes. In [60] findet keine entscheidungstheoretische Bewertung statt.

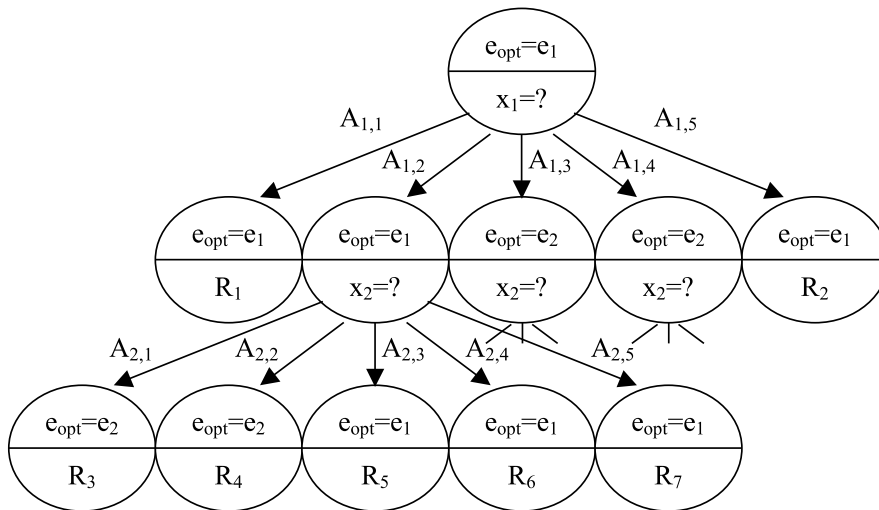


Bild 3.1: Ausschnitt aus einem Entscheidungsbaum.

Die Regelhypothesen für das Regelpruning werden aus den Endknoten der Entscheidungsbaume extrahiert. Dazu werden die Prämissen der Regelhypothesen durch 'UND' Verknüpfungen der Bedingungen auf den einzelnen Ebenen der Bäume gebildet. Die Regeln des Baumes aus Bild 3.1 und Tabelle 3.1 sind in Bild 3.2 grafisch durch Rahmen und die Markierung r . Regel dargestellt. Regeln mit der Konklusion $C_r = e_1$ sind grau hinterlegt. Die anderen Regeln haben die Konklusion $C_r = e_2$.

Es besteht die Option, mehrere Entscheidungsbaume mit unterschiedlichen Merkmalen im Ursprung zu generieren. Dazu werden die Merkmale gesperrt, die für die Auftrennung

	Ursprung	$\hat{L}_D = 0.83$			$e_{opt} = e_1$	
1. Ebene	Auftrennung x_1	$A_{1,1}$	$A_{1,2}$	$A_{1,3}$	$A_{1,4}$	$A_{1,5}$
	C_i	e_1	e_1	e_2	e_2	e_1
	Fehl./Anz.	1/10	94/120	5/131	4/83	0/16
	$\hat{L}_{D,A_{1,i}}$	0.1	0.78	0.38	0.48	0
	\hat{L}_{D,x_1}	0.51 → Auftrennung				
	Auftrennung x_2	$A_{2,1}$	$A_{2,2}$	$A_{2,3}$	$A_{2,4}$	$A_{2,5}$
	C_i	e_1	e_2	e_1	e_1	e_1
	Fehl./Anz.	24/27	9/103	79/97	87/107	16/26
	$\hat{L}_{D,A_{2,i}}$	0.89	0.87	0.81	0.81	0.61
	\hat{L}_{D,x_2}	0.82 → keine Auftrennung				
2. Ebene	$A_{1,1}$	$\hat{L}_D = 0.1$			$e_{opt} = e_1$	
	Auftrennung x_2	$A_{2,1}$	$A_{2,2}$	$A_{2,3}$	$A_{2,4}$	$A_{2,5}$
	C_i	e_1	e_1	e_1	e_1	e_1
	Fehl./Anz.	0/0	1/2	0/2	0/4	0/2
	$\hat{L}_{D,A_{2,i}}$	0.0	0.5	0.0	0.0	0.0
	\hat{L}_{D,x_2}	0.1 → keine Auftrennung (Regel 1)				
	$A_{1,2}$	$\hat{L}_D = 0.78$			$e_{opt} = e_1$	
	Auftrennung x_2	$A_{2,1}$	$A_{2,2}$	$A_{2,3}$	$A_{2,4}$	$A_{2,5}$
	C_i	e_2	e_2	e_1	e_1	e_1
	Fehl./Anz.	0/16	0/48	19/27	9/22	2/7
	$\hat{L}_{D,A_{2,i}}$	0.0	0.0	0.70	0.41	0.29
\hat{L}_{D,x_2}	0.25 → Auftrennung (Regel 3-7)					
	⋮					

Tabelle 3.1: Kosten und Auftrennung für die erste Ebene und für Teile der zweiten Ebene des Entscheidungsbaumes aus Bild 3.1.

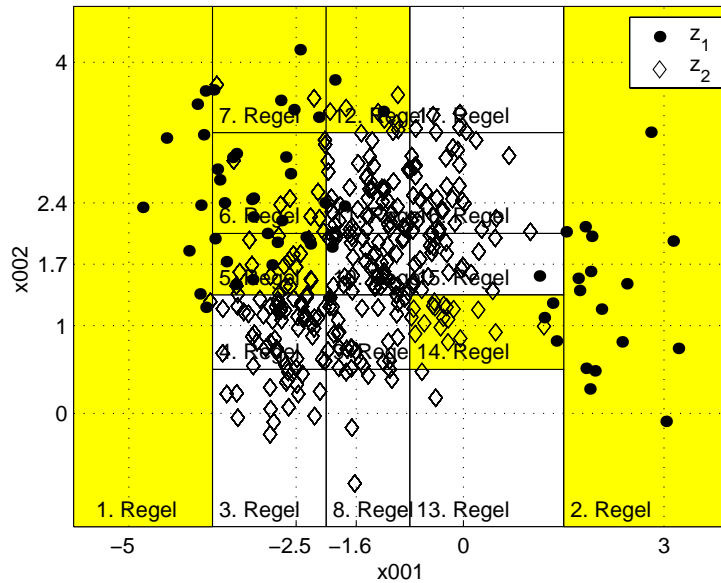


Bild 3.2: Darstellung der Regeln aus den Endknoten des Baumes (Ausschnitt in Bild 3.1). Grau: Entscheidung e_1 , Rest e_2 .

der Ursprungsknoten bereits erzeugter Bäume verwendet wurden. Dadurch wird die Zahl unterschiedlicher Regeln als Basis für das nachfolgende Regelpruning vergrößert.

Das Ergebnis ist meist eine recht hohe Anzahl von spezialisierten Einzelregeln, die nur wenige Beispiele abdecken. Diese werden im folgenden Schritt - dem Pruning - generalisiert.

Merkmalskosten können bereits bei der Entwicklung von Entscheidungsbäumen berücksichtigt werden. Eine Reduktion der Entscheidungskosten durch ein zusätzliches Merkmal kann jedoch durch die zusätzlichen Merkmalskosten aufgehoben werden. Eine Auftrennung des Baumes findet dann nicht statt. Dies kann dazu führen, dass aussichtsreiche Kandidaten für die Regelgeneralisierung nicht zur Verfügung stehen. Im Baum des Beispieldatensatzes wird für Merkmalskosten von $L_C = 0.2$ für x_1 und x_2 zum Beispiel die Auftrennung des Knotens $x_1 = A_{1,4}$ nicht vorgenommen. Dadurch steht die 14. Regel des Baumes für das Pruning nicht zu Verfügung. Aus dieser Regel geht jedoch die Regel 3 aus Tabelle 2.4 hervor. Durch die Berücksichtigung von Merkmalskosten beim Generieren der Entscheidungsbäume können für die spätere Regelbasis wichtige Einzelregeln verhindert werden. Daher werden die Merkmalskosten beim Generieren von Entscheidungsbäumen nicht verwendet.

3.2.2 Verbesserungsverfahren (Pruning)

Ziel des Prunings ist es, aus der Menge der spezialisierten Regelhypothesen durch Generalisierung (Vergrößerung der Gebiete der Prämissen) die Kandidaten für die folgende Suche einer Regelbasis zu bilden. Die Einzelregeln werden während des Prunings nach (2.102)

bewertet.

In [32] werden Pruningverfahren für Entscheidungsbäume unter Berücksichtigung der Entscheidungskosten diskutiert. In der vorliegenden Arbeit werden die Einzelregeln und nicht die gesamten Bäume für das Pruning verwendet. Dies führt in der Regel dazu, dass sich die Einzelregeln eines Baumes nach dem Pruning überlappen.

Die Untersuchung jeder möglichen Einzelregel ist wegen deren exponentiell wachsender Zahl nicht sinnvoll. Beispielsweise ergeben sich für ein Problem mit fünf Merkmalen und zwei möglichen Entscheidungen mit $m_l = 5$ bereits ca. 1.5 Millionen verschiedene Regeln. Jedes weitere Merkmal und jede zusätzliche Entscheidung erhöhen diese Zahl erheblich. Von jeder Teilprämisse können ein bis fünf Terme aktiviert sein. Dies führt zu insgesamt 15 möglichen Varianten jeder Teilprämisse. Die aktivierten Terme einer Teilprämisse müssen entsprechend der Definition des Fuzzy-Systems benachbart sein (siehe Abschnitt 2.3.1). Daraus resultiert für m_l Terme und $N_a \leq m_l$ aktivierten Teilprämissen die Anzahl möglicher Teilprämissen $N_{TP}(m_l, N_a)$:

$$N_{TP}(m_l, N_a) = m_l - (N_a - 1). \quad (3.5)$$

Über alle möglichen $N_a : \{1 \dots m_l\}$ ergibt sich die Anzahl der möglichen Teilprämissen zu:

$$N_{TP}(m_l) = \sum_{j=1}^{m_l} m_l - (j - 1). \quad (3.6)$$

Da alle Teilprämissen unabhängig voneinander gestaltet werden können, ergibt sich die Gesamtzahl möglicher Prämissen N_{Pr} als Produkt der möglichen Teilprämissen:

$$N_{Pr} = \prod_{l=1}^{N_x} \sum_{j=1}^{m_l} m_l - (j - 1). \quad (3.7)$$

Die Anzahl möglicher Regeln ergibt sich dann durch Multiplikation mit der Anzahl möglicher Entscheidungen. Für den Beispieldatensatz mit zwei Merkmalen sind theoretisch 450 Regeln möglich:

$$N_{R,th} = \underbrace{2}_{N_e} \cdot \underbrace{(5 + 4 + 3 + 2 + 1)}_{x_1} \cdot \underbrace{(5 + 4 + 3 + 2 + 1)}_{x_2} = 450 \quad (3.8)$$

Beim Pruning der Regelhypothesen aus dem Baum wird meist nur ein Teil der theoretisch möglichen Regeln untersucht. Beim Beispieldatensatz sind das ca. 210 Regeln. Der prozentuale Anteil der untersuchten Regeln an den theoretisch möglichen nimmt mit zunehmender Anzahl an Merkmalen ab.

Das Pruning wird sequentiell für die Hypothesen durchgeführt. Zunächst werden aus der aktuellen Regel generalisierte Hypothesen abgeleitet. Dabei stehen zwei Optionen zur Generalisierung zur Verfügung. Entweder wird eine Teilprämisse aus der Regel gestrichen, oder zu einer Teilprämisse werden benachbarte Terme hinzugefügt [105]. Aus den so erzeugten Hypothesen wird die beste für den nächsten Pruningschritt ausgewählt. Das Pruning einer Einzelregel ist beendet, wenn keine Verbesserung mehr eintritt.

Für die sechste Regel aus Bild 3.2 ergeben sich die in Tabelle 3.2 aufgeführten Hypothesen für den ersten Pruningschritt. Bild 3.3 zeigt die Hypothesen aus Tabelle 3.2 über den

r	WENN	DANN	$\hat{L}_{D,P_r,Z}$	Anz./Fehl.
1	$(x_1 = A_{1,2}) \text{ UND } (x_2 = A_{2,3})$	$C_1 = e_1$	0.80	12/17
2	$x_2 = A_{2,3}$	$C_2 = e_1$	0.83	79/97
3	$x_1 = A_{1,2}$	$C_3 = e_1$	0.80	94/120
4	$(x_1 = A_{1,1} \text{ ODER } A_{1,2}) \text{ UND } (x_2 = A_{2,3})$	$C_4 = e_1$	0.77	17/26
5	$(x_1 = A_{1,2} \text{ ODER } A_{1,3}) \text{ UND } (x_2 = A_{2,3})$	$C_5 = e_1$	0.88	47/54
6	$(x_1 = A_{1,2}) \text{ UND } (x_2 = A_{2,2} \text{ ODER } A_{2,3})$	$C_6 = e_1$	0.91	60/66
7	$(x_1 = A_{1,2}) \text{ UND } (x_2 = A_{2,3} \text{ ODER } A_{2,4})$	$C_7 = e_1$	0.67	21/37

Tabelle 3.2: Hypothesen im ersten Pruningschritt zu Regel 5 (hier Nr. 1) aus Bild 3.2 inklusive Bewertung.

Merkmale x_1 und x_2 und Bild 3.4 zeigt die entsprechenden Hypothesen im Bewertungsgraphen. Die zweite Regel in Bild 3.3 entsteht durch Weglassen der Teilprämisse $x_2 = A_{2,3}$ der ersten Regel.

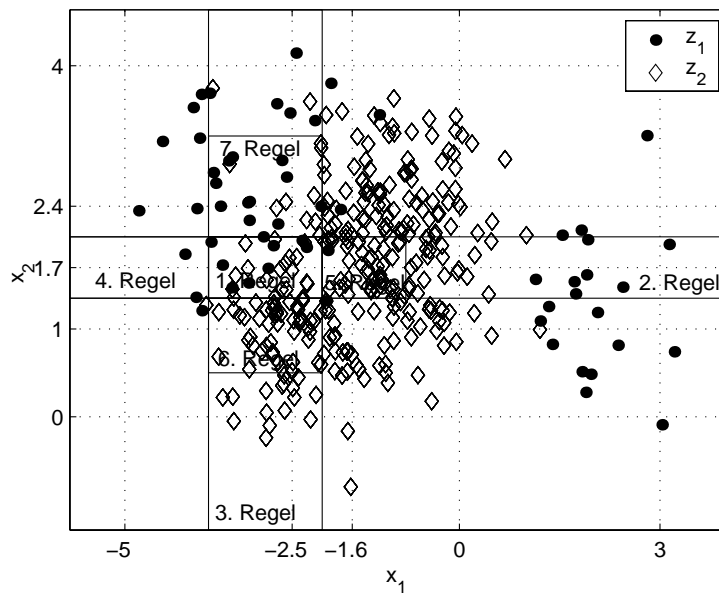


Bild 3.3: Darstellung der Hypothesen aus Tabelle 3.2. $C_i = e_1$.

Die siebte Hypothese ist die beste, da für sie der Wert von $\hat{L}_{D,P_r,Z}$ im Vergleich mit den anderen Hypothesen minimal ist (in Tabelle 3.2 fett markiert). Diese Information ist auch aus Bild 3.4 zu entnehmen. Die durchgezogenen Linien zeigen die Linien gleicher Kosten, zur oberen linken Ecke hin nehmen die Kosten ab. Die siebte Hypothese weist auch hier die niedrigsten Kosten auf. Die siebte Hypothese wird für den nächsten Pruningschritt übernommen. Die anderen Hypothesen werden gelöscht.

Für Regel 5 aus Bild 3.2 sind in Bild 3.5 die besten Regeln aus allen Pruningschritten dargestellt. Im ersten Pruningschritt wird der Teilprämisse $x_2 = A_{2,3}$ die Teilprämiss-

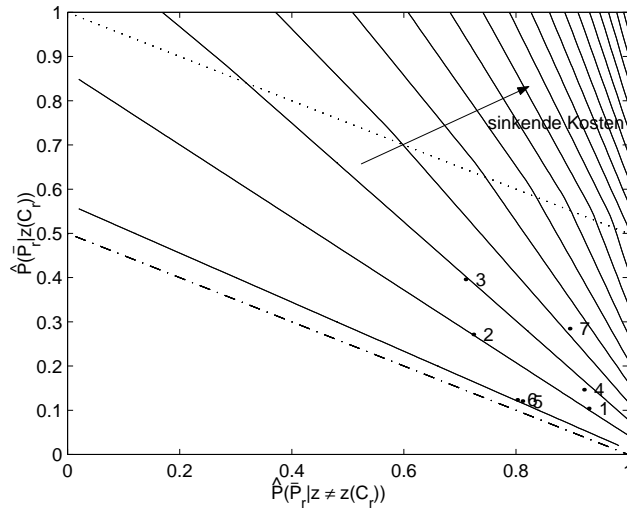


Bild 3.4: Darstellung der Hypothesen aus Tabelle 3.2 im Bewertungsgraphen.

se $x_2 = A_{2,4}$ hinzugefügt. Im zweiten Schritt erfolgt die Hinzunahme von $x_1 = A_{1,1}$ und schließlich wird im letzten Schritt noch die Teilprämisse $x_2 = A_{2,5}$ angefügt. Jede weitere Generalisierung führt zu einer Verschlechterung der Regel nach (2.102).

Pruning mit Merkmalskosten

Die Minimierung der Merkmalskosten des Entscheidungssystems erfordert deren Berücksichtigung beim Pruning von Einzelregeln. Diese Aussage lässt sich anhand des Beispieldatensatzes nachvollziehen. Dazu seien die Merkmalskosten von x_2 höher als die Kosten einer Default-Entscheidung. Für die Auswahl einer Regelbasis kommen nur solche Regeln in Frage, deren Prämissen das Merkmal x_2 nicht auswerten. Wird dennoch eine Regel mit x_2 ausgewählt, sind die Kosten pro Entscheidung in jedem Fall höher als bei einer Default-Entscheidung.

Im einfachsten Fall werden die Merkmalskosten einer Regel $L_{C,R}$ direkt zu den Entscheidungskosten $\hat{L}_{D,P_r,Z}$ hinzuaddiert:

$$\hat{L}_{T,R} = \hat{L}_{D,P_r,Z} + L_{C,R} \tag{3.9}$$

Dies kann allerdings dazu führen, dass Regeln zu weit generalisiert werden. Dieser Fall tritt ein, wenn die Reduktion der Merkmalskosten durch das Weglassen einer Teilprämisse größer ist als eine dadurch verursachte Erhöhung der Entscheidungskosten. Die Regel wird dann besser bewertet und die vorhergehende gelöscht. Es handelt sich somit um ein Optimierungsproblem mit den zwei Kriterien:

$$\min \hat{L}_{D,P_r,Z} \tag{3.10}$$

$$\min \hat{L}_{T,R} . \tag{3.11}$$

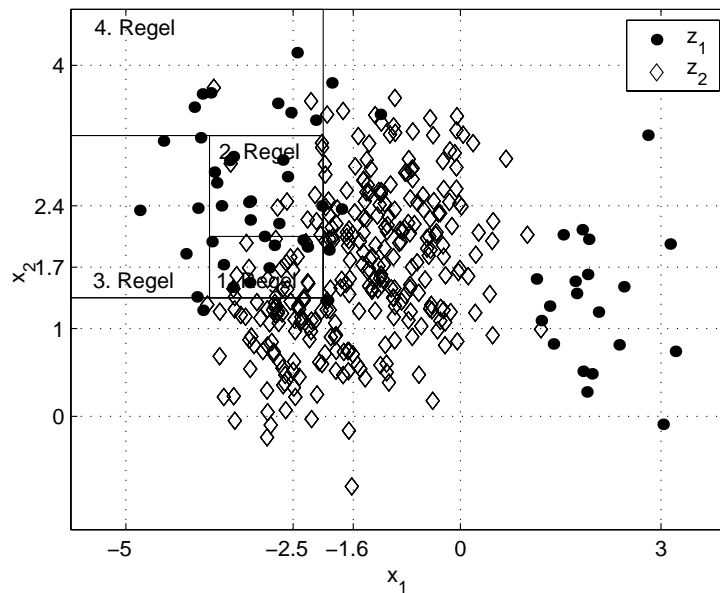


Bild 3.5: Alle Pruningschritte für Regel 5 aus Bild 3.2. Regel 1 entspricht Regel 5.

Beim Pruning der einzelnen Regeln ist noch ungewiss, welche Merkmale in der zu generierenden Regelbasis verwendet werden. Die Generalisierung einer Regel zugunsten geringerer Merkmalskosten kann somit überflüssig sein. Betrachtet wird eine Regel, bei der die Streichung einer Teilprämisse zu höheren Entscheidungskosten $\hat{L}_{D,P_r,Z}$, aber niedrigeren Gesamtkosten $\hat{L}_{T,R}$ führt. Wenn das Merkmal der gestrichenen Teilprämisse von einer anderen in die Regelbasis aufgenommenen Regel verwendet wird, war die Generalisierung überflüssig.

Aus den angeführten Überlegungen wird ein neues Dominanzkriterium für das Pruning von Einzelregeln bei Berücksichtigung von Merkmalskosten abgeleitet. Dieses Kriterium erlaubt es, eine oder mehrere zusätzliche Regeln aus einem Pruningschritt für das weitere Pruning zu übernehmen. Die übernommenen Regeln sollen von keiner anderen Regel im aktuellen Pruningschritt dominiert werden. Dazu sind folgende Fälle für die Regelhypothesen eines Pruningschrittes im Vergleich zur besten Hypothese des Pruningschrittes zu unterscheiden:

- bei gleichen Merkmalskosten liegen unterschiedliche Entscheidungskosten vor,
- bei gleichen Entscheidungskosten liegen unterschiedliche Merkmalskosten vor,
- bei niedrigeren Merkmalskosten liegen niedrigere Entscheidungskosten vor,
- bei niedrigeren Merkmalskosten liegen höhere Entscheidungskosten vor.

Die ersten drei Fälle sind eindeutig. Bei gleichen Merkmalskosten dominiert die Regel mit den niedrigeren Entscheidungskosten. Bei identischen Entscheidungskosten ist die Regel mit den niedrigeren Merkmalskosten dominant. Sind bei einer Regel beide Kostenarten

niedriger, so ist diese Regel ebenfalls dominant. Im vierten Fall sind zusätzlich die Gesamtkosten zu berücksichtigen. Betrachtet werden zwei Regeln R_1 und R_2 . Dabei sei R_1 die beste Regel des aktuellen Pruningschrittes nach (2.102). Liegt die folgende Situation vor:

$$\begin{aligned}\hat{L}_{D,P_1,Z} &< \hat{L}_{D,P_2,Z} \\ L_{C,R_1} &> L_{C,R_2} \\ \hat{L}_{T,R_1} &> \hat{L}_{T,R_2},\end{aligned}\tag{3.12}$$

ist ohne Berücksichtigung von Merkmalskosten (nur \hat{L}_D) R_1 vorzuziehen. Bei der Berücksichtigung von Merkmalskosten nach (3.9) ist dagegen R_2 zu wählen, da die Gesamtkosten niedriger sind. In diesem Fall sollen beide Regeln für das weitere Pruning verwendet werden.

Zu betrachten ist nun, wann der oben beschriebene Fall auftreten kann. Die diskutierten Zusammenhänge sind auch in Bild 3.6 grafisch dargestellt.

Hinzufügen benachbarter Terme

Durch das Hinzufügen benachbarter Terme werden die Merkmalskosten einer Regel nicht verändert. Eine Regel, deren Entscheidungskosten sinken, dominiert somit die vorhergehende Regel ohne den benachbarten Term. Steigen die Entscheidungskosten, so wird die generalisierte Regel von der vorhergehenden dominiert. Als Kriterium für die Entscheidung, welche Regel für das weitere Pruning zu übernehmen ist, reichen die Entscheidungskosten \hat{L}_D aus.

Weglassen von Teilprämissen

Führt das Weglassen einer Teilprämisse zu niedrigeren Entscheidungskosten, so dominiert diese Regel ihre Vorgängerin, da die Merkmalskosten entweder gleich bleiben oder sinken. Führt das Weglassen einer Teilprämisse dagegen zu höheren Entscheidungskosten, so können die Gesamtkosten durch verringerte Merkmalskosten dennoch niedriger sein. Diese Regel dominiert die ursprüngliche allerdings nicht. In diesem Fall sind beide Regeln für das weitere Pruning vorzusehen.

Bild 3.6 zeigt die Dominanzbeziehungen grafisch. Auf den Achsen sind die Entscheidungskosten und die Merkmalskosten aufgetragen. Dadurch entsprechen Geraden mit der Steigung $m = -1$ Linien gleicher Gesamtkosten L_T . Die durch den grauen Punkt markierte Regel sei die Ursprungshypothese eines Pruningschrittes. Durch Hinzufügen von benachbarten Termen entstehen die durch die Zeichen \oplus und \odot markierten Regeln mit den selben Merkmalskosten wie die Ursprungshypothese. Die mit \oplus markierte Regel dominiert die Ursprungshypothese und die mit \odot markierte Regel. Das Weglassen unterschiedlicher Teilprämissen führt zu weiteren mit den Zeichen \oplus und \odot gekennzeichneten Regeln, die allerdings niedrigere Merkmalskosten haben. Die mit \oplus markierten Regeln dominieren die Ursprungshypothese. Die mit \odot markierte nicht. Die drei mit \oplus gekennzeichneten Regeln dominieren jede für sich die Ursprungshypothese. Untereinander liegt für je zwei der mit \oplus gekennzeichneten Regeln der Fall (3.12) vor. Zu den Kandidaten für das weitere Pruning gehören die drei mit \oplus gekennzeichneten Regeln. Verallgemeinert werden stets die Regeln

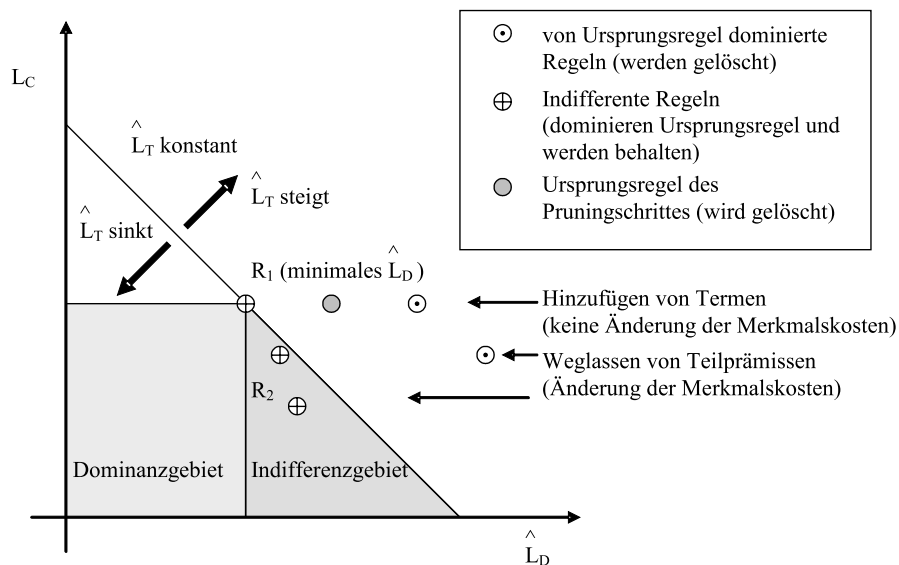


Bild 3.6: Darstellung der Dominanzsituationen beim Pruning von Einzelregeln. Der Vergleich bezieht sich stets auf die Regel mit dem schwarzen Punkt.

r	WENN	DANN	$\hat{L}_{D,P_r,Z}$	$\hat{L}_{T,P_r,Z}$	Anz./Fehl.
1	$(x_1 = A_{1,1}$ ODER $A_{1,2})$ UND $(x_2 = A_{2,3}$ ODER $A_{2,4}$ ODER $A_{2,5})$	$C_1 = e_1$	0.63	1.03	29/63
2	$x_1 = A_{1,1}$ ODER $A_{1,2}$	$C_2 = e_1$	0.76	0.96	95/130
3	$x_2 = A_{2,3}$ ODER $A_{2,4}$ ODER $A_{2,5}$	$C_3 = e_1$	0.80	1.00	182/230

Tabelle 3.3: Pruningergebnis von Regel 5 aus Bild 3.2 bei Berücksichtigung von Merkmalskosten und dem Dominanzkriterium.

übernommen, die im Indifferenzgebiet der besten Regel gemäß (3.10) des aktuellen Pruningsschrittes liegen. Das Indifferenzgebiet wird durch eine senkrechte Gerade zur Abszisse, die Abszisse und die Isokostenlinie durch die Markierung der besten Regel aufgespannt.

Bild 3.7 zeigt exemplarisch die resultierenden Regeln des Prunings bei Berücksichtigung von Merkmalskosten und Verwendung des Dominanzkriteriums für den Beispieldatensatz. Ausgangsregel ist Nr. 5 aus dem Baum (Bild 3.2). Die Merkmalskosten für x_1 und x_2 betragen jeweils $L_C = 0.2$. Das Pruning mit Merkmalskosten liefert in diesem Fall zusätzlich zum Ergebnis des Prunings ohne Merkmalskosten (in Bild 3.7 mit 1. Regel markiert) zwei zusätzliche Regeln (2. Regel und 3. Regel). Die zusätzlichen Regeln verwenden jeweils nur ein Merkmal und haben dadurch niedrigere Merkmalskosten als Regel 1. In Tabelle 3.3 sind die Regeln und ihre Bewertungen zusammengestellt. Ohne das Dominanzkriterium resultiert lediglich die stark generalisierte Regel 2 aus dem Pruning von der fünften Regel des Baumes. Die zusätzliche Einführung des neuen Dominanzkriteriums führt somit zum Erhalt wertvoller Regeln für die spätere Regelsuche.

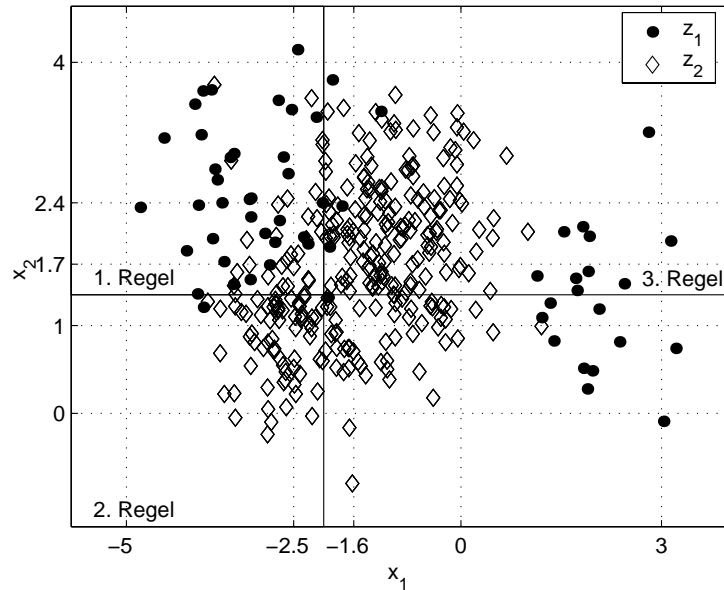


Bild 3.7: Pruningergebnis von Regel 5 aus Bild 3.2 bei Berücksichtigung von Merkmalskosten und dem Dominanzkriterium.

Die Generierung der Fuzzy-Einzelregeln ist nun abgeschlossen. Im nächsten Abschnitt werden modifizierte entscheidungstheoretische Verfahren zur Auswahl geeigneter Regeln aus der Menge der Einzelregeln und zur Bildung von Regelbasen vorgestellt.

3.3 Hypothesengenerierung - einstufige Regelbasen

3.3.1 Eröffnungsverfahren

Wie beim Pruning ist es bei der Suche nach Regelbasen nicht möglich, alle denkbaren Kombinationen aus den Regelhypothesen zu evaluieren. Aus $N_{R,Hyp} = 50$ Regeln ergeben sich ca. 2.1 Millionen Möglichkeiten für die Auswahl von $N_R = 5$ Regeln:

$$N_{RB} = \binom{N_{R,Hyp}}{N_R} = \frac{N_{R,Hyp}!}{N_R! \cdot (N_{R,Hyp} - N_R)!}. \quad (3.13)$$

Eine Überprüfung aller möglichen Kombinationen führt zu inakzeptabel langen Rechenzeiten. Für die Suche einer Regelbasis ist daher wie beim Pruning ein heuristisches Verfahren anzuwenden.

Das Eröffnungsverfahren beginnt mit einer Default-Entscheidung (SONST-Regel) für alle Beispiele des Lerndatensatzes und einer leeren Regelbasis ($N_R = 0$). Anschließend werden der Regelbasis Schritt für Schritt Regeln hinzugefügt. Dabei wird stets die Regel R_i^{Hyp} aus der Menge aller Regelhypothesen ausgewählt, die Kosten \hat{L}_T pro Entscheidung minimiert. Die ausgewählte Regel wird an der Stelle $N_R + 1$ in die Regelbasis einge-

fügt:

$$R_{N_R+1}^{RB} = \arg \min_{i \in \{1 \dots N_{R,Hyp}, \emptyset\}} \hat{L}_T(R_1^{RB}, \dots, R_{N_R}^{RB}, R_i^{Hyp}). \quad (3.14)$$

Dabei ist $N_{R,Hyp}$ die Anzahl der Regeln nach dem Pruning. Die Merkmalskosten L_C pro Entscheidung können optional Null gesetzt werden. Die Suche nach weiteren Regeln bricht ab, wenn keine Reduktion der Kosten durch eine weitere Regel eintritt. Um große Regelbasen zu vermeiden, kann eine Schranke für die minimale Kostenreduktion einer zusätzlichen Regel gesetzt werden. Zusätzlich ist eine Obergrenze für die Anzahl der Regeln möglich. Für die Auswahl einer Regelbasis mit $N_R = 5$ Regeln aus $N_{R,Hyp} = 50$ möglichen Regeln werden dann 285 Regelbasen untersucht, wenn das Verfahren im sechsten Durchlauf keine zusätzliche Regel aufnimmt. In jedem Durchlauf ist eine Regelbasis weniger zu untersuchen. Daher ergibt sich der Aufwand der Heuristik allgemein zu:

$$N_{RB} = \sum_{k=1}^{N_R+1} N_{R,Hyp} - (k - 1). \quad (3.15)$$

Die Bewertung der Regelbasen erfolgt entweder als vollständige oder als unvollständige Regelbasis (siehe Abschnitt 2.5). Soll die Bewertung als vollständige Regelbasis erfolgen, sind die bereits gewählten Prämissen um eine SONST-Regel zu erweitern. Die Konklusion dieser Regel stellt einen zusätzlichen Parameter der Suche dar. Für die Bewertung als unvollständige Regelbasis sind die Kosten \hat{L}_{RW} der bisher nicht entschiedenen Beispiele festzulegen. Insgesamt stehen die folgenden Optionen zur Verfügung:

1. Die Bewertung erfolgt als vollständige Regelbasis. Eine SONST-Regel mit vom Anwender vorgegebener Default-Entscheidung wird der Regelbasis hinzugefügt.
2. Die Bewertung erfolgt als vollständige Regelbasis. Eine SONST-Regel mit vom Algorithmus festgelegter Entscheidung wird der Regelbasis hinzugefügt. Dies entspricht der pessimistischen Annahme, dass keine weitere Regel mehr zu finden ist.
3. Die Bewertung erfolgt als unvollständige Regelbasis. Eine SONST-Regel mit der Rückweisungsentscheidung wird der Regelbasis hinzugefügt. Die Kosten \hat{L}_{RW} je Beispiel für diese Entscheidung werden vom Algorithmus bestimmt.
4. Wie 3., aber für die restlichen Beispiele werden vom Anwender vorgegebene Kosten angesetzt. Diese Option ist für die Evidenz-Inferenz konzipiert.

Nach Abbruch der Regelsuche wird die fest vorgegebene oder eine vom Algorithmus bestimmte Entscheidung für die restlichen Beispiele getroffen (SONST DANN...). Bei der Evidenz-Inferenz wird für die restlichen Beispiele keine Entscheidung getroffen. Die Konklusion der SONST-Regel bleibt die Rückweisungsentscheidung. Die einzelnen Optionen werden im Folgenden anhand des Beispieldatensatzes diskutiert.

Feste Default-Entscheidung

Den Beispielen, die von der SONST-Regel abgedeckt werden, wird eine manuell vorgegebene Default-Entscheidung zugewiesen. Dieses Vorgehen ist sinnvoll, wenn insbesondere

Regeln für eine seltene Klasse in die Regelbasis aufgenommen werden sollen. Als Beispiel sollen Regeln bevorzugt ausgewählt werden, deren Konklusion der optimalen Entscheidung für den Umweltzustand z_1 entspricht. Die feste Default-Entscheidung wird dann auf die beste Entscheidung für die Beispiele des Umweltzustandes z_2 gesetzt. Alle Regeln, deren Konklusionen identisch zur Default-Entscheidung sind, führen nicht zu einer Reduktion der Kosten. Für die von den Prämissen dieser Regeln abgedeckten Beispiele wird über die Default-Entscheidung bereits die beste Entscheidung getroffen. Sollen für alle Umweltzustände Regeln ausgewählt werden, ist die feste Default-Entscheidung nicht geeignet. Die Prämissen von Regelbasen, die mit fester Default-Entscheidung zusammengestellt werden, decken in der Regel nicht alle Beispiele des Lerndatensatzes ab. Allerdings ist die Zahl der ausgewählten Regeln tendenziell niedriger als bei den anderen Optionen. Bild 3.8 zeigt das Ergebnis für den Beispieldatensatz für die feste Default-Entscheidung. Links ist e_1 die Default-Entscheidung, rechts e_2 . Für dieses Beispiel sind die Ergebnisse der beiden Regelbasen strukturell identisch. Der Bereich im linken Teil, der bei e_1 als Default-Entscheidung zur SONST-Regel wird, wird bei e_2 als Default-Entscheidung von den drei Prämissen mit jeweils der Konklusion e_1 erfasst. Umgekehrt wird der Bereich im rechten Teil, der bei e_2 als Default-Entscheidung zur SONST-Regel wird, bei e_1 als Default-Entscheidung von zwei Prämissen mit der Entscheidung e_2 abgedeckt.

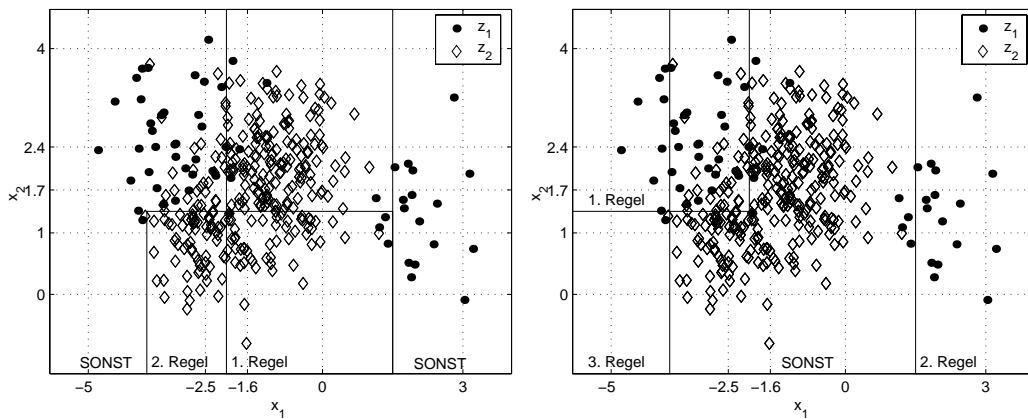


Bild 3.8: Links: Suche einer Regelbasis mit fester Default-Entscheidung e_1 . Rechts: Suche einer Regelbasis mit fester Default-Entscheidung e_2 .

Automatische Default-Entscheidung

Den Beispielen, die von der SONST-Regel abgedeckt werden, wird ebenfalls eine Default-Entscheidung zugewiesen. Diese wird allerdings vom Algorithmus für jede mögliche zusätzliche Regel neu berechnet [101]:

$$e_{opt,def} = \arg \min_{e_k} \sum_{j=1}^{N_z} L(e_k, z_j) \cdot \hat{P}(z_j | \overline{\bigcup_{r=1}^{N_R} P_r}) \quad (3.16)$$

Der Regelbasis werden bevorzugt solche Regeln hinzugefügt, deren Konklusionen von der aktuellen Default-Entscheidung verschieden sind. Die Prämissen einer Regelbasis, die mit

automatischer Default-Entscheidung erstellt werden, decken ebenfalls häufig nicht alle Beispiele des Lerndatensatzes ab. Durch das Hinzufügen einer weiteren Regel mit der Konklusion der aktuellen Default-Entscheidung, die keine widersprüchliche Überlappung mit den bereits ausgewählten Regeln aufweist, reduzieren sich die Kosten nicht. Bild 3.9 zeigt das Ergebnis für den Beispieldatensatz für die automatische Default-Entscheidung. Das Gebiet der SONST-Regel ist deutlich kleiner als bei fester Default-Entscheidung. In die Regelbasis werden vier Regeln aufgenommen. Eine Änderung der Kosten tritt hier nicht ein, wenn zum Beispiel die Regel:

$$\text{WENN } (x_1 = A_{1,1} \text{ ODER } A_{1,2}) \text{ UND} \\ (x_2 = A_{2,3} \text{ ODER } A_{2,4} \text{ ODER } A_{2,5}) \text{ DANN } C_1 = e_1,$$

der Regelbasis zusätzlich hinzugefügt wird, da durch die zusätzliche Regel für kein Beispiel die Entscheidung geändert wird.

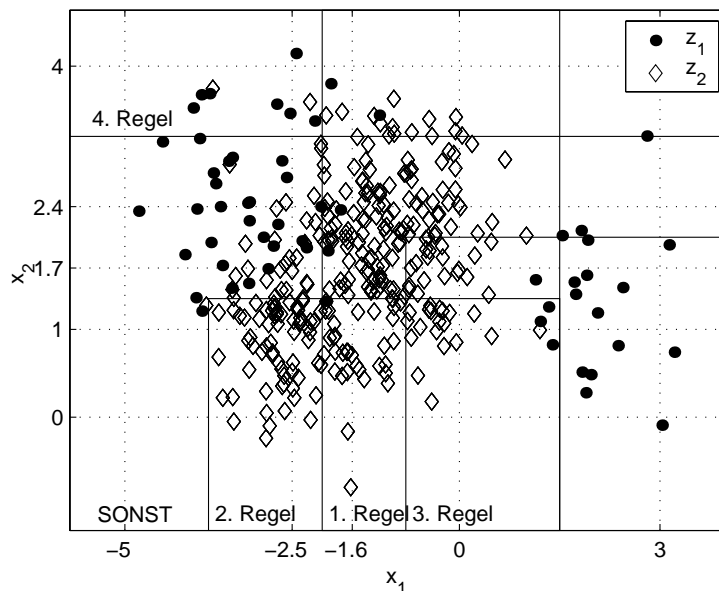


Bild 3.9: Suche einer Regelbasis mit automatischer Default-Entscheidung.

Rückweisungsentscheidung

Für die restlichen Beispiele wird eine Rückweisungsentscheidung getroffen. Diese Beispiele werden allerdings nicht von der Bewertung ausgenommen. Es ist jedoch noch ungewiss, ob noch aufzunehmende Regeln für diese Beispiele die besten Entscheidungen treffen werden oder nicht. Wird keine weitere Regel aufgenommen, bricht die Suche ab. Dann wird den Beispielen in der Rückweisung über die SONST-Regel eine Entscheidung zugewiesen. Die beste Entscheidung $e_{opt,def}$ für die restlichen Beispiele wird nach (3.16) bestimmt. Bild 3.10 zeigt die Aufteilung der Kosten. Zu Beginn der Regelsuche ist noch keine Regel in die Basis aufgenommen und wird für alle Beispiele die Rückweisungsentscheidung

getroffen. Daher ist $\hat{P}_{RW} = 1$. Die tatsächlichen Kosten liegen im Intervall zwischen 0 und \hat{L}_{RW} . Für jede zusätzliche Regel werden für einen Teil der Beispiele $(1 - \hat{P}_{RW})$ Entscheidungen getroffen und damit die Kosten $\hat{L}_{D,P_{RB}}$ für diese Beispiele über eine Risikosituation festgelegt. Die Gesamtkosten sinken $\hat{L}_{D,P_{RB},RW} < \hat{L}_{RW}$ und es gilt: $\hat{P}_{RW} < 1$. Somit variiert die Aufteilung in Kosten der Prämisse und Kosten der Rückweisung. Nach Abschluss der Regelsuche verschwindet der Teil für die Rückweisungsentscheidung. Die Kosten werden nun für alle Beispiele über eine Risikosituation bestimmt. Eine zusätzliche Regel vergrößert den Anteil der nicht entschiedenen Beispiele nicht. Festzulegen sind nun die Kosten \hat{L}_{RW} für die Beispiele in der Rückweisung.

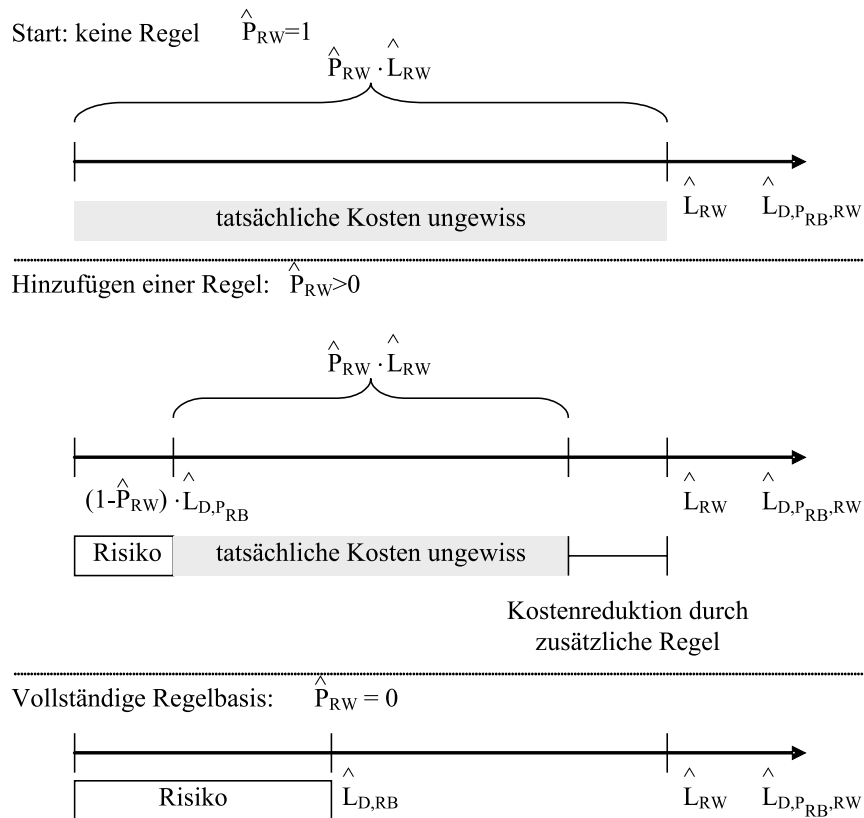


Bild 3.10: Aufteilung der Entscheidungskosten bei der Regelbasissuche. Die Kosten der Rückweisung \hat{L}_{RW} sind als konstant angenommen.

Eine Möglichkeit für \hat{L}_{RW} sind die Kosten pro Entscheidung, die entstehen, wenn für die restlichen Beispiele die aktuell beste Default-Entscheidung $e_{opt,def}$ nach (3.16) getroffen wird:

$$\hat{L}_{RW,def} = \sum_{j=1}^{N_Z} L(e_{opt,def}, z_j) \cdot \hat{P}(z_j | \mu(P_{RB}) < 0.5) \tag{3.17}$$

Als zweite Möglichkeit stehen die Kosten der besten Default-Entscheidung bezogen auf den gesamten Datensatz gemäß (2.16) zur Verfügung:

$$\hat{L}_{RW,0} = \sum_{j=1}^{N_j} L(e_{opt}, z_j) \cdot \hat{P}(z_j). \quad (3.18)$$

Bei der ersten Möglichkeit besteht eine Abhängigkeit zwischen \hat{L}_{RW} und der aktuell betrachteten zusätzlichen Regel. Bei der zweiten bleibt \hat{L}_{RW} während der gesamten Suche konstant. Wie bei der automatischen Rückweisung werden Regeln nicht aufgenommen, die keine widersprüchliche Überlappung mit bereits ausgewählten Prämissen haben und $e_{opt,def}$ nicht ändern.

Bild 3.11 zeigt das Ergebnis für den Beispieldatensatz für die Rückweisungsentscheidung. Links wird \hat{L}_{RW} für jede zusätzliche Regel nach (3.17) berechnet. Rechts ist \hat{L}_{RW} konstant (3.18). Bis zur dritten ausgewählten Regel ist der Suchvorgang für beide Optionen identisch. Die Bewertungen für die die Regelbasen mit R_a beziehungsweise R_b als vierte Regel sind in Tabelle 3.4 zusammengefasst. Zusätzlich sind in der Tabelle die Anzahl der getroffenen Entscheidungen und die Anzahl Beispiele in der Rückweisung für z_1 und z_2 zusammengestellt. Weiterhin sind die Kosten der Prämissen $\hat{L}_{D,PRB}$ angegeben. Durch die geringeren Kosten für die nicht erfassten Beispiele wird die speziellere Regel R_a bei der Verwendung von $\hat{L}_{RW,def}$ bevorzugt. Im anderen Fall erhält R_b die bessere Bewertung. $\hat{L}_{RW,0}$ ist für beide Regeln gleich. Tendenziell erhöht sich die Prämissenabdeckung bei der Bewertung von Regelbasen mit $L_{RW,0}$.

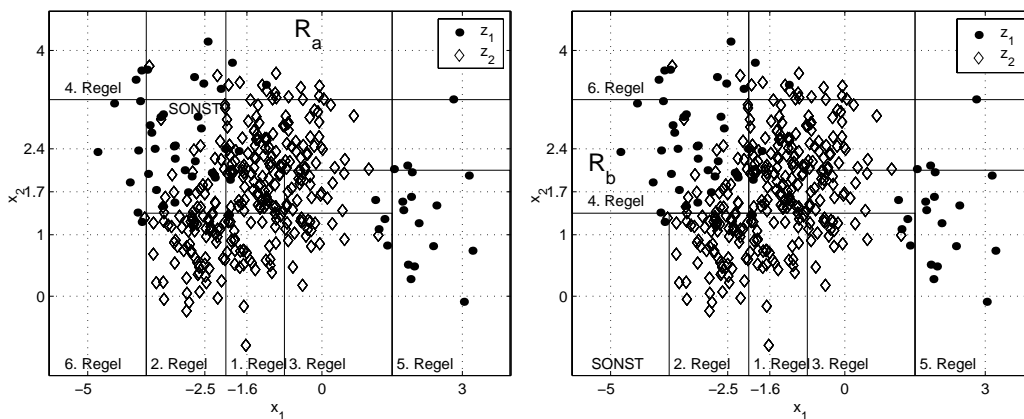


Bild 3.11: Links: Ergebnis der Suche einer Regelbasis mit Rückweisungsentscheidung $\hat{L}_{RW,def}$. Rechts: Ergebnis der Suche einer Regelbasis mit Rückweisungsentscheidung $\hat{L}_{RW,0}$.

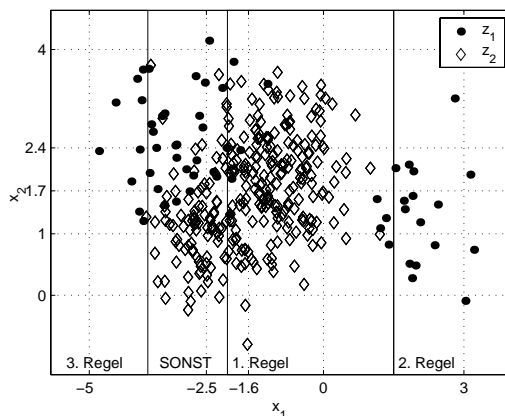
Merkmalskosten

Die Berücksichtigung der Merkmalskosten ist optional. Bild 3.12 links zeigt das Ergebnis der Regelbasissuche mit Rückweisungsentscheidung bei Berücksichtigung von Merkmalskosten. Die Merkmalskosten betragen dabei $L_{C,1} = L_{C,2} = 0.3$. Für die Regelbasis werden

	R_a		R_b	
	z_1	z_2	z_1	z_2
e_1	26	5	51	32
e_2	5	266	6	268
RW	29	29	3	10
$\hat{L}_{D,PRB}$	0.18		0.26	
$\hat{L}_{RW,def}$	0.5		0	
$\hat{L}_{RW,0}$	0.83		0.83	
$\hat{P}(RW)$	0.16		0.01	
$\hat{L}_{D,PRB,RW,def}$ (3.17)	0.23		0.26	
$\hat{L}_{D,PRB,RW,0}$ (3.18)	0.29		0.26	

Tabelle 3.4: Bewertung der vierten Regel (R_a, R_b) für \hat{L}_{RW} nach (3.17) beziehungsweise (3.18). Fettdruck: ausgewählte Regel.

nur Regeln, die ausschließlich Teilprämissen des Merkmals x_1 enthalten, ausgewählt. Die Reduktion der Entscheidungskosten durch Regeln, die zusätzlich das Merkmal x_2 verwenden, wird über die höheren Merkmalskosten mit dem Merkmal wieder ausgeglichen. In Bild 3.12 rechts erfolgt ein Vergleich der getroffenen Entscheidungen und der Kostenarten mit und ohne Berücksichtigung von Merkmalskosten. Die Gesamtkosten $L_{T,RB}$ sind in diesem Fall vergleichbar. Da mit x_1 nur ein Merkmal verwendet wird, ist die Regelbasis bei Berücksichtigung von Merkmalskosten hier deutlich einfacher aufgebaut, verursacht aber erheblich mehr Fehlentscheidungen. Die Entscheidungskosten $\hat{L}_{D,RB} = 0.51$ sind wesentlich höher als ohne Berücksichtigung von Merkmalskosten ($\hat{L}_{D,RB} = 0.22$).



	mit L_C		ohne L_C	
	z_1	z_2	z_1	z_2
e_1	51	9	56	4
e_2	95	205	38	262
$\hat{L}_{D,RB}$	0.51		0.22	
\hat{L}_C	0.3		0.6	
$\hat{L}_{T,RB}$	0.81		0.82	

Bild 3.12: Links: Regelbasis bei Berücksichtigung von Merkmalskosten. Rechts: Ergebnisse für die Regelbasissuche mit Rückweisungsentscheidung mit und ohne Merkmalskosten. $L_{C,1} = L_{C,2} = 0.3, L_R = 10$.

Evidenz-Inferenz

Bei Anwendung der Evidenz-Inferenz kann auch für Beispiele mit:

$$\mu(P_{RB})[n] \geq 0.5,$$

die Rückweisungsentscheidung getroffen werden. Im Gegensatz zur bisherigen Strategie mit Rückweisungsentscheidung wird diese Rückweisung nicht im letzten Suchschritt wieder durch eine Default-Entscheidung ersetzt, sondern zur nachfolgenden Entscheidung durch einen Experten beibehalten. Die Rückweisung eines Beispiels liegt entweder an einem zu geringen Glaubensgrad für die einzelnen Entscheidungen oder an einem zu großen Widerspruch der aktivierten Regeln. Daher ist es möglich, dass eine zusätzliche Regel den Anteil der nicht entschiedenen Beispiele erhöht. Die Kosten $\hat{L}_{RW,ev}$ sind vom Anwender vorzugeben. Es handelt sich dabei um die Gesamtkosten pro Entscheidung des Experten. Die Kosten können zum Beispiel aus Merkmalskosten und Entscheidungskosten je Beispiel für den Experten zusammengesetzt sein. Die Kosten einer Regelbasis ergeben sich dann zu:

$$L_{T,R_{NR},Ev} = (1 - \hat{P}_{RW}) \cdot \hat{L}_{D,P_{RB}} + \hat{P}_{RW} \cdot \hat{L}_{RW,ev}. \tag{3.19}$$

Der Ablauf der Suche ist identisch zur Suche mit Rückweisungsentscheidung. Bild 3.13 zeigt das Ergebnis der Regelsuche mit Evidenz-Inferenz. Als Kosten des Experten sind dabei $\hat{L}_{RW,ev} = 0.3$ angesetzt. Links ist die Regelbasis grafisch dargestellt und in der Tabelle rechts sind die getroffenen Entscheidungen und die Bewertungen zusammengestellt. Bei dem Beispieldatensatz wird ein großer Teil der Beispiele nicht entschieden $\hat{P}(RW) = 0.59$. Die Kosten pro Entscheidung für die entschiedenen Beispiele sind mit 0.27 niedriger als die Kosten des Experten. Dieses Ergebnis ist zu erwarten, da bei Entscheidungskosten $\hat{L}_{D,RB} > \hat{L}_{RW,ev}$ der Experte die Entscheidungen mit niedrigeren Kosten treffen kann.

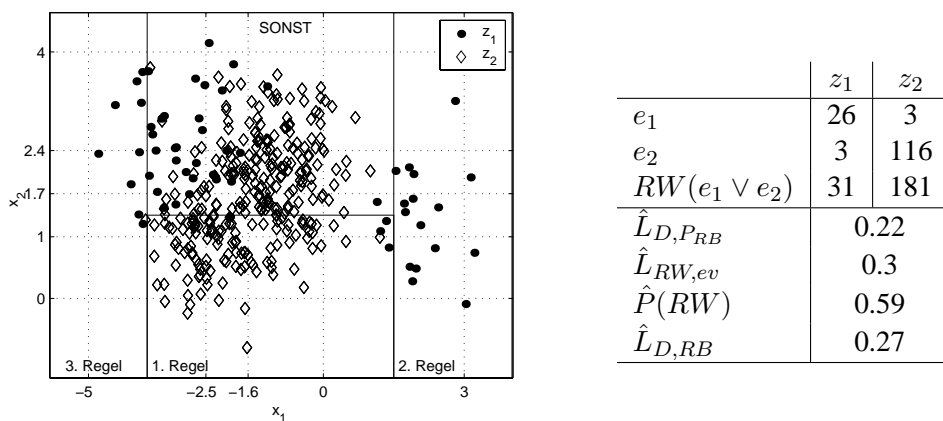


Bild 3.13: Links: Suche einer Regelbasis mit Evidenz-Inferenz. $\hat{L}_{RW,ev} = 0.3$. Rechts: Ergebnisse.

	$\hat{P}(RW)$	$\hat{L}_{D,RB}$	Bild
Feste Default-Entscheidung e_1	0.26	0.34	3.8
Feste Default-Entscheidung e_2	0.73	0.34	3.8
Automatische Default-Entscheidung	0.19	0.22	3.9
Rückweisungsentscheidung $\hat{L}_{RW,def}$	0.13	0.22	3.11
Rückweisungsentscheidung $\hat{L}_{RW,0}$	0.02	0.22	3.11
Evidenz-Inferenz $\hat{L}_{RW,ev} = 0.3$	0.59	0.27	3.13

Tabelle 3.5: Übersicht der Ergebnisse der Regelsuche für den Beispieldatensatz und einem Kostenverhältnis von $L_R = 10$.

Vergleich

Tabelle 3.5 fasst die Ergebnisse für die verschiedenen Suchoptionen der Regelbasissuche nochmals zusammen. Danach führt die Option Rückweisungsentscheidung in Verbindung mit konstanten Default-Kosten $\hat{L}_{RW,0}$ zur Regelbasis mit der größten Prämissenabdeckung. Diese Tendenz bestätigen Untersuchungen mit weiteren Datensätzen. Zum Beispiel ergibt sich für den Heart-Disease Datensatz (siehe Abschnitt 4.2.2) bei der Regelbasissuche mit der Option Rückweisungsentscheidung und konstanten Default-Kosten $\hat{L}_{RW,0}$ ein Wert von $\hat{P}_{RW} = 0.04$, während bei der Verwendung von $\hat{L}_{RW,def}$ die Wahrscheinlichkeit $\hat{P}_{RW} = 0.10$ ist. Der Anteil der nicht entschiedenen Beispiele ist bei fester Default-Entscheidung höher als bei der Rückweisungsentscheidung. Auch diese Tendenz bestätigen Untersuchungen mit weiteren Datensätzen. Beim Heart-Disease stellen sich für die automatische Default-Entscheidung ($\hat{P}_{RW} = 0.26$) und die Feste Default-Entscheidung e_2 ($\hat{P}_{RW} = 0.56$) ebenfalls deutlich höhere Werte ein. Die feste Default-Entscheidung liefert insbesondere bei stark ungleichmäßigen Wahrscheinlichkeiten der Umweltzustände gute Ergebnisse (siehe Abschnitt 4.3). Liegt kein besonderes Ungleichgewicht vor, wird die Verwendung der Rückweisungsentscheidung empfohlen. Dabei wird die Verwendung von $\alpha_{RB} = 1$ vorgeschlagen. Eine eingehende Untersuchung der Zusammenhänge zwischen dem Wert von α_{RB} und der Güte der gefundenen Regelbasis steht noch aus.

3.3.2 Verbesserungsverfahren - Genetischer Algorithmus

Viele Probleme bei der Generierung von Regelbasen können auf das iterative Hinzufügen von Einzelregeln zurückgeführt werden, das zu suboptimalen Lösungen führt. In diesem Abschnitt wird untersucht, ob ein Verfahren, dessen Hypothesen vollständige Regelbasen sind, vorteilhaft ist.

Als Verbesserungsverfahren für eine Lösung, die durch iteratives Hinzufügen von Regeln gefunden wurde, wird in diesem Abschnitt ein genetischer Algorithmus vorgeschlagen. Genetische Algorithmen haben ihren Ursprung in den Arbeiten von JOHN HOLLAND aus den 60er Jahren des zwanzigsten Jahrhunderts. 1975 hat HOLLAND seine Arbeiten in [67] veröffentlicht.

Genetische Algorithmen versuchen den Lauf der Evolution nachzuahmen. Gültige Lösungen eines gegebenen Problems werden dabei in sogenannte Genotypen übersetzt. Ein

Genotyp ist bei HOLLAND eine binäre Zahlenfolge. Später wurde die Definition auch auf kontinuierliche Zahlenfolgen erweitert. Die entsprechende Darstellung des Genotyps im Lösungsraum des Problems wird Phänotyp genannt. Aus einer Startpopulation von Genotypen werden die Eltern für die nachfolgende Generation zufällig oder anhand der Güte der Lösungen ausgewählt. Mittels verschiedener genetischer Operatoren (Selektion, Mutation und Rekombination) werden neue Genotypen erzeugt. Die Zulässigkeit der erzeugten Lösungen ist zu prüfen. Die Gütewerte beziehungsweise die *Fitness* der neuen gültigen Lösungen werden ermittelt und für die Anpassung der Population verwendet.

In [43] werden Anwendungen von genetischen Algorithmen zum Entwurf regelbasierter Fuzzy-Systeme vorgestellt. Dabei ist wiederum der Pittsburgh- und der Michigan-Ansatz (siehe Abschnitt 3.1) zu unterscheiden. Beim Pittsburgh-Ansatz repräsentiert jedes Element der Population eine vollständige Regelbasis [5], beim Michigan-Ansatz eine Einzelregel [66]. Bei letzterem entspricht die gesamte Population einer Regelbasis. In [5] beinhaltet der Genotyp die Parameter der einzelnen Regeln. Die einzelnen Regeln können somit vom genetischen Algorithmus verändert werden. In [66] werden Einzelregeln optimiert und sequentiell einer Regelbasis hinzugefügt.

Der in dieser Arbeit vorgeschlagene Ansatz basiert auf den im Pruning generalisierten Einzelregeln. Damit ist der Suchraum der potentiellen Regelbasen bereits erheblich eingeschränkt, da die einzelnen Regeln nicht mehr verändert werden. Die Elemente der Population entsprechen vollständigen Regelbasen, denen beliebige Kombinationen der Einzelregeln angehören. Eine ähnliche Definition des Genotyps ist in [44] für das Ersetzen einer Regel einer bestehenden Regelbasis durch eine oder mehrere andere Kandidaten beschrieben. Dabei enthält der Genotyp die Information, welche Regeln die ausgewählte ersetzen sollen.

Für die Regelbasissuche wird in dieser Arbeit ein sogenannter Turnier-Algorithmus [42] verwendet. Dabei werden in jedem Iterationsschritt aus der aktuellen Population zufällig zwei Genotypen ausgewählt und ihre Gütewerte verglichen. Die Wahrscheinlichkeit der Auswahl ist für alle Elemente der Population gleich. Der 'Sieger' dieses Turniers bleibt erhalten, der 'Verlierer' erfährt eine Mutation und eine Rekombination mit dem 'Sieger'. Die Anzahl der Genotypen in der Population N_P bleibt dadurch stets konstant. Die Gütewerte sind die Entscheidungskosten oder die Gesamtkosten entsprechend dem gewählten Inferenzverfahren und der Bewertung als vollständige oder unvollständige Regelbasis.

Aufbau der Genotypen

Die Genotypen der Population sind Zeilenvektoren \mathbf{g}_p $p \in [1 \dots N_P]$. N_P ist die Größe der Population. Die Länge der Genotypen ergibt sich aus der Anzahl der zur Verfügung stehenden Regeln nach dem Pruning $N_{R,P}$.

Dann bedeutet:

$$g_{p,r} = 1, \quad p \in [1 \dots N_P], \quad r \in [1 \dots N_{R,P}],$$

dass in der zu Genotyp p gehörenden Regelbasis die Regel R_r enthalten sein soll. Entsprechend bedeutet

$$g_{p,r} = 0, \quad p \in [1 \dots N_P], \quad r \in [1 \dots N_{R,P}],$$

das Gegenteil: die Regel R_r ist nicht in der zu Genotyp p gehörenden Regelbasis enthalten. An der Stelle $g_{p,N_{R,P}+1}$ wird die Entscheidung der SONST-Regel der Regelbasis ange-

hängt. Bild 3.14 zeigt den Zusammenhang zwischen Genotyp und Phänotyp grafisch. Werden die Regelbasen als unvollständige Regelbasen bewertet, wird an die Stelle $g_{p,N_{R,P}+1}$ die Rückweisungsentscheidung eingetragen.

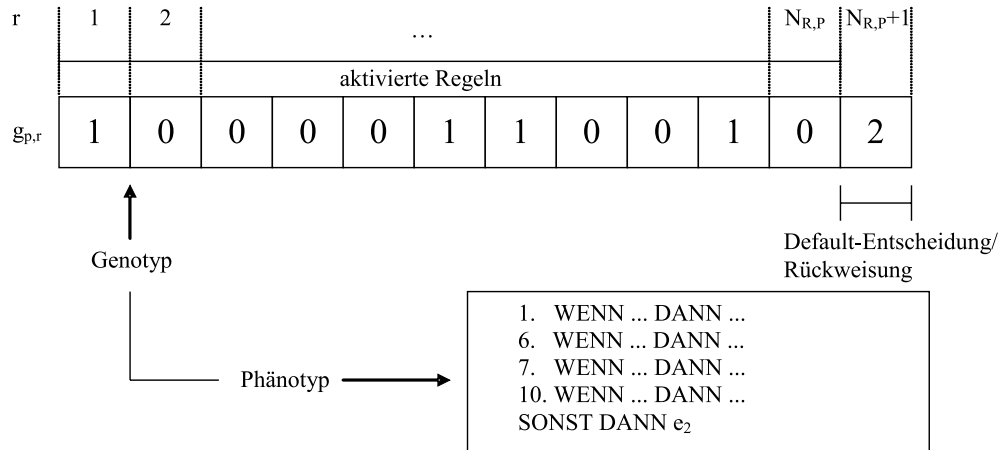


Bild 3.14: Darstellung eines Genotypen und dem daraus abgeleiteten Phänotyp.

Es können Randbedingungen vorgegeben werden. Zum Beispiel ist die maximale Anzahl aktivierter Regeln begrenzt. Die Auswertung der Regelbasis kann sowohl über die Fuzzy-Inferenz als auch die Evidenz-Inferenz erfolgen.

Erzeugen der Startpopulation

Zur Erzeugung der Startpopulation stehen zwei Möglichkeiten zur Verfügung. Die Genotypen können entweder über einen Zufallsgenerator erzeugt werden oder durch wiederholte Regelsuche mittels des in Abschnitt 3.3.1 vorgestellten Verfahrens. Dabei wird die jeweils zuerst ausgewählte Regel der vorherigen Suche aus der Liste der Regeln gelöscht. Dadurch wird gewährleistet, dass die Startpopulation nicht ausschließlich aus identischen Genotypen besteht. Die zufällig gewählten Genotypen sind darauf zu überprüfen, ob die zulässige Anzahl Regeln überschritten wurde. Gegebenenfalls werden dann Regeln zufällig wieder aus dem Genotyp gelöscht. Genotypen mit weniger Regeln sind zulässig.

Bewertung der Hypothesen

Für die Bewertung der Hypothesen stehen alle Optionen zur Bewertung von Regelbasen zur Verfügung. Bei der festen Default-Entscheidung legt der Benutzer die Default-Entscheidung fest. Bei der automatischen Default-Entscheidung bestimmt diese der genetische Algorithmus zufällig. Für die Optionen Rückweisung und Evidenz-Inferenz ist die Default-Entscheidung Rückweisung während der Suche vorgegeben. Nach deren Abschluss geht die Rückweisung entweder in die beste Entscheidung für die nicht erfassten Beispiele über oder bleibt bestehen (Evidenz-Inferenz).

Mutation

Bei der Mutation wird vom Algorithmus zufällig eine Stelle des Genotyps ausgewählt und diese von 0 auf 1 beziehungsweise von 1 auf 0 geändert. Wird die letzte Stelle im Genotyp ausgewählt, so wird bei automatischer Default-Entscheidung zufällig eine neue Default-Entscheidung aus der Liste der zulässigen Entscheidungen ausgewählt. Bei den anderen Optionen ist die Default-Entscheidung vorgegeben und nicht veränderbar. Der mutierte Genotyp muss wiederum eine zulässige Lösung sein. Dies bedeutet insbesondere, dass die zulässige Anzahl aktivierter Regeln durch die Mutation nicht überschritten werden darf.

Rekombination

Nach der Mutation des 'Verlierers' wird anschließend eine Rekombination der beiden Turnierteilnehmer durchgeführt. Hierzu wird zufällig eine Anzahl Gene bestimmt, die vom 'Sieger' auf den 'Verlierer' kopiert werden sollen. Ferner wird die Startposition, ab der die Gene eingefügt werden, zufällig bestimmt. Auf die Gültigkeit der Lösung ist wiederum zu achten. Die Wahl der Genanzahl und der Position wird so lange wiederholt, bis der rekombinierte Genotyp zulässig ist.

Ergebnisse

Für den Beispieldatensatz liefert der genetische Algorithmus für alle Optionen außer der Evidenz-Inferenz ähnliche Regelbasen vergleichbar zu denen in Bild 3.11. Die erwarteten Kosten sind dabei stets $\hat{L}_{D, RB} = 0.22$. Erst bei höher dimensionalen Problemen mit deutlich mehr Einzelregeln resultieren Regelbasen mit niedrigeren Kosten als bei der iterativen Regelsuche.

Für die Option Evidenz-Inferenz findet der genetische Algorithmus die Regelbasis aus Bild 3.15 links. Im selben Bild rechts sind die getroffenen Entscheidungen und die Bewertungen zusammengestellt. In der Regelbasis sind mit sechs Regeln deutlich mehr Regeln enthalten als bei der iterativen Suche. Die Regelbasis erhält im Vergleich zur iterativ gefundenen eine deutlich bessere Bewertung. Sowohl die Kosten $\hat{L}_{D, PRB}: 0.14$ zu 0.22 der entschiedenen Beispiele als auch die Anzahl der nicht entschiedenen Beispiele: $23 + 71 = 94$ zu $31 + 181 = 212$ fallen deutlich niedriger aus. Hier stellt sich die bessere Abdeckung des Suchraumes durch den genetischen Algorithmus als Vorteil heraus. Allerdings ist der Berechnungsaufwand größer.

3.3.3 Verbesserungsverfahren - Optimierung der Zugehörigkeitsfunktionen

In diesem Abschnitt wird untersucht, ob die nachträgliche Anpassung der Zugehörigkeitsfunktionen an die gefundene Regelbasis zu einer Verbesserung der Ergebnisse führt.

Bei der Initialisierung der Zugehörigkeitsfunktionen wird die Kostenmatrix nicht berücksichtigt. Die Zugehörigkeitsfunktionen werden über die Häufigkeiten der Merkmalswerte bestimmt.

Zunächst werden die Auswirkungen der Optimierung der Parameter $a_{l,i}$ der Zugehörigkeitsfunktionen an einem einfachen Beispiel gezeigt. Anschließend wird das Vorgehen bei der Optimierung vorgestellt.

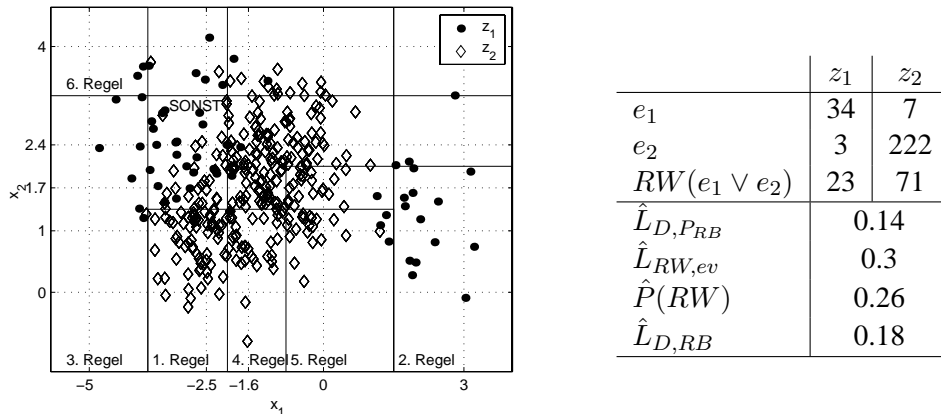


Bild 3.15: Links: Ergebnis der Regelbasissuche des genetischen Algorithmus mit Optimierung Evidenz-Inferenz. Rechts: Zahlenwerte, Optionen: $L_R = 10$, keine Startlösung, Populationsgröße 25.

Sei x das einzige Merkmal zu den Beispielen der beiden Umweltzustände z_1 und z_2 . Es handelt sich um zufällig über Normalverteilungen erzeugte Daten für die Umweltzustände z_1 und z_2 . Im Bereich $1.5 \leq x \leq 2.5$ befinden sich Beispiele beider Umweltzustände. Der Algorithmus initialisiert die Parameter a_i der fünf Zugehörigkeitsfunktionen A_i zu:

$$a_i = \{0; 1; 2; 3; 4\}$$

Bild 3.16 zeigt die Daten für die beiden Umweltzustände z_1 und z_2 und die Zugehörigkeitsfunktionen. Die gestrichelten senkrechten Linien markieren die Stellen x_k , an denen der Term mit der größten Aktivierung bei diesen Zugehörigkeitsfunktionen wechselt:

$$x_k = \{0.5; 1.5; 2.5; 3.5\}$$

An welcher Stelle der Wechsel des maximal aktivierten Terms zur Änderung der Entscheidung führt, hängt vom vorliegenden Kostenverhältnis L_R und der damit entworfenen Regelbasis ab. Für ein Kostenverhältnis von $L_R = 2$ kippt die Entscheidung zum Beispiel an der Stelle $x = 2.5$ (Fehlentscheidungen für z_1 sind teurer). Die vom Algorithmus gefundene Regelbasis beinhaltet dann die beiden Regeln:

r	WENN	DANN
1	$x = A_1$ ODER A_2 ODER A_3	$C_1 = e_1$
2	$x = A_4$ ODER A_5	$C_2 = e_2$
	SONST	$C_3 = e_1$

Für $L_R = 0.5$ kippt die Entscheidung an der Stelle $x = 1.5$ (Fehlentscheidungen für z_2 sind teurer) und die von Algorithmus gefundene Regelbasis lautet dann:

r	WENN	DANN
1	$x = A_1$ ODER A_2	$C_1 = e_1$
2	$x = A_3$ ODER A_4 ODER A_5	$C_2 = e_2$
	SONST	$C_3 = e_1$

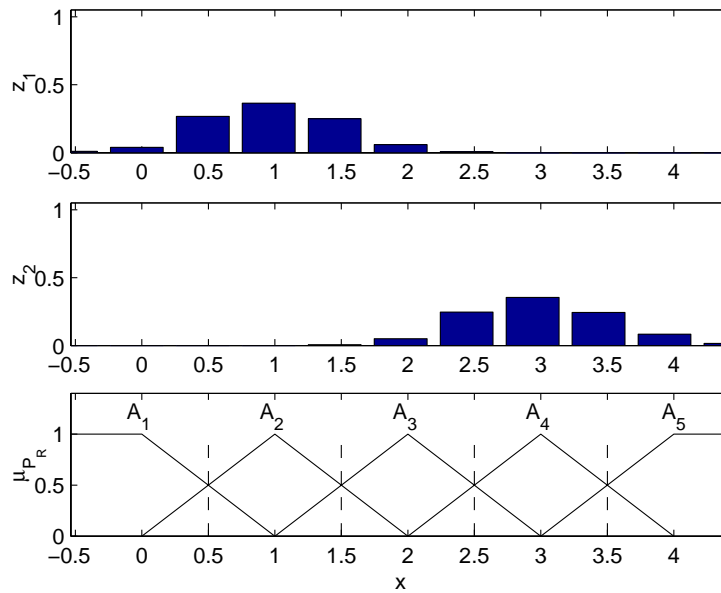


Bild 3.16: Beispielproblem für die Optimierung der Zugehörigkeitsfunktionen (500 Beispiele je Umweltzustand).

Eine Variation der Parameter der Zugehörigkeitsfunktionen bei Beibehaltung der Regelbasen zeigt jedoch, dass die vom Algorithmus gewählten initialen Parameter nicht immer optimal sind. Für den Parameter a_3 ist die Variation in Bild 3.17 für die beiden Kostenverhältnisse eingezeichnet. Dabei ist auf der Abszisse der Wert für a_3 und auf der Ordinate die Entscheidungskosten L_D eingezeichnet. Im linken Teil für $L_R = 2$ führt eine Verschiebung von a_3 auf den Wert $a_3 = 1$ zu einer deutlichen Kostenreduktion. Die Entscheidung kippt dadurch an der Stelle $x = 2$. Somit wird für deutlich weniger Beispiele von z_2 die Entscheidung e_1 getroffen. Entsprechend führt für $L_R = 0.5$ im rechten Teil von Bild 3.17 eine Verschiebung von a_3 auf den Wert $a_3 = 2.9$ zu einer Kostenreduktion. Hierdurch kippt die Entscheidung bei $x = 1.95$. In Tabelle 3.6 sind die Entscheidungen der Regelbasen für $L_R = 2$ und $L_R = 0.5$ sowie ihre Bewertung $\hat{L}_{D, RB}$ vor und nach der Variation zusammengestellt.

Um die niedrigsten Kosten für eine gegebene Regelbasis zu erreichen, muss der Wert für x , an dem die Entscheidung von e_1 nach e_2 wechselt, frei wählbar sein. Da bei der Bestimmung der Parameter $a_{l,i}$ der Zugehörigkeitsfunktionen die Kosten nicht berücksichtigt werden und auch die entworfene Regelbasis noch nicht bekannt ist, ist die kostenoptimale Regelbasis nicht unbedingt realisierbar. Es kann somit sinnvoll sein, die Zugehörigkeitsfunktionen an die entworfene Regelbasis anzupassen. Weiterhin kann die Optimierung sinnvoll sein, um die entworfene Regelbasis an geänderte Kostenmatrizen anzupassen.

Eine direkte Berechnung der optimalen Parameter ist wegen der Diskretisierung der Verteilungsdichtefunktionen der Merkmale in Abhängigkeit von den Umweltzuständen nicht möglich. Es wird daher ein heuristisches Verfahren zur Optimierung angewandt. Die Parameter der Zugehörigkeitsfunktionen werden nacheinander solange mit einem Gradienten-

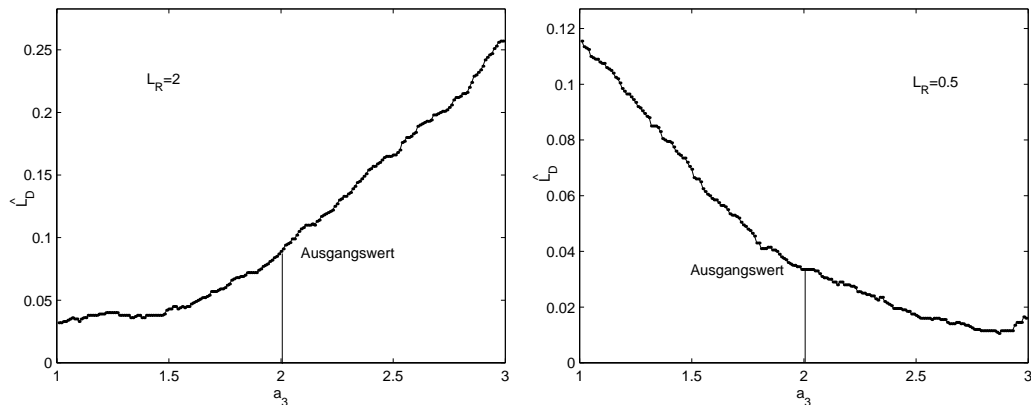


Bild 3.17: Variation von Parameter a_3 . Links: Kostenverhältnis $L_R = 2$; Rechts: Kostenverhältnis $L_R = 0.5$.

	$L_R = 2$		$L_R = 0.5$	
Initialisierung	$a_3 = 2$		$a_3 = 2$	
	z_1	z_2	z_1	z_2
e_1	499	87	435	1
e_2	1	413	65	499
$\hat{L}_{D, RB}$	0.09		0.03	
Variation	$a_3 = 1.0$		$a_3 = 2.9$	
	z_1	z_2	z_1	z_2
e_1	490	12	489	6
e_2	10	488	11	494
$\hat{L}_{D, RB}$	0.03		0.01	

Tabelle 3.6: Ergebnisse Optimierung der Zugehörigkeitsfunktionen (siehe Bild 3.17).

abstiegsverfahren verändert, bis sich der Güterwert der Regelbasis nicht mehr ändert. Als Güterwerte stehen die Entscheidungskosten, die Gesamtkosten aber auch die Klassifikationsgüte zur Auswahl.

Zur Reduktion der Rechenzeit werden zunächst die Terme bestimmt, deren Änderung einen Einfluss auf die getroffenen Entscheidungen haben können. Die Änderung der Zugehörigkeitsfunktion eines Merkmals, das für die Entscheidungen nicht verwendet wird, hat keinen Einfluss auf die getroffene Entscheidung. Die Terme von nicht verwendeten Merkmalen werden vollständig von der Optimierung ausgenommen.

Die Terme der Merkmale, die in der Regelbasis verwendet werden, sind ebenfalls nicht alle bei der Optimierung zu berücksichtigen. Bild 3.18 zeigt eine Prämisse mit drei benachbarten ODER-verknüpften Termen eines Merkmals x_l . Die breite schwarze Linie zeigt die Aktivierung $\mu(P_r)$ dieser Prämisse. Ein Einfluss auf die Entscheidung erfolgt nur, wenn sich $\mu(P_r)$ ändert. Bild 3.18 zeigt sowohl die Verschiebung des Parameters $a_{l,1}$ des Terms $A_{l,1}$ als auch des Parameters $a_{l,3}$ des Terms $A_{l,3}$. Die Verschiebung von $a_{l,3}$ führt nicht zu einer Änderung der Aktivierung. Die Verschiebung von $a_{l,1}$ dagegen schon. Für

		$a_{l,1}$	$a_{l,2}$	$a_{l,3}$	$a_{l,4}$	$a_{l,5}$		z_1	z_2	$L_{D,RB}$
vor	x_1	-5	-2.5	-1.6	0	3	e_1	56	38	0.22
	x_2	0	1	1.7	2.4	4	e_2	4	262	
nach	x_1	-5.5	-1.91	-1.65	0	2.67	e_1	58	37	0.16
	x_2	0	1.36	2.04	2.05	6.16	e_2	2	263	

Tabelle 3.7: Werte der Parameter $a_{l,i}$ vor und nach der Optimierung der Zugehörigkeitsfunktionen für die Regelbasis Bild 3.11 rechts sowie Entscheidungen und Kosten.

die dargestellte Prämisse wirkt sich eine Verschiebung der Parameter $\{a_1, a_2, a_4, a_5\}$ aus.

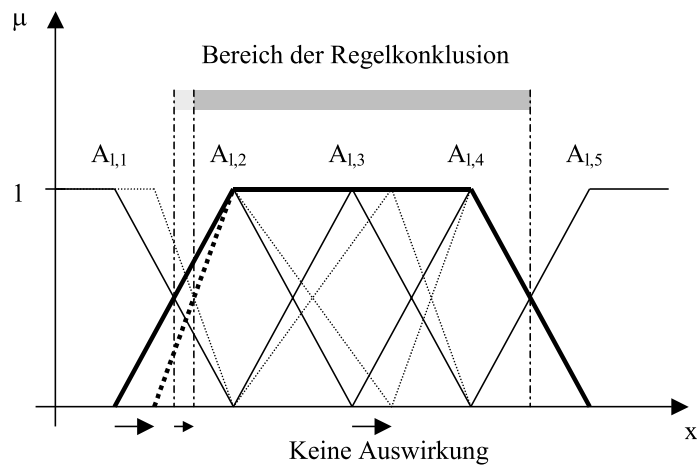


Bild 3.18: Zu optimierende Terme ($A_{l,1}, A_{l,2}, A_{l,4}, A_{l,5}$) am Beispiel der Prämisse: WENN $x_l = (A_{l,2} \vee A_{l,3} \vee A_{l,4})$. ($i_{min} = 2, i_{max} = 4$)

Allgemein sind die folgenden Terme einer Teilprämisse zu optimieren:

$$\begin{aligned}
 & A_{l,i_{min}-1} \quad \text{für } i_{min} > 1 \\
 & A_{l,i_{min}} \\
 & A_{l,i_{max}} \\
 & A_{l,i_{max}+1} \quad \text{für } i_{max} < m_l.
 \end{aligned} \tag{3.20}$$

Für den Beispieldatensatz zeigt Bild 3.19 einen Ausschnitt aus dem Gütegebirge der Regelbasis aus Bild 3.11. Dabei werden lediglich die Parameter $a_{1,2}$ und $a_{2,3}$ der beiden Merkmale verändert. Auf der Ordinate sind die Kosten pro Entscheidung der Regelbasis $\hat{L}_{D,RB}$ aufgetragen. Ein Gradientenabstiegsverfahren mit ausreichend großen Suchschritten ist in der Lage, das globale Minimum zu finden. Tabelle 3.7 zeigt die Ausgangswerte der Parameter $a_{l,i}$ und die Werte nach der Optimierung.

Abschließend werden in Tabelle 3.8 die Optionen des Algorithmus zusammengestellt. Bei der Generierung der Bäume ist es vorteilhaft, die Merkmalskosten zu vernachlässigen.

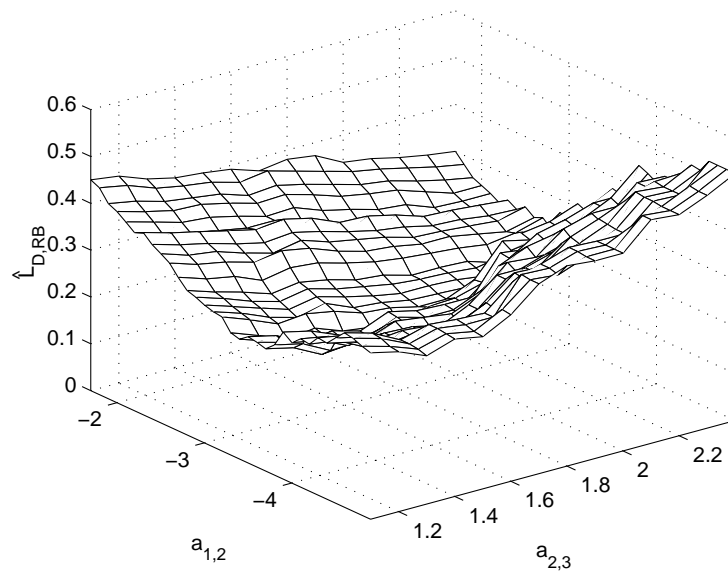


Bild 3.19: Ausschnitt aus dem Gütegebirge für die Parameter $a_{1,2}$ und $a_{2,3}$ und der Regelbasis aus Bild 3.11 rechts.

Die Prämissen im Baum werden stets über einen α -Schnitt festgelegt. Beim Pruning der Einzelregel besteht die Wahl zwischen α -Schnitt oder der Aktivierung $\mu(P_r)$. Zusätzlich besteht die Option, Merkmalskosten zu berücksichtigen. Die wesentlichen Optionen bei Regelbasen sind die Art der Inferenz und die Wahl bezüglich der Festlegung der Default-Entscheidung. Die Aktivierung der Prämissen wird stets über $\mu(P_r)$ bestimmt. Als Verbesserungsverfahren stehen die Optionen Genetischer Algorithmus beziehungsweise Optimierung der Zugehörigkeitsfunktionen zur Verfügung. Sinnvolle Standardeinstellungen für den Entwurf eines Entscheidungssystems mit Merkmalskosten sind in Tabelle 3.8 fett gedruckt. Die Empfehlungen basieren auf den Erkenntnissen aus den vorgestellten Beispielen und den in Kapitel 4 untersuchten Anwendungen.

Damit ist die entscheidungstheoretische Erweiterung bestehender Verfahren zum Entwurf von einstufigen Regelbasen mit den in Kapitel 2 entwickelten Bewertungsmaßen und die Ergänzung um neuartige Verbesserungsverfahren abgeschlossen. Im folgenden Abschnitt werden mehrstufige Entscheidungssysteme betrachtet.

3.4 Hypothesengenerierung - mehrstufige Regelbasen

3.4.1 Überblick

Das Pruning von Einzelregeln ist stets ein Kompromiss aus Generalisierung und der Güte der Entscheidungen der Regel. Weiterhin steht nur ein Bruchteil der möglichen Regeln zur Integration in die Regelbasis bereit. Daher ist nicht immer gewährleistet, dass eine ideal passende Regel zur Ergänzung der Regelbasis vorhanden ist. Häufig werden somit suboptimale Regeln in die Regelbasis aufgenommen. In der Regel entstehen Bereiche der

Einzelregel (ER)	Eröffnungsverfahren	Verbesserungsverfahren
Aktivierung μ Merkmalskosten	Baum $\mu \in \{0, 1\}$ (α -Schnitt) optional (nicht sinnvoll)	Pruning $\mu(\mathbf{P}_r)$ oder (α -Schnitt) optional (Dominanzkriterium)
Regelbasis (RB)	Regelbasis-Suche	GA/ ZGF-Optimierung
Aktivierung μ Inferenz Merkmalskosten Default- Entscheidung	$\mu(P_r)$ Fuzzy / Evidenz (Ev) optional fest / auto / RW	$\mu(P_r)$ Fuzzy / Evidenz optional GA: fest / Zufall / RW ZGF: fest aus RB / RW bei Ev

Tabelle 3.8: Optionen des Algorithmus für die Erstellung von Fuzzy-Systemen. GA: Genetischer Algorithmus, RW: Rückweisungsentscheidung, ZGF: Optimierung der Zugehörigkeitsfunktionen, Ev: Evidenz.

Regelbasis, in denen Widerspruch von zwei oder mehreren Regeln vorliegt. In diesen Bereichen kann sich die Güte der Entscheidungen durch die Überlappung verbessern oder verschlechtern. Nun ist die Situation möglich, in der eine zusätzliche Regel für einen Teil der noch nicht abgedeckten Beispiele die richtige Entscheidung trifft, jedoch im Bereich der Prämissen der bereits ausgewählten Regeln zu zusätzlichen Fehlern führt. Die Regel wird aufgenommen, wenn die Bewertung der Regelbasis insgesamt besser wird, und keine bessere alternative Regel zur Auswahl steht. In dieser Situation ist es vorteilhaft, die zusätzliche Regel nur für die Beispiele auszuwerten, die noch nicht von einer bereits ausgewählten Prämisse erfasst werden. Bild 3.20 veranschaulicht dies an einem schematischen Beispiel.

Es wird angenommen, dass nach dem Pruning die Regeln 1 bis 3 zur Verfügung stehen. Die Regeln sind durch unterschiedliche Rahmen in Bild 3.20 markiert. Es liege ein Kostenverhältnis $L_R > 1$ vor, so dass Fehlentscheidungen für Beispiele von z_1 teurer sind als für Beispiele von z_2 . Regel 1 erfasst fehlerfrei alle Beispiele von z_2 . Regel 3 erfasst lediglich einen Teil von z_1 . Regel 2 erfasst alle Beispiele von z_1 , aber auch Beispiele von z_2 . Somit führt die Aufnahme der Regeln 1 und 2 in die Regelbasis zu einem Bereich, in dem sich Regeln mit unterschiedlicher Konklusion überlappen. Es werden dann Fehlentscheidungen für Beispiele von z_2 getroffen. Eine Regelbasis aus den Regeln 1 und 3 trifft für einen Teil der Beispiele von z_1 keine Entscheidung. Die SONST-Regel trifft dann die restlichen Entscheidungen.

Eine Regelbasis mit mehreren Ebenen ist in der Lage, dieses Problem zu lösen. Dazu wird zunächst ausschließlich die Regel 1 ausgewertet (Bild 3.21 links). Anschließend führt Regel 2 für alle noch nicht entschiedenen Beispiele zur richtigen Entscheidung (Bild 3.21 links).

Eine mehrstufige Regelbasis kann zusätzlich zur Reduktion der variablen Merkmalskosten beitragen. Es sind stets nur die Merkmale, die zur Auswertung der Regelbasis der aktuellen Ebene notwendig sind, zu erfassen. Dies reduziert die Kosten gegenüber der Erfassung aller verwendeten Merkmale für alle Beispiele, da einige Merkmale nur für einen Teil der Beispiele auszuwerten sind.

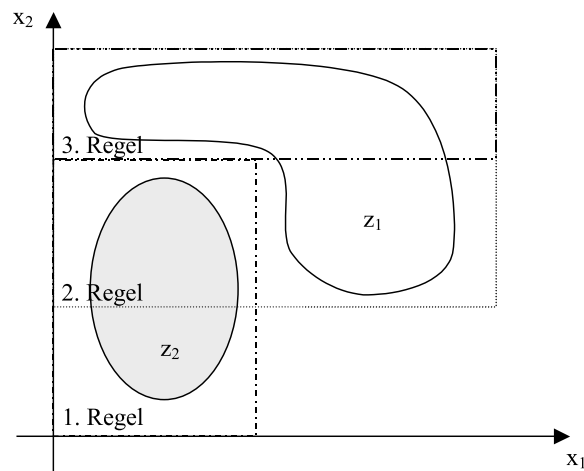


Bild 3.20: Schematisches Beispiel, fehlerfreie Entscheidung mit gegebenen Regeln nur über SONST-Regel möglich.

Diverse Autoren haben sich mit hierarchischen Strukturen für Entscheidungssysteme befasst. Eine hierarchische Klassenstruktur wird in [118] zur Modellierung von nicht kompakten Umweltzuständen vorgeschlagen. Eine hierarchische Aufteilung des Merkmalsraumes mit unterschiedlich feiner Aufteilung der Zugehörigkeitsfunktionen wird in [3] vorgestellt. Zur Verringerung der Kosten für Messungen wird in [2] ein sequentielles Durchführen von Messungen vorgeschlagen. Die nachfolgende Messung erfolgt in Abhängigkeit der Ergebnisse der vorherigen. In [63] werden Merkmale ebenfalls nacheinander ausgewertet. Das Entscheidungssystem kann jedoch nach der Auswertung jedes Merkmals eine Entscheidung treffen oder einen weiteren Test durchführen. Ebenfalls sequentiell durchlaufen werden die Regelbasen in [155]. Die Ausgangsgrößen der Regelbasen sind unscharf. Jedes Beispiel durchläuft alle Ebenen. Ein vorzeitiger Abbruch der Auswertung ist nicht vorgesehen. Bei hierarchischen Fuzzy-Reglern legt ebenfalls die letzte Ebene den Reglerausgang fest [89, 156].

Im Folgenden werden zwei neue Strategien für den Entwurf von regelbasierten Entscheidungssystemen vorgestellt, die mehrstufig hierarchisch aufgebaut sind. Ein hierarchisches regelbasiertes Entscheidungssystem im Sinne dieser Arbeit hat mindestens zwei Entscheidungsebenen, die sequentiell durchlaufen werden. Auf jeder Ebene kann eine Entscheidung getroffen werden, auf der letzten Ebene muss eine Entscheidung getroffen werden. Die Auswertung des Entscheidungssystems für ein Beispiel wird abgebrochen, sobald eine Entscheidung getroffen wurde. Die erste Ebene wird somit in jedem Fall ausgewertet, alle weiteren Ebenen nur, wenn die übergeordnete keine Entscheidung getroffen hat. Bild 3.22 zeigt diesen Ablauf exemplarisch für ein Entscheidungssystem mit drei Ebenen. Die Regelbasen der Ebenen setzen sich aus den Regeln zusammen, die nach dem Pruning zur Verfügung stehen. Die SONST-Regeln auf allen Ebenen bis auf die letzte haben die Konklusion Rückweisung (*RW*). Andernfalls werden alle Beispiele bereits auf der ersten Ebene

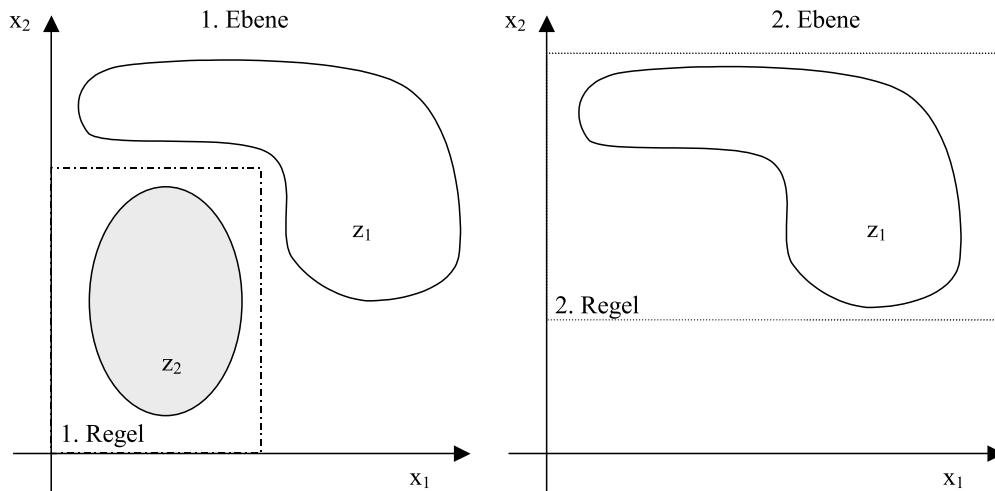


Bild 3.21: Schematisches Beispiel für sequentielle Entscheidung, fehlerfreie Entscheidung möglich.

entschieden. Jede Regel kann nur maximal einmal verwendet werden. Die vorgeschlagenen Strategien basieren auf dem in Abschnitt 3.3.1 vorgestellten Verfahren zur Suche von Regelbasen.

Bewertung hierarchischer Regelbasen

Die Bewertung einer hierarchischen Regelbasis (*HRB*) erfolgt analog zur Bewertung von einstufigen Regelbasen (2.105) beziehungsweise (2.108). Neben den Merkmalskosten lassen sich die Kosten einer hierarchischen Regelbasis wie die Kosten einer einstufigen Regelbasis in Entscheidungskosten $\hat{L}_{D,HRB}$ und die Kosten der auf der letzten Ebene nicht entschiedenen Beispiele \hat{L}_{RW} aufteilen.

Merkmalskosten bei hierarchischen Regelbasen

Auf der ersten Ebene ($rh = 1$) einer hierarchischen Regelbasis *HRB* sind alle verwendeten Merkmale zu berechnen:

$$L_{C,l,rh=1} = L_{C,l} \tag{3.21}$$

Für die weiteren Ebenen entfallen die Kosten der auf einer übergeordneten Ebenen bereits ausgewerteten Merkmale:

$$L_{C,l,rh+1} = 0 \quad \text{für } l \in \bigcup_{k=1}^{rh} l_{RB_k} \tag{3.22}$$

Dabei sind l_{RB_k} die Indices der Merkmale x_l , die auf der Ebene k ausgewertet werden. Für die erwarteten Merkmalskosten pro Entscheidung bezogen auf alle Ebenen ist der variable

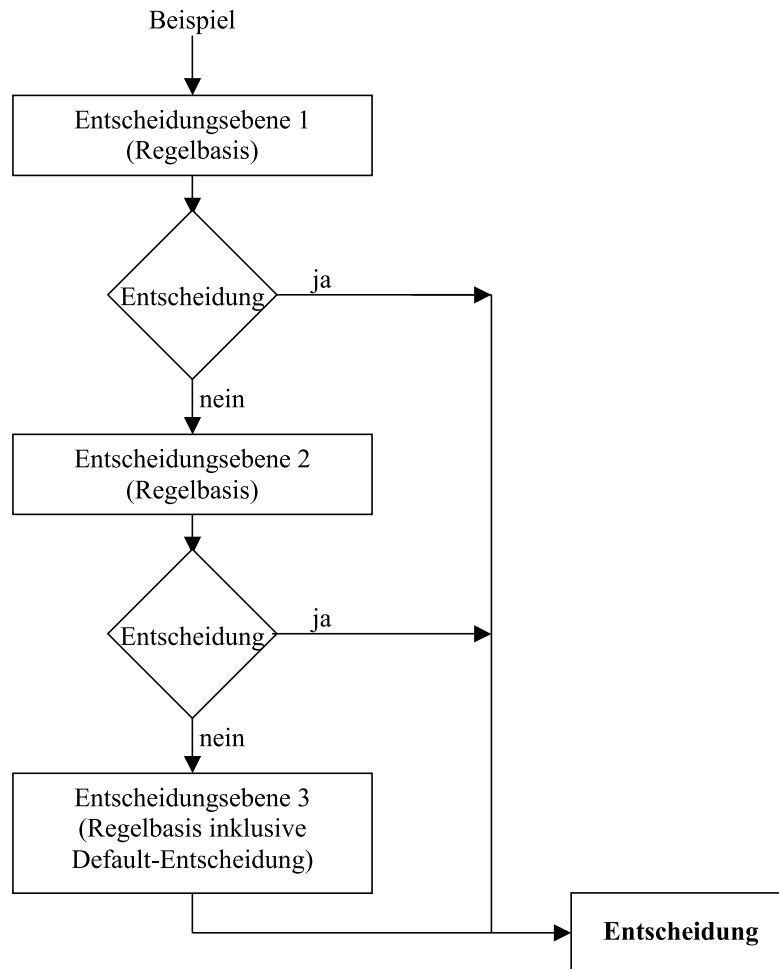


Bild 3.22: Ablauf einer dreistufigen Entscheidung.

Teil $L_{C,l,var}$ der Merkmalskosten nur anteilig zu berücksichtigen. Ist $N_{RW,rh}$ die Anzahl der Beispiele aus dem Lerndatensatz, die auf der Ebene rh nicht entschieden werden, dann ist:

$$\hat{P}_{RW,rh} = \frac{N_{RW,rh}}{N} \quad (3.23)$$

eine Schätzung für die Wahrscheinlichkeit, dass ein Beispiel bis einschließlich der Ebene rh noch nicht entschieden wird. Dies entspricht dem Anteil an Beispielen, für die die variablen Kosten eines Merkmals, das auf der nächsten Ebene $rh+1$ erstmals benötigt wird, zu berechnen sind. Damit folgt für die variablen Merkmalskosten auf der Ebene $rh+1$:

$$\hat{L}_{C,l,var,rh+1} = \begin{cases} 0 & \text{für } l \in \bigcup_{k=1}^{rh} l_{RB_k} \\ \hat{P}_{RW,rh} \cdot L_{C,l,var} & \text{sonst.} \end{cases} \quad (3.24)$$

Der fixe Teil $L_{C,l,fix}$ ist unabhängig von der Ebene bei der ersten Verwendung voll zu berechnen:

$$\hat{L}_{C,l,fix,rh+1} = \begin{cases} 0 & \text{für } l \in \bigcup_{k=1}^{rh} l_{RB_k} \\ L_{C,l,fix} & \text{sonst.} \end{cases} \quad (3.25)$$

Damit ergibt sich als Schätzung für die Merkmalskosten bis zur Ebene rh :

$$\hat{L}_{C,rh} = \sum_{l \in \bigcup_{k=1}^{rh} l_{RB_k}} \left[L_{C,l,fix} + \sum_{j=1}^{rh} \hat{L}_{C,l,var,j} \right]. \quad (3.26)$$

3.4.2 Permutationsstrategie

Ziel dieser Strategie ist es, durch Permutation der Positionen von Einzelregeln eine möglichst gute hierarchische Regelbasis zu erzeugen. Dazu wird die Regelbasis der jeweils letzten Ebene der aktuell besten hierarchischen Regelbasis gelöscht und durch eine neu ausgewählte Regelbasis ersetzt. Zunächst wird ausschließlich für die Beispiele der letzten Ebene eine neue einstufige Regelbasis mit $N_{R,soll}$ Regeln gesucht. Die erste Alternative zur aktuellen hierarchischen Regelbasis entsteht durch Einfügen der ausgewählten Regeln als Regelbasis auf der letzten Ebene. Weitere Hypothesen ergeben sich durch Hinzufügen einer weiteren Ebene und alle möglichen Aufteilungen der ausgewählten Regeln auf die aktuelle und die zusätzliche Ebene. In Bild 3.23 sind die aktuelle hierarchische Regelbasis und die drei möglichen Hypothesen für $N_{R,soll} = 2$ dargestellt.

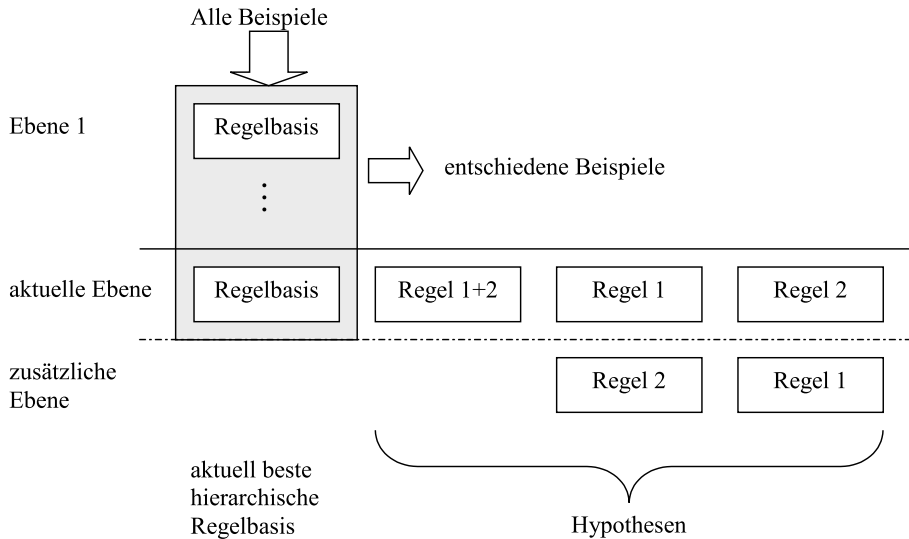


Bild 3.23: Hypothesenbildung bei der Permutationsstrategie.

Allgemein ergeben sich:

$$N_{hyp} = 2^{N_{R,soll}} - 1 \quad (3.27)$$

Hypothesen, da jede Regel entweder auf der aktuellen oder auf der zusätzlichen Ebene eingefügt werden kann. Das Einfügen aller Regeln auf der zusätzlichen Ebene wird ausgeschlossen, da sonst eine Ebene ohne Regeln entsteht.

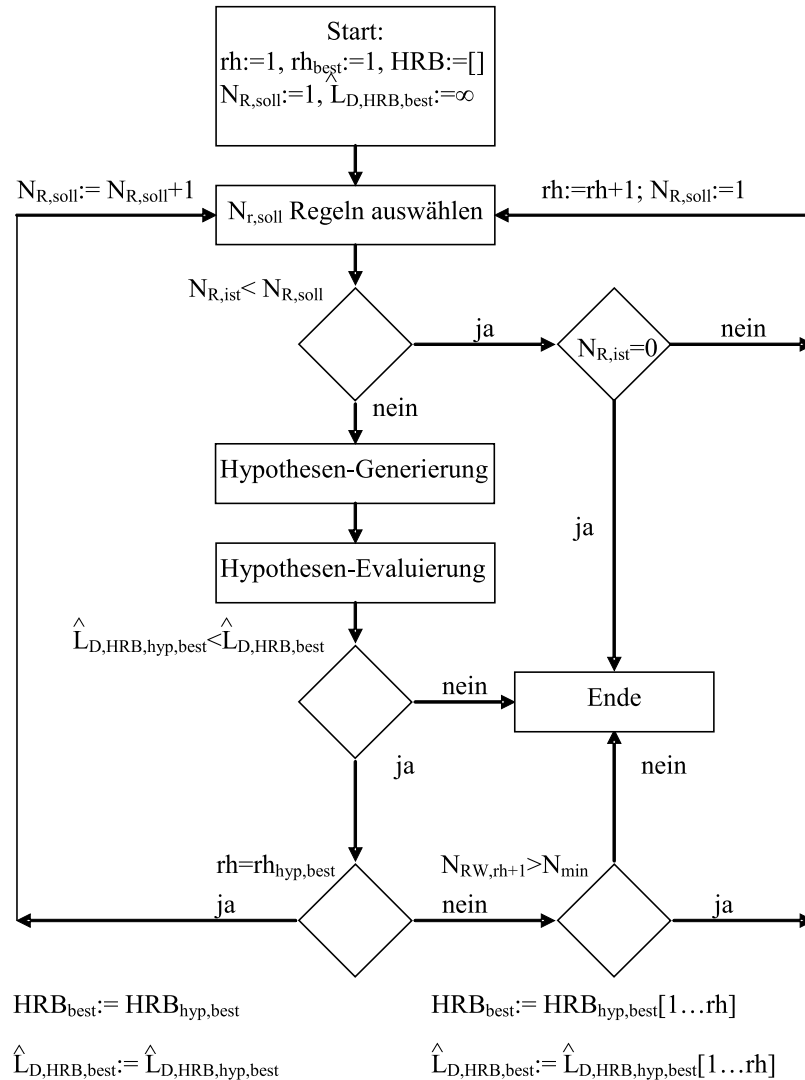


Bild 3.24: Permutationsstrategie: Algorithmus.

Bild 3.24 zeigt den Algorithmus der Permutationsstrategie. Zu Beginn ist $N_{R,soll} = 1$. Die Regelauswahl erfolgt als Regelbasissuche für die Beispiele der Ebene rh , die von den übergeordneten Ebenen noch nicht entschieden wurden. Es stehen die Regeln zur Verfügung, die das Pruning durchlaufen haben. Findet die Regelauswahl $N_{R,ist} = N_{R,soll}$ Regeln, werden die Hypothesen für die neue hierarchische Regelbasis generiert und anschließend evaluiert. Erreicht eine Hypothese niedrigere Kosten als die aktuell beste, wird die

aktuell beste Hypothese in Abhängigkeit von der Anzahl Ebenen der neuen besten Hypothese aktualisiert. Verfügt die neue beste hierarchische Regelbasis über die selbe Anzahl Ebenen wie ihre Vorgängerin, sind alle $N_{R,soll}$ Regeln auf der aktuellen Ebene eingefügt worden. In diesem Fall ersetzt die neue beste Hypothese die alte vollständig. Für die weitere Suche wird die Anzahl zu suchender Regeln $N_{R,soll}$ auf der selben Ebene rh um Eins erhöht. Dadurch erhöht sich ebenfalls die Anzahl der Hypothesen. Der Algorithmus wechselt dennoch die Ebene, wenn die erhöhte Anzahl Regeln in der nächsten Runde nicht gefunden wird ($N_{R,ist} < N_{R,soll}$ und $N_{R,ist} \neq 0$).

Verfügt die beste Hypothese hingegen über eine zusätzliche Ebene $rh + 1$ ($rh \neq rh_{hyp,best}$), wird als neue beste hierarchische Regelbasis die beste Hypothese bis zur Ebene rh festgelegt. Die Kosten werden ebenfalls nur bis zur Ebene rh berücksichtigt. Damit sind die Regeln auf der Ebene rh fixiert. Im nächsten Durchlauf wird die Ebene $rh + 1$ gegebenenfalls wiederum durch andere Regeln ersetzt. Die weitere Suche konzentriert sich auf die Beispiele, die auf der Ebene rh nicht entschieden werden.

Für die Suche auf der nächsten Ebene $rh + 1$ wird $N_{R,soll}$ auf den Wert Eins zurückgesetzt, wenn die Mindestanzahl Beispiele für die Ebene $rh + 1$ erreicht ist. Wird die Mindestanzahl Beispiele unterschritten, bricht der Algorithmus ab. Ebenso erfolgt ein Abbruch, wenn keine Hypothese besser ist als die aktuell beste hierarchische Regelbasis. Ein zusätzliches Abbruchkriterium beendet den Algorithmus, wenn die Regelauswahl für die noch nicht entschiedenen Beispiele keine passende Regel liefert. Die Konklusion der SONST-Regel auf der letzten Ebene der hierarchischen Regelbasis wird nach Abschluss der Suche festgelegt. Bei der Anwendung der Evidenz-Inferenz bleibt die Konklusion der SONST-Regel die Rückweisungsentscheidung.

3.4.3 Alpha-Strategie

Die Alpha-Strategie basiert auf der Variation des Risiko-Faktors α_{RB} (Abschnitt 2.5), (2.108). Je kleiner α_{RB} gewählt wird, desto weniger Fehlentscheidungen verursachen die ausgewählten Regeln. Allerdings steigt auch der Anteil Beispiele, für die keine Entscheidung getroffen wird. Diesen Effekt nutzt die Alpha-Strategie. Der Faktor α_{RB} wird für jede neue Ebene erhöht. Beginnend mit Ebene 1 und allen Beispielen wird für den aktuellen Wert von α_{RB} eine Regelbasis mit Rückweisungsentscheidung als SONST-Regel gesucht. Für die noch nicht entschiedenen Beispiele wird auf der nächsten Ebene mit einem erhöhten α_{RB} die Suche nach Regeln fortgesetzt. Die Erhöhung von α_{RB} wird so lange fortgesetzt, bis keine Regeln mehr gefunden werden, oder die Anzahl noch nicht entschiedener Beispiele auf der letzten Ebene einen Grenzwert unterschreitet. Die Erhöhung von α_{RB} erfolgt gemäß:

$$\alpha_{RB}[k + 1] = \frac{\alpha_{RB}[k] + 1}{2}. \quad (3.28)$$

Dadurch diese Formel nähert sich α_{RB} asymptotisch dem Wert 1. Bild 3.25 zeigt den Ablauf für $\alpha_{RB}[1] = 0.5$. Die dritte Ebene entscheidet die restlichen Beispiele. Der Entwurf einer hierarchischen Regelbasis mit der Alpha-Strategie verursacht deutlich weniger Rechenaufwand als der Entwurf mit der Permutationsstrategie.

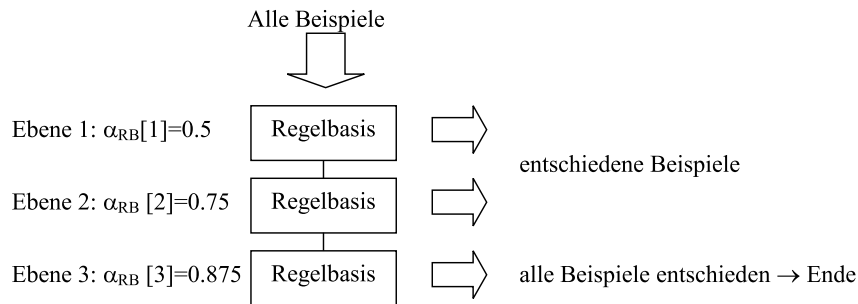


Bild 3.25: Alpha-Strategie mit $\alpha_{RB}[1] = 0.5$. Auf der dritte Ebene wird in diesem Beispiel der Rest entschieden.

Ergebnis für den Beispieldatensatz

Beide Strategien liefern für den Beispieldatensatz bei Berücksichtigung von Merkmalskosten identische Ergebnisse. Als variable Merkmalskosten werden $L_{C,1,var} = L_{C,2,var} = 0.3$ angesetzt. Fixe Kosten entstehen für beide Merkmale nicht. Das Kostenverhältnis ist wiederum $L_R = 10$. Bild 3.26 zeigt die beiden Ebenen der resultierenden hierarchischen Regelbasis. Die erste Ebene ist identisch zur einstufigen Regelbasis bei Berücksichtigung von Merkmalskosten aus Bild 3.12 links. Im rechten Teil von Bild 3.26 sind alle bereits auf der ersten Ebene entschiedenen Beispiele als Rechtecke eingezeichnet. Die beiden Regeln auf der zweiten Ebene treffen für alle noch nicht entschiedenen Beispiele eine Entscheidung. Beide Regeln erfordern neben der Auswertung des Merkmals x_1 auch die Auswertung von x_2 . Die variablen Kosten pro Entscheidung von x_2 fallen nur für die Beispiele an, die auf der ersten Ebene nicht entschieden werden. Die getroffenen und zurückgewiesenen Entscheidungen, Merkmalskosten auf den beiden Ebenen und die Gesamtkosten pro Entscheidung sind in Tabelle 3.9 zusammengefasst. Das Merkmal x_2 wird nur für ein Drittel der Beispiele ausgewertet und verursacht somit bezogen auf alle Beispiele nur Kosten in Höhe von 0.1. Gegenüber der einstufigen Regelbasis unter Berücksichtigung von Merkmalskosten ergibt sich eine deutliche Reduktion der Entscheidungskosten von 0.51 auf 0.34 pro Entscheidung. Gleichzeitig steigen die Merkmalskosten wegen des zusätzlichen verwendeten Merkmals x_2 an. Der Anstieg ist allerdings kleiner als die Reduktion der Entscheidungskosten. Somit ergibt sich insgesamt ein Kostenvorteil der hierarchischen Regelbasis gegenüber der einstufigen. Der Entwurf von mehrstufigen Regelbasen erfordert mehr Rechenzeit als der Entwurf von einstufigen Regelbasen.

3.5 Implementierung

Die entwickelten Bewertungsmaße und Verfahren werden für die Evaluierung in das am Institut für Angewandte Informatik in der Programmierumgebung Matlab entwickelte Programmpaket KAFKA [101] eingebunden. Bild 3.27 zeigt die Benutzerschnittstelle von KAFKA am Beispiel der Optionen für die Suche von Einzelregeln.

Das Programmpaket beinhaltet bereits die Strukturen für den Entwurf von Fuzzy-

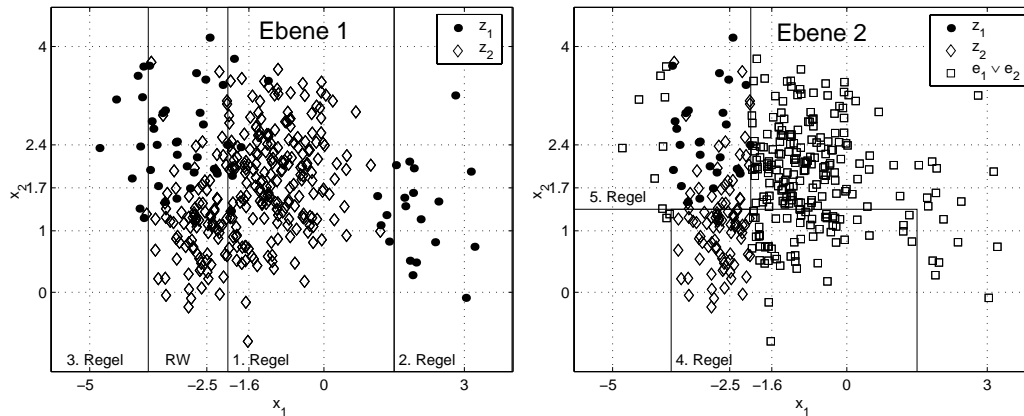


Bild 3.26: Hierarchische Regelbasis. Links: Regeln der ersten Ebene. Rechts: Regeln der zweiten Ebene. Rechtecke: auf der ersten Ebene entschiedene Beispiele.

		z_1	z_2
1. Ebene	e_1	25	9
	e_2	1	205
	$RW(e_1 \vee e_2)$	26	94
	$\hat{P}(RW, 1)$	0.66	
	$\hat{L}_{C,rh=1}$	0.3	
2. Ebene	e_1	51	9
	e_2	34	266
	$RW(e_1 \vee e_2)$	0	0
	$\hat{P}(RW, 2)$	0	
	$\hat{L}_{C,2,var,rh=2}$	$0.33 \cdot 0.3 = 0.1$	
Gesamt	$\hat{L}_{C,rh=2}$	0.4	
	$\hat{L}_{D,HRB}$	0.34	
	$\hat{L}_{T,HRB}$	0.74	

Tabelle 3.9: Kosten der hierarchischen Regelbasis für den Beispieldatensatz.

$$L_{C,1,var} = L_{C,2,var} = 0.3, L_R = 10, \alpha_{RB} = 0.8.$$

Systemen gemäß der Definition aus Abschnitt 2.3.1. Darüber hinaus sind einige statistische Verfahren zur Klassifikation implementiert. Der Quellcode umfasst Dateien mit einer Gesamtgröße von ca. einem Megabyte. Nach dem Start des Programmes befinden sich Variablen mit einem Platzbedarf von ca. 14 kilobyte im Speicher. Der Speicherbedarf beim Entwurf von Entscheidungssystemen hängt von der Größe des Datensatzes ab. Die Werte der Merkmale des Beispieldatensatzes sind zum Beispiel in einem Feld von ca. 11 kilobyte Größe gespeichert. Insgesamt sind nach dem Entwurf eines Entscheidungssystems beim Beispieldatensatz ca. 230 kilobyte im Speicher abgelegt. Probleme mit mehr Daten benötigen mehr Speicherplatz. Für den Entwurf einer Regelbasis für den Beispieldatensatz benötigt ein Rechner mit einem Pentium IV Prozessor mit 2.0 Gigahertz etwa vier Sekunden.

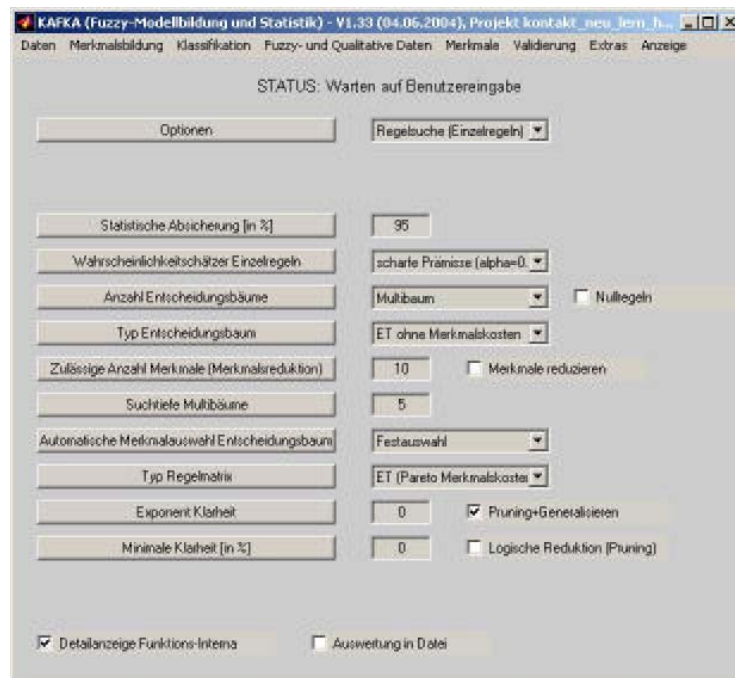


Bild 3.27: Benutzerschnittstelle von Kafka für die Optionen der Regelsuche.

Für den Entwurf einer Regelbasis beim Heart-Disease Datensatz (siehe Abschnitt 4.2.2) werden auf dem selben Rechner etwa 95 Sekunden benötigt. Dieser Datensatz umfasst 270 Beispiele und 13 Merkmale.

Das Programmpaket wird um alle in der vorliegenden Arbeit beschriebenen Bewertungsmaße und Verfahren erweitert. Im einzelnen sind das:

- die entscheidungstheoretische Bewertung bei Bäumen, Einzelregeln und Regelbasen,
- die Option Merkmalskosten in jedem Schritt zu berücksichtigen,
- das Dominanzkriterium beim Pruning mit Merkmalskosten,
- die Evidenz-Inferenz,
- die Optimierung der Zugehörigkeitsfunktionen für eine bestehende Regelbasis,
- die Regelbasissuche über einen genetischen Algorithmus und
- das Entwurfsverfahren für hierarchische Regelbasen.

Die Maßnahmen werden durch Erweiterungen bestehender Programmteile oder Hinzufügen neuer Programmteile realisiert. Die entscheidungstheoretischen Bewertungsmaße sind in einer Funktion zusammengefasst. Diese Funktion wird von den Entwurfsvorgängen für die verschiedenen Bewertungsmaße gerufen.

Für den Entwurf der mehrstufigen Regelbasen wird zum Beispiel eine Steuerungsfunktion ergänzt, die je nach Ebene den entsprechenden Teil des Datensatzes und den Wert für α_{RB} an die Entwurfsvorgänge für einstufige Regelbasen übergibt.

In Bild 3.28 ist das Menü 'Fuzzy- und Qualitative Daten' von KAFKA dargestellt. Über dieses Menü werden die Verfahren zum Entwurf von Entscheidungssystemen aufgerufen. In der ersten Zeile des Menü's ist die Automatik-Funktion zu sehen. Bei Aktivierung dieses Menüpunktes wird der gesamte Entwurfsprozess auf der Basis der eingestellten Optionen durchlaufen. Dies sind im einzelnen die folgenden Schritte:

1. Entwurf der Zugehörigkeitsfunktionen,
2. Entwicklung eines oder mehrerer Entscheidungsbäume,
3. Pruning der Regelkandidaten und
4. Suchen einer einstufigen Regelbasis.

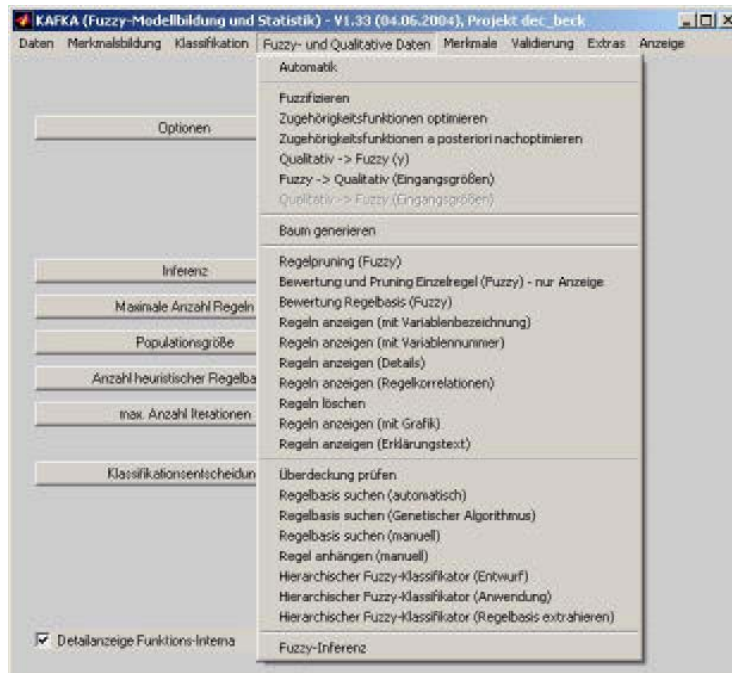


Bild 3.28: Menü 'Fuzzy- und Qualitative Daten' von KAFKA. Über dieses Menü werden die Verfahren zum Entwurf von Entscheidungssystemen aufgerufen.

Für den Vergleich mit statistischen Verfahren wird das Programm zusätzlich um ein Modul ergänzt, das eine entscheidungstheoretische Bewertung für statistische Verfahren zulässt.

Die wesentlichen Erkenntnisse aus Kapitel 3 werden im Folgenden nochmals kurz zusammengefasst. In Abschnitt 3.2 wird der Entwurf von Entscheidungsbäumen und das Pruning von Einzelregeln um die entscheidungstheoretische Bewertung erweitert. Dabei wird

für das Pruning ein neues Dominanzkriterium zur zweikriteriellen Bewertung (Entscheidungskosten und Merkmalskosten) der Einzelregeln vorgestellt. In Abschnitt 3.3 werden die Bewertungsmaße auf den Entwurf von einstufigen Regelbasen angepasst. Eine Suchstrategie für den Entwurf von Regelbasen bei Verwendung des auf der Evidenztheorie basierenden Inferenzverfahrens wird vorgestellt. Zusätzlich werden modifizierte Verbesserungsverfahren für Regelbasen vorgeschlagen (Optimierung der Zugehörigkeitsfunktionen und Regelbasissuche über einen genetischen Algorithmus). In Abschnitt 3.4 wird ein Konzept für mehrstufige Regelbasen und Strategien für den Entwurf solcher Regelbasen diskutiert. Mit der Implementierung der neuen Maße und Verfahren sowie der Erweiterung des Programmpaketes KAFKA steht ein leistungsfähiges Werkzeug zum Entwurf von Entscheidungssystemen auf der Basis unscharfer Regeln zur Verfügung.

4 Anwendungen

4.1 Überblick

In diesem Kapitel werden Anwendungen der diskutierten neuartigen Algorithmen zum Entwurf von Fuzzy-Entscheidungssystemen vorgestellt. Um einen Vergleich mit etablierten Verfahren ziehen zu können, werden Benchmarkdatensätze verwendet. Darüber hinaus wird eine praxisnahe Anwendung für die Kontakterkennung einer künstlichen Hand untersucht.

Die Anwendungen sind im einzelnen:

- **German-Credit:** Der Datensatz wurde im Rahmen des Statlog-Projektes [98] untersucht. Ziel ist es, anhand von personenbezogenen Daten die Kreditwürdigkeit der Person vorherzusagen (Abschnitt 4.2.1).
- **Heart-Disease:** Der Datensatz wurde ebenfalls im Rahmen des Statlog-Projektes untersucht. Hier ist auf der Basis von Anamnesedaten und weiteren Untersuchungsergebnissen das Vorliegen einer Herzkrankheit zu prognostizieren. Für den Datensatz liegen Informationen über die variablen Kosten der einzelnen Untersuchungen vor [145]. Damit können die Merkmalskosten in den Entwurfsprozess integriert werden (Abschnitt 4.2.2).
- **Kfz-Aggregate Benchmark:** Auf der Basis von akustischen Messdaten soll vorhergesagt werden, ob ein Kfz-Getriebe defekt ist oder nicht. Der Hersteller der Getriebe stellt als Anforderung an das Entscheidungssystem, 70% der fehlerhaften Getriebe zu erkennen und dabei eine Glaubwürdigkeit von 30% zu erreichen [140]. Kosten der Merkmale liegen nicht vor (Abschnitt 4.3).
- **Kontakterkennung bei einer künstlichen Hand:** Der Datensatz ist selbst erstellt und beinhaltet Führungs-, Mess- und Stellgrößen der Fingergelenkregelung einer künstlichen Hand. Aus den Daten soll erkannt werden, ob Kontakt zwischen einem einzelnen Finger der Hand und einem Objekt besteht (Abschnitt 4.4).

Die verwendeten Abkürzungen für die Optionen während des Entwurfes sind in Tabelle 4.1 zusammengestellt. Für die Risikoparameter gilt sofern nicht anders angegeben: $\alpha_{ER} = \alpha_{RB} = 1$.

Die Evaluierung der Datensätze German-Credit und Heart-Disease erfolgt über Kreuzvalidierung (*crossvalidation*). Dabei wird sichergestellt, dass die Validierungsdaten beim Entwurf des Entscheidungssystems nicht berücksichtigt werden. Bei der Kreuzvalidierung

Ev	Evidenz-Inferenz aktiviert
GA250	genetischer Algorithmus mit 250 Iterationen ohne Startlösung
MK	Merkmalskosten berücksichtigt
HRB	hierarchische Regelbasis
ZGF	Optimierung der Zugehörigkeitsfunktionen (a posteriori)

Tabelle 4.1: Zusammenstellung der verwendeten Abkürzungen.

wird der Datensatz zufällig in N_c -Teile mit identischer Anzahl Beispiele geteilt:

$$N_C = \frac{N}{N_c}.$$

Das Verhältnis der Umweltzustände zueinander ist in allen Teilen identisch. Für den Entwurf des Entscheidungssystems stehen $N_c - 1$ Teile des Datensatzes zur Verfügung. Die Validierung erfolgt mittels des letzten Teils des Datensatzes. Der Vorgang wird N_c -mal wiederholt. Jedes Beispiel des Datensatzes ist somit genau einmal an der Validierung und $N_c - 1$ -mal am Entwurf beteiligt. Bild 4.1 zeigt schematisch den Ablauf einer Kreuzvalidierung. Der Datensatz wird dabei in fünf Teile aufgeteilt.

Die Evaluierung der Anwendungen Kfz-Benchmark und Kontakterkennung erfolgt über unabhängige Validierungsdatensätze zur jeweiligen Problemstellung. Es steht mindestens ein Datensatz zum Entwurf des Entscheidungssystems und mindestens ein weiterer zur Evaluierung zur Verfügung. Damit ist ebenfalls sichergestellt, dass Beispiele, die zum Entwurf verwendet werden, bei der Validierung unberücksichtigt bleiben.

4.2 Statlog-Datensätze

4.2.1 German-Credit

Der 'German-Credit Data Set' wurde am Institut für Statistik und Ökonometrie der Universität Hamburg erstellt. Der Datensatz besteht aus 1000 Beispielen, die in zwei Umweltzustände eingeteilt sind:

1. 700 Beispiele der Klasse 'guter Kreditkunde'
2. 300 Beispiele der Klasse 'schlechter Kreditkunde'

Der Original-Datensatz besteht aus 20 qualitativen und quantitativen Merkmalen für die einzelnen Kreditkunden. Es existiert eine Variante des Datensatzes mit 24 Merkmalen. Die Merkmale 21 bis 24 sind zur Transformation der qualitativen in quantitative Merkmale angefügt. In der vorliegenden Arbeit wird zur Gewährleistung der Vergleichbarkeit wie in [98] der Datensatz mit 24 Merkmalen verwendet. Ebenso wird die Kostenmatrix aus [98] verwendet:

$$\mathbf{L} = \begin{pmatrix} 0 & 5 \\ 1 & 0 \end{pmatrix}.$$

Die erste Zeile entspricht der Entscheidung 'Kredit gewähren' (e_1), die zweite der Entscheidung 'Kredit ablehnen' (e_2). Entsprechend bedeutet Spalte 1 'guter Kreditkunde' und

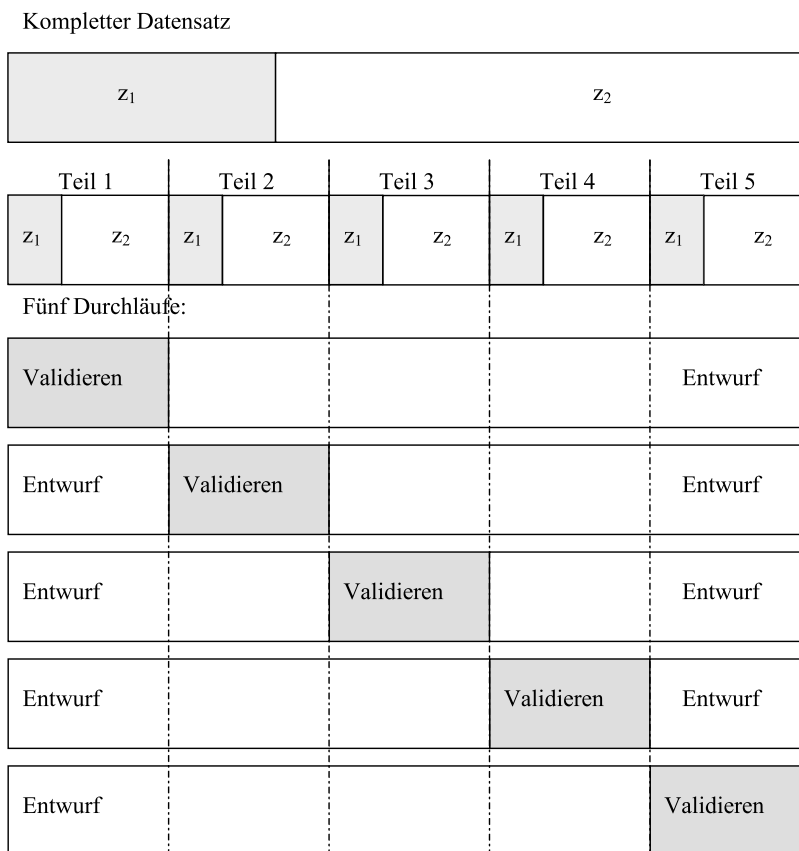


Bild 4.1: Schema der Kreuzvalidierung.

Spalte 2 'schlechter Kreditkunde'. Demnach verursacht das Gewähren eines Kredites an einen schlechten Kreditkunden fünfmal höhere Kosten – der Kredit wird ausfallen –, als das Ablehnen des Kreditantrages eines guten Kunden – dieser kann zum Beispiel zu einer anderen Bank gehen. Die Kostenmatrix für dieses Problem wurde in Abschnitt 2.2.1 eingehender diskutiert.

Die beste Default-Entscheidung ist e_2 . Für diese Entscheidung entstehen Kosten in Höhe von $L_{D,def} = 0.7$ pro Entscheidung. Merkmalskosten liegen für den Datensatz nicht vor und werden daher beim Entwurf der Entscheidungssysteme auch nicht berücksichtigt. Die Ergebnisse werden über eine 10-fache Kreuzvalidierung mit zehn Wiederholungen berechnet.

Tabelle 4.2 zeigt die Ergebnisse für die lineare Diskriminanzanalyse und einstufige Fuzzy-Regelbasen. Bei dem statistischen Verfahren werden die Entscheidungen ebenfalls auf der Basis einer entscheidungstheoretischen Bewertung der Wahrscheinlichkeiten der Zustände getroffen. Die Fuzzy-Regelbasen werden mit der Option Rückweisung gesucht. Das beste Verfahren ist die lineare Diskriminanzanalyse. Die besten Ergebnisse der Fuzzy-Entscheidungssysteme ergeben sich für die Regelbasissuche mittels des genetischen Al-

	Kreuzvalidierung				Lerndaten		Rang
	\hat{L}_D	σ	Fehler %	σ	\hat{L}_D	Fehler %	
Diskriminanzanalyse Ergebnis aus [98]	0.535				0.509		1
eigene Implementierung	0.543	0.009	37.5	0.4	0.521	32.2	(3)
Regelbasierte Verfahren [98]							
CN2	0.856				0.000		16
ITrule	0.879				*		18
Fuzzy α – Schnitt	0.593	0.026	37.2	0.8	0.433	31.3	(5)
Fuzzy $\mu(P_r)$	0.589	0.018	37.3	1.4	0.412	32.0	(5)
Fuzzy α – Schnitt GA	0.587	0.018	41.5	0.8	0.569	32.9	(5)
Fuzzy $\mu(P_r)$ GA	0.572	0.028	41.6	0.9	0.526	32.2	(3)

Tabelle 4.2: Ergebnisse German-Credit Datensatz. (): theoretische Platzierung. *: Wert steht nicht zur Verfügung.

gorithmus, wenn bei der Bewertung der Einzelregeln kein α -Schnitt zur Festlegung des Einzugsgebietes erfolgt. Es entstehen dann Kosten in Höhe von 0.572 pro Entscheidung.

In der Rangliste der Verfahren für diesen Datensatz [98] liegt das Fuzzy-Verfahren damit hinter der linearen ($\hat{L}_D = 0.535$) und der logistischen Diskriminanzanalyse ($\hat{L}_D = 0.538$) bei diesem Datensatz auf Platz drei. Damit erreicht das Fuzzy-Verfahren keine besseren Ergebnisse als die statistischen Verfahren. Als Vorteil gegenüber den statistischen Verfahren kann die Interpretierbarkeit der entworfenen Regelbasen angeführt werden. Das in dieser Arbeit vorgestellte regelbasierte Verfahren bringt jedoch eine deutliche Verbesserung im Vergleich zu den in [98] untersuchten regelbasierten Entwurfsverfahren. Diese belegen die Plätze 16 ($\hat{L}_D = 0.856$) und 18 ($\hat{L}_D = 0.879$) von 23 untersuchten Verfahren.

4.2.2 Heart-Disease

Der Heart-Disease Datensatz wurde von der Cleveland Clinic Foundation erstellt [98]. Er ist Teil des UCI Machine Learning Repository, einer Sammlung von Datensätzen, die in der Literatur [98] als Benchmarkdatensätze für Maschinelles Lernen verwendet werden. Der Datensatz wurde ausgewählt, da ebenfalls Ergebnisse für Entscheidungssysteme, die die unterschiedlichen Konsequenzen für Fehlentscheidungen berücksichtigen, vorliegen [98]. Zusätzlich liegen auch die variablen Kosten für die Merkmale vor [145].

Der Datensatz besteht aus 270 Beispielen mit den Werten für je 13 Merkmale. Zwei Umweltzustände werden unterschieden:

1. 150 Beispiele der Klasse 'Herzkrankheit liegt nicht vor'
2. 120 Beispiele der Klasse 'Herzkrankheit liegt vor'

Die Entscheidungen entsprechen den Umweltzuständen: 'Herzkrankheit liegt nicht vor' (e_1) oder 'Herzkrankheit liegt vor' (e_2). Folgende Kostenmatrix wird in [98] verwendet:

$$\mathbf{L} = \begin{pmatrix} 0 & 5 \\ 1 & 0 \end{pmatrix}.$$

Für ein Beispiel, bei dem die Herzkrankheit nicht vorliegt, die Entscheidung 'Herzkrankheit liegt vor' zu treffen, verursacht Kosten von 1. Im umgekehrten Fall verursacht die Diagnose 'Herzkrankheit liegt nicht vor' Kosten von 5, sofern die Krankheit tatsächlich vorliegt. Wichtiger als die absolute Höhe der angesetzten Kosten für die Fehlentscheidungen ist deren Verhältnis zueinander. Die Kostenmatrix bedeutet dann, dass eine falsche Diagnose bei Vorliegen der Krankheit fünfmal schlimmer ist als bei deren Abwesenheit. Bild 4.2 gibt anhand von zwei Merkmalen einen Eindruck von der schlechten Trennbarkeit der Umweltzustände im Datensatz.

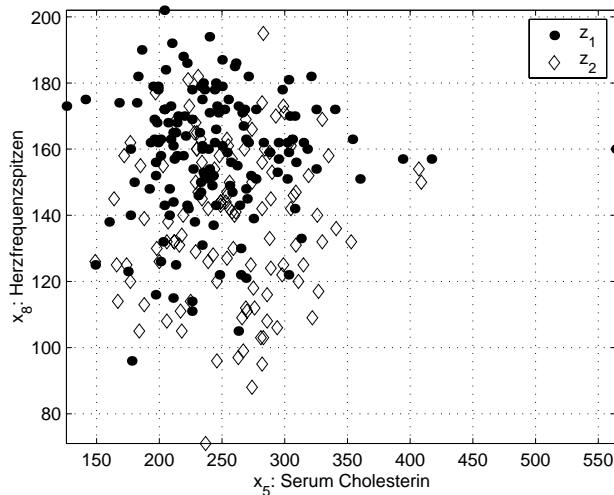


Bild 4.2: Heart-Disease Datensatz über den beiden Merkmalen mit der höchsten Relevanz.

Die Kosten der Merkmale liegen vor und betragen zwischen 1 und 102.9 Geldeinheiten [145] (siehe Anhang 6.2.3). Die Merkmalskosten sind alle variable Kosten, das heißt die Kosten fallen nur an, wenn das entsprechende Merkmal ausgewertet wird. Das Verhältnis zwischen Entscheidungskosten und Merkmalskosten ist damit für den Entwurf von Entscheidungssystemen ungeeignet. Teilweise übersteigen die Kosten der Merkmale die Entscheidungskosten erheblich. Für den Entwurf der Entscheidungssysteme müssen die Kosten der Merkmale niedriger sein, als die Kosten einer Default-Entscheidung. Bei der gegebenen Verteilung der Umweltzustände im Lerndatensatz von 150:120 ergibt sich e_2 als beste Default-Entscheidung. Die Default-Kosten betragen dann $L_{D,def} = 0.56$. Die Kosten der Default-Entscheidung sind somit niedriger als das günstigste Merkmal. Um das Verhältnis zu ändern, wird die Kostenmatrix mit dem selbst gewählten Faktor 1000 multipliziert:

$$\mathbf{L} = \begin{pmatrix} 0 & 5000 \\ 1000 & 0 \end{pmatrix}.$$

Jeder andere Wert ist ebenfalls möglich. Dadurch ändern sich die Entscheidungssysteme nicht, die ohne Berücksichtigung der Merkmalskosten entworfen werden (siehe Abschnitt 2.2.1). Durch die Änderung der Kostenmatrix wird der Entwurf mit Merkmalskosten möglich.

Zur Evaluierung der Entscheidungssysteme wird wie in [98] eine 9-fache Kreuzvalidierung mit zehn Wiederholungen verwendet. Die Entscheidungssysteme mit Fuzzy-Inferenz werden alle mit der Option Rückweisung entworfen. Neben der Fuzzy-Inferenz wird auch die Evidenz-Inferenz untersucht. Als Kosten des Experten wird $L_{RW,ev} = 400$ gewählt. Tabelle 4.3 zeigt die Ergebnisse für Entscheidungssysteme mit Fuzzy-Inferenz und Tabelle 4.4 zeigt die Ergebnisse mit Evidenz-Inferenz. In beiden Fällen ist $\alpha_{ER} = \alpha_{RB} = 1$.

Die Entscheidungen über Diskriminanzfunktionen zu treffen, liefert wie beim German-Credit Datensatz bessere Ergebnisse als die Fuzzy-Entscheidungssysteme. Zu bemerken ist, dass bei Berücksichtigung von Merkmalskosten die entworfenen Regelbasen weniger Regeln enthalten. Die Kosten über Validierungsdaten sinken. Allerdings nimmt die Anzahl der Fehlentscheidungen deutlich zu. Die Ursache dafür ist, dass die Entscheidungssysteme durch die Berücksichtigung der Merkmalskosten weniger speziell sind. Dadurch tritt zusätzlich der Effekt des 'Overfitting' [98, 120] weniger stark in Erscheinung. Overfitting bedeutet, dass aus den vorhandenen Daten zum Entwurf des Entscheidungssystems mehr strukturelle Annahmen extrahiert werden, als in den Daten vorhanden sind. Die Folge sind in der Regel viele Fehler bei den Validierungsdaten. Die Optimierung der Zugehörigkeitsfunktionen bringt nur für die Lerndaten eine Verbesserung. Eine deutliche Verbesserung erfolgt durch Variation der Parameter α_{ER} und α_{RB} (siehe unten).

Ebenfalls zu erwähnen sind die Effekte bei der Verwendung des genetischen Algorithmus zur Regelbasissuche. Wie beim German-Credit Datensatz sinken die Entscheidungskosten gegenüber der iterativen Regelbasissuche. Die größte Verbesserung ergibt sich ohne Berücksichtigung der Merkmalskosten bei Verwendung der Option $\mu(P_r)$ für die Aktivierung der Einzelregeln. Die Entscheidungskosten sinken dadurch von 488.9 auf 403. Werden die Merkmalskosten berücksichtigt, fallen diese deutlich höher aus als bei der iterativen Suche. Die Ursache dafür ist, dass eine rein zufällig erzeugte Startpopulation verwendet wird und dabei keine Kosten berücksichtigt werden. Eine deutliche Verbesserung der Entscheidungskosten findet bei Berücksichtigung von Merkmalskosten nicht statt.

Die hierarchischen Regelbasen werden mit der Alpha-Strategie entworfen ($HRB\alpha$). Die Permutationsstrategie führt bei deutlich höherem Rechenaufwand zu vergleichbaren Ergebnissen. Die Entscheidungskosten bei der Kreuzvalidierung liegen bei den hierarchischen Regelbasen auf einem vergleichbaren Niveau zu den einstufigen Regelbasen. Allerdings nimmt die Schwankung der Ergebnisse tendenziell zu. Die Merkmalskosten nehmen, wenn sie beim Entwurf berücksichtigt werden, im Vergleich zu einstufigen Regelbasen entgegen der Erwartung leicht zu. Der Grund hierfür sind die zusätzlich in der Regelbasis verwendeten Merkmale. Sie werden nicht für alle Beispiele ausgewertet und nicht für alle Beispiele verrechnet. Beim Entwurf einer hierarchischen Regelbasis über dem gesamten Datensatz werden bei der Option $\mu(P_r)$ im Vergleich zur einstufigen Regelbasis zusätzlich die Merkmale x_5 und x_7 verwendet. Für beide zusammen entstehen Kosten in Höhe von 20.7, wenn sie für alle Beispiele verwendet werden. Da sie nur für einen Teil der Beispiele ausgewertet werden, steigen die Merkmalskosten lediglich um 3.9 von 9.2 auf 13.1. Die Anzahl ausgewählter Regeln N_R und damit die Spezialisierung ist bei den hierarchischen Regelbasen höher als bei einstufigen Regelbasen.

Die Evidenz-Inferenz senkt sowohl die Kosten für Fehlentscheidungen als auch die Anzahl Fehler im Vergleich zur Fuzzy-Inferenz sowohl für Regelbasen mit einer Ebene als auch für hierarchische Regelbasen deutlich (Tabelle 4.4). Die Kosten sinken bei der

Verfahren	Kreuzvalidierung						Lerndaten					
	\hat{L}_D	σ	L_C	σ	Fehl. %	σ	\hat{L}_D	L_C	Fehl. %	$N_R/r/h$	N_x	Rang
Naive Bayes [98]	374						351					1
Diskriminanzanalyse												
Ergebnis aus [98]	393						315					2
eigene Implementierung	398.5	26.4	600.5	0	23.8	1.2	337	600.5	21.8			(4)
Regelbasierte Verfahren [98]												
CN2	707						206					21
ITrule	515						*					13
Fuzzy α -Schnitt	499.2	36.6	449.7	21.9	26.8	2.8	181.5	323.1	15.19	10	10	(12)
Fuzzy $\mu(P_r)$	488.9	51.0	440.3	29.7	26.4	1.9	259.3	323.1	23.0	7	10	(12)
Fuzzy MK α -Schnitt	476.7	44.1	10.5	1.6	33.1	1.7	325.9	16.5	29.6	6	5	(10)
Fuzzy MK $\mu(P_r)$	454.8	28.4	6.7	1.1	33.6	2.0	292.6	9.2	26.3	7	5	(9)
Fuzzy MK $\mu(P_r)$, $\alpha_{ER} = 0.8, \alpha_{RB} = 0.75$	398.5	20.5	5.3	0.7	32.1	1.4	314.8	4	28.5	4	4	(4)
Fuzzy GA250 α -Schnitt	451.5	42.0	440.0	23.6	28.4	2.0	288.9	425.0	25.9	10	10	(8)
Fuzzy GA250 $\mu(P_r)$	403.0	40.8	448.6	37.8	26.5	1.8	251.8	508.1	20.7	10	11	(4)
Fuzzy MK GA250 α -Schnitt	458.5	25.8	194.8	24.2	33.4	2.6	363.0	119.4	24.4	10	7	(9)
Fuzzy MK GA250 $\mu(P_r)$	445.5	42.5	213.5	21.5	31.7	3.0	414.8	132.9	26.7	10	8	(8)
Fuzzy ZGF α -Schnitt	519.3	63.1	458.7	24.3	25.4	2.3	159.3	323.1	12.9	10	10	(14)
Fuzzy ZGF $\mu(P_r)$	500.0	64.0	432.4	33.2	25.1	2.9	229.6	323.1	18.5	7	10	(12)
Fuzzy MK ZGF α -Schnitt	487.7	56.9	10.9	1.6	32.9	2.2	266.7	16.5	23.7	6	5	(12)
Fuzzy MK ZGF $\mu(P_r)$	471.1	46.3	7.1	1.8	32.6	2.1	288.9	9.2	35.9	7	5	(10)
Fuzzy HRB α α -Schnitt	510.4	59.2	434.8	20.1	23.9	2.4	177.8	310.3	14.8	10/5	10	(13)
Fuzzy HRB α $\mu(P_r)$	494.1	75.1	394.5	18.2	26.7	2.9	192.6	372.0	19.2	11/5	12	(12)
Fuzzy HRB α MK α -Schnitt	490.7	45.9	10.1	1.3	34.5	2.8	281.5	8.3	25.2	9/5	6	(12)
Fuzzy HRB α MK $\mu(P_r)$	447.0	17.6	10.0	2.1	33.0	1.9	281.5	13.1	25.2	7/5	7	(8)

Tabelle 4.3: Ergebnisse des Heart-Disease Datensatzes, Durchschnitt und Standardabweichung aus 10 Realisierungen einer 9-fachen Kreuzvalidierung (240 Datensätze Lernen / 30 Datensätze Testen). Fuzzy-Inferenz. (): theoretische Platzierung. *: Wert steht nicht zur Verfügung.

Verfahren	Kreuzvalidierung						Lerndaten					
	\hat{L}_D	σ	L_C	σ	Fehl. %	σ	\hat{P}_{RW}	σ	\hat{L}_D	L_C	Fehl. %	\hat{P}_{RW}
Fuzzy EV α -Schnitt	360.4	62.2	470.1	26.2	14.1	1.7	26.5	3.1	83.3	307.6	8.3	24.4
Fuzzy EV $\mu(P_r)$	352.6	49.8	440.5	19.8	14.0	1.8	27.4	1.8	72	426	7.2	27.8
Fuzzy EV MK α -Schnitt	355.5	38.9	9.6	1.2	20.2	1.0	34.6	2.1	178.6	16.5	12.8	27.4
Fuzzy EV MK $\mu(P_r)$	314.6	35.0	7.1	1.4	19.2	1.2	35.8	3.3	196.1	4	19.6	24.4
Fuzzy EV GA250 $\mu(P_r)$	313.4	64.9	455.4	30.0	15.5	1.4	35.2	3.0	140.7	485.3	10.0	26.3
Fuzzy EV MK GA250 $\mu(P_r)$	327.9	37.6	202.7	24.0	23.4	2.8	37.0	5.4	230.3	127.7	23.03	38.9
Fuzzy HRB α Ev α -Schnitt	421.9	71.8	334.9	20.8	16.9	2.5	20.3	3.1	41.8	387.6	4.2	11.5
Fuzzy HRB α Ev $\mu(P_r)$	387.6	52.7	329.9	24.8	16.0	1.7	20.7	3.1	87.4	341.2	8.7	23.7
Fuzzy HRB α Ev MK α -Schnitt	384.9	44.6	7.5	0.8	21.3	1.1	26.4	3.3	174.4	8.6	13.4	19.3
Fuzzy HRB α Ev MK $\mu(P_r)$	344.2	53.0	6.1	0.8	20.0	1.7	32.3	2.8	187.5	8.6	14.5	28.8

Tabelle 4.4: Ergebnisse des Heart-Disease Datensatzes, Durchschnitt und Standardabweichung aus 10 Realisierungen einer 9-fachen Kreuzvalidierung (240 Datensätze Lernen / 30 Datensätze Testen). Evidenz-Inferenz. \hat{L}_D entspricht den Kosten der entschiedenen Beispiele.

Berücksichtigung von Merkmalskosten nochmals. Allerdings steigt dann auch der Anteil nicht entschiedener Beispiele. Bei einstufigen Regelbasen werden ohne Merkmalskosten durchschnittlich 27.4 % der Beispiele der Validierungsdaten nicht entschieden. Mit Merkmalskosten steigt der Anteil nicht entschiedener Beispiele auf 35.8 % an, während die Kosten für Fehlentscheidungen von 352.6 auf 314.6 sinken.

Für den genetischen Algorithmus sind ähnliche Effekte zu beobachten wie bei der Fuzzy-Inferenz. Ohne Berücksichtigung von Merkmalskosten sinken die Entscheidungskosten. Die Merkmalskosten fallen im Vergleich zur iterativen Regelsuche deutlich höher aus, wenn sie im Entwurf berücksichtigt werden.

Bei den hierarchischen Regelbasen, die mit der Option Evidenz-Inferenz entworfen werden, nimmt die Zahl der nicht entschiedenen Beispiele bei der Kreuzvalidierung tendenziell ab, die Zahl der Fehler nimmt dafür zu. Mit den Optionen $\mu(P_r)$ ohne Merkmalskosten sinkt \hat{P}_{RW} von 27.4 % auf 20.7 %, dafür steigt die Fehlerrate von 14.0 % auf 16.0 %. Dementsprechend nehmen auch die Kosten der entschiedenen Beispiele L_D zu. Zu bemerken ist, dass die Merkmalskosten ohne Berücksichtigung beim Entwurf bei den hierarchischen Regelbasen gegenüber den einstufigen deutlich sinken, für die Option $\mu(P_r)$ von 440.5 auf 329.9. Der Grund hierfür ist wiederum, dass die Merkmale teilweise nicht für alle Beispiele ausgewertet werden.

Mit der Option $\mu(P_r)$ für die Aktivierung der Einzelregeln ergeben sich tendenziell bessere Ergebnisse als mit der Option α -Schnitt.

Variation von α_{ER} und α_{RB}

Für die Optionen mit den niedrigsten Kosten (einstufige Regelbasis mit Merkmalskosten und vollständiger Prämissenaktivierung) aus Tabelle 4.3 soll der Einfluss der Variation der Parameter α_{ER} und α_{RB} auf die Entscheidungskosten untersucht werden. Die beiden Parameter variieren jeweils im Intervall $[0.5 \dots 1]$. Bild 4.3 zeigt die Ergebnisse für die erwarteten Kosten (rechts) und die prozentualen Fehler (links).

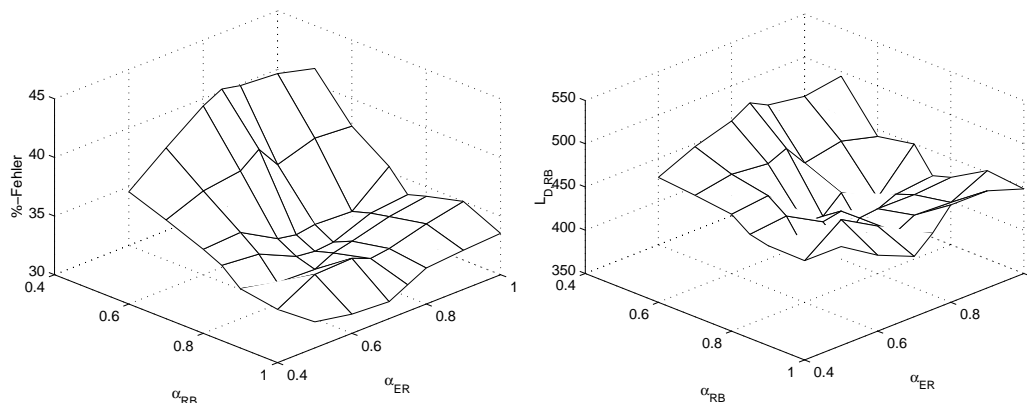


Bild 4.3: Variation von α_{ER} und α_{RB} . Links: Prozent Fehler. Rechts: erwartete Entscheidungskosten. Kreuzvalidierung bei Berücksichtigung von Merkmalskosten.

Folgende Zusammenhänge lassen sich aus den Bildern ableiten. Die Variation des

Parameters α_{ER} hat nur einen geringen Einfluss auf die Fehler und die Kosten. Dagegen sinken die Kosten und die Anzahl der Fehler mit der Zunahme des Parameters α_{RB} . Für $\alpha_{RB} > 0.8$ nehmen die Kosten wieder zu. Die niedrigsten erwarteten Kosten ergeben sich mit $L_D = 398$ für $\alpha_{ER} = 0.8$ und $\alpha_{RB} = 0.75$ (Fehlerrate 32.1%). Die Kosten liegen dann auf dem Niveau der Diskriminanzanalyse.

Es empfiehlt sich beim Entwurf eines Entscheidungssystems mit $\alpha_{ER} = \alpha_{RB} = 1$ zu beginnen. Sind die Ergebnisse nicht zufriedenstellend, ist eine Variation der Parameter α_{ER} und α_{RB} sinnvoll. Das Ergebnis kann sich dadurch verbessern.

Zusammenfassung

Für diesen Datensatz sind mit allen Verfahren des Maschinellen Lernens nur wenig befriedigende Ergebnisse zu erzielen, weil er offensichtlich zu wenig informationstragende Merkmale enthält [98]. Dennoch erreicht der entscheidungstheoretische Ansatz Ergebnisse, die auf dem Niveau der besten statistischen Verfahren liegen. Hinzu kommt die bessere Interpretierbarkeit des Fuzzy-Entscheidungssystems.

Je spezieller die Entscheidungssysteme beim Heart-Disease Datensatz entworfen werden, desto schlechter schneidet das Verfahren bei einer Kreuzvalidierung ab. Der Datensatz ist sehr heterogen. Bei einer starken Spezialisierung des Entscheidungssystems auf den Teil der Daten, der zum Lernen verwendet wird, steigt die Fehlerquote für die Validierungsdaten. Daher führt die Optimierung der Zugehörigkeitsfunktionen bei der Fuzzy-Inferenz zu schlechteren Ergebnissen. Das Verfahren führt tendenziell zu spezielleren Entscheidungssystemen. Der genetische Algorithmus bringt Vorteile, wenn die Merkmalskosten beim Entwurf nicht zu berücksichtigen sind. Die hierarchischen Regelbasen erreichen insgesamt ebenfalls keine besseren Ergebnisse. Im Zusammenhang mit der Evidenz-Inferenz wird allerdings deutlich, dass die hierarchischen Regelbasen weniger Merkmalskosten verursachen als einstufige. Die Eignung der Evidenz-Inferenz zur Identifikation von Beispielen, für die keine Entscheidung getroffen werden soll, kann für den Heart-Disease Datensatz ebenfalls gezeigt werden.

4.3 Kfz-Aggregate Benchmark

Bei diesem Datensatz handelt es sich um einen Benchmarkdatensatz der Universität Dortmund und des GMA-Fachausschusses 5.22 für die Evaluierung von automatischen Verfahren zur Modellbildung [65, 100, 117, 140]. Die Daten stammen von Prüfläufen von Kfz-Getrieben. Ziel ist die automatische Erkennung von defekten Aggregaten anhand von akustischen Signalen. Weitere Beispiele für Diagnoseaufgaben im Automotive-Bereich sind zum Beispiel in [73, 94, 112, 139] zu finden.

Der Benchmarkdatensatz besteht aus den drei Teilen A, B und C mit 2612, 4373 und 10613 Beispielen. Aus den akustischen Signalen des Prüflaufs eines Aggregates werden 304 Merkmale abgeleitet [65]. Der Anteil der defekten Aggregate in den einzelnen Teilen des Datensatzes beträgt jeweils ca. 1%. Die beiden Umweltzustände und entsprechend auch die beiden Entscheidungen sind 'kein Defekt (io)' (z_1/e_1) oder 'Defekt (nio)' (z_2/e_2).

Zur Beschreibung der Güte der Entscheidungssysteme verwendet der Hersteller der Getriebe die Erkennungsrate ER der defekten Aggregate sowie die Glaubwürdigkeit G [65,

100]. Die Erkennungsrate ist der Prozentsatz der vom Entscheidungssystem erkannten und tatsächlich defekten Beispiele:

$$ER = \hat{P}(e_2|z_2). \quad (4.1)$$

Die Glaubwürdigkeit entspricht dem Prozentsatz der Beispiele, die als defekt eingestuft werden und auch tatsächlich defekt sind:

$$G = \frac{\hat{P}(e_2 \wedge z_2)}{\hat{P}(e_2 \wedge z_1) + \hat{P}(e_2 \wedge z_2)} = \frac{\hat{P}(e_2 \wedge z_2)}{\hat{P}(e_2)}. \quad (4.2)$$

Tendenziell geht eine Erhöhung der Erkennungsrate mit einer sinkenden Glaubwürdigkeit einher. Der Hersteller der Aggregate erwartet eine Erkennungsrate ER_{soll} von mehr als 70% bei einer Glaubwürdigkeit G_{soll} von mehr als 30% [65].

Aus der geforderten Glaubwürdigkeit G_{soll} lässt sich ein Anhaltspunkt für das Kostenverhältnis L_R der beiden Fehlentscheidungen ableiten. Folgende Ungleichung ist zu erfüllen:

$$\frac{\hat{P}(e_2 \wedge z_2)}{\hat{P}(e_2 \wedge z_2) + \hat{P}(e_2 \wedge z_1)} > G_{soll}. \quad (4.3)$$

Dividieren von Zähler und Nenner der linken Seite durch die Wahrscheinlichkeit für den Umweltzustand z_2 liefert:

$$\frac{\hat{P}(e_2|z_2)}{\hat{P}(e_2|z_2) + \frac{\hat{P}(e_2 \wedge z_1)}{\hat{P}(z_2)}} = \frac{\hat{P}(e_2|z_2)}{\hat{P}(e_2|z_2) + \underbrace{\frac{\hat{P}(e_2 \wedge z_1)}{\hat{P}(z_1)} \cdot \frac{\hat{P}(z_1)}{\hat{P}(z_2)}}_{\hat{P}(e_2|z_1)}} > G_{soll}. \quad (4.4)$$

Umstellen und Verwendung von $\hat{P}(e_2|z_1) = 1 - \hat{P}(e_1|z_1)$ führt zu:

$$\hat{P}(e_1|z_1) > 1 - \frac{1 - G_{soll}}{G_{soll}} \cdot \frac{\hat{P}(z_2)}{\hat{P}(z_1)} \cdot \hat{P}(e_2|z_2) \quad (4.5)$$

Der Koeffizientenvergleich von (4.5) mit (2.35) zeigt, dass $\frac{1-G}{G}$ als Kostenverhältnis $\frac{L(e_1, z_2) - L(e_2, z_2)}{L(e_2, z_1) - L(e_1, z_1)}$ aufgefasst werden kann.

Bild 4.4 zeigt den Verlauf des äquivalenten Kostenverhältnisses bei der Variation von G .

Für $G_{soll} = 0.3$ ergibt sich ein Kostenverhältnis von $L_R = 7/3$. Aus der geforderten Erkennungsrate der defekten Aggregate lässt sich kein zusätzlicher Zusammenhang zum Kostenverhältnis der Fehlentscheidungen ableiten.

Zur Anonymisierung wurden die Merkmalswerte innerhalb jedes Teildatensatzes vom Hersteller der Aggregate auf den Bereich $[0 \dots 1]$ normiert. Die Normierung hängt wegen den heterogenen Fehlerbeispielen stark von den zufällig auftretenden Minimal- und Maximalwerten im Datensatz ab [65]. Daher sind die Werte der Merkmale in den Teilen des Datensatzes trotz identischer Wertebereiche nicht vergleichbar. Zur Herstellung der Vergleichbarkeit werden die Merkmale der drei Teile des Datensatzes auf einheitliche Mittelwerte $x^*_l = 0$ und Standardabweichung $\sigma^*_l = 1$ renormiert. Dies erfolgt gemäß [65]:

$$x_l^* = \frac{x_l - \bar{x}_l}{\sigma_l} \quad (4.6)$$

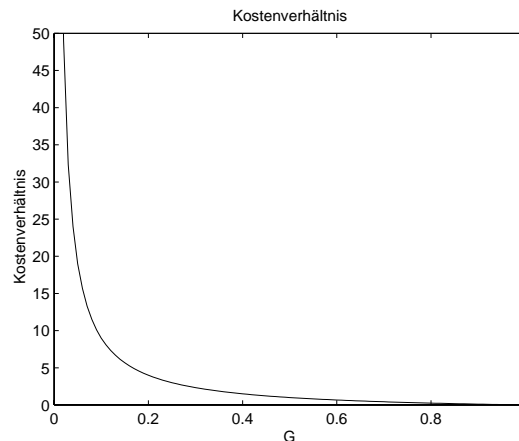


Bild 4.4: Kostenverhältnis in Abhängigkeit von G .

Zur Evaluierung der Algorithmen werden Fuzzy-Entscheidungssysteme mit unterschiedlichen Kostenverhältnissen L_R der folgenden Kostenmatrix entworfen:

$$\mathbf{L} = \begin{pmatrix} 0 & L_R \\ 1 & 0 \end{pmatrix} \quad (4.7)$$

Ein Teil des Datensatzes dient dem Entwurf (Teil C) und die beiden anderen (Teile A und B) zur Validierung. Der Vergleich mit einem herkömmlichen Verfahren ohne entscheidungstheoretische Bewertung wird ebenfalls gezogen.

Wegen der starken Asymmetrie der Verteilung der beiden Umweltzustände wird für die Suche der Regelbasis die Option feste Default-Entscheidung verwendet. Als Default-Entscheidung wird 'io'(e₁) eingestellt. Somit reduzieren sich die Kosten der Regelbasis durch Hinzufügen von geeigneten Regeln mit der Konklusion 'nio'(e₂).

In den Bildern 4.5 und 4.6 sind die Ergebnisse des Entwurfs von Fuzzy-Systemen bei verschiedenen Kostenverhältnissen dargestellt. Bild 4.5 zeigt die Ergebnisse für die Pruning-Option 'vollständige Prämisse ($\mu(P_r)$)'. Bild 4.6 zeigt die Ergebnisse für die Pruning-Option ' α -Schnitt'.

Für beide Optionen nimmt die Glaubwürdigkeit über den Lerndaten mit steigendem Kostenverhältnis tendenziell ab. Im Gegensatz dazu nimmt die Erkennungsrate der defekten Aggregate erwartungsgemäß mit steigendem Kostenverhältnis tendenziell zu. Die niedrigeren Erkennungsraten bei der Option 'vollständige Prämisse' für kleine Kostenverhältnisse resultieren aus den weniger stark generalisierten Einzelregeln. Der un stetige Verlauf der Kurven hängt mit der Suboptimalität des Verfahrens und der relativ kleinen Zahl an defekten Beispielen in den Datensätzen zusammen.

Die Kurven spiegeln sich für die Teile A und B der Daten, die zur Validierung dienen, nicht eindeutig wieder. Eine mögliche Ursache hierfür ist, stark unterschiedliche Verteilung verschiedener Fehlersubgruppen in den drei Teilen des Datensatzes [100]. In Tabelle 4.5 sind die Ergebnisse für einige Kostenverhältnisse zusammengefasst.

Ein Ergebnis, das den Anforderungen über allen Datensätzen (Lernen mit Teil C) relativ nahe kommt, stellt sich für $L_R = 20$ ein.

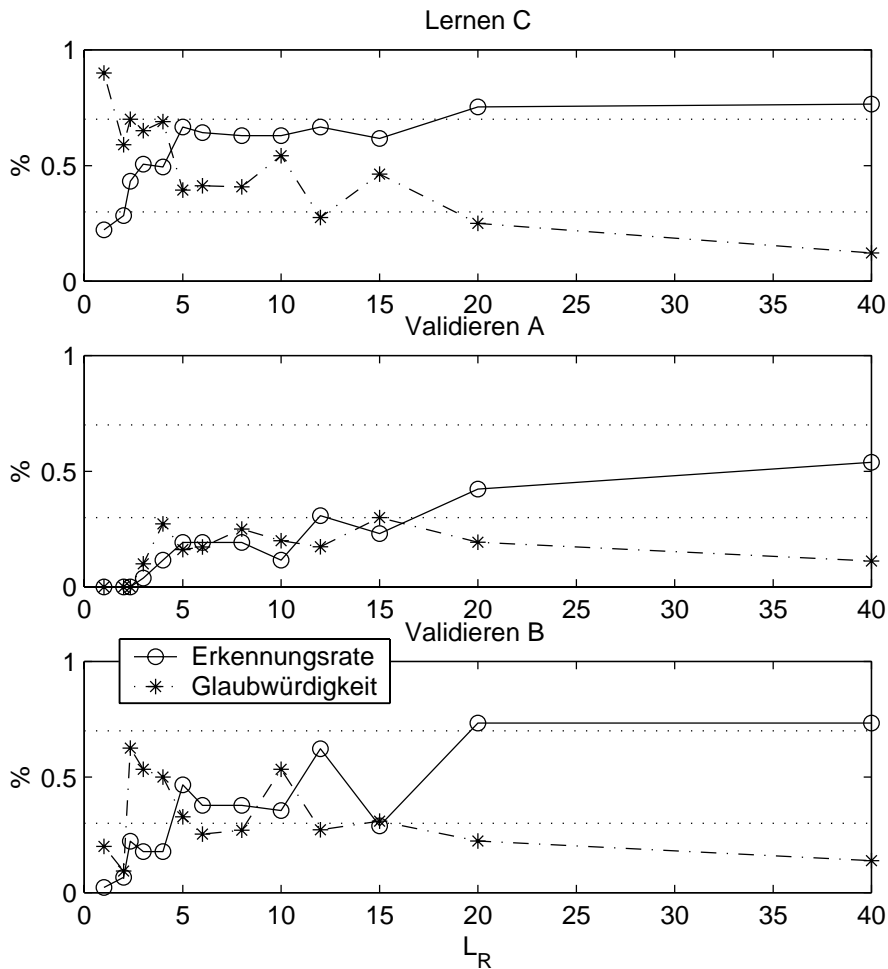


Bild 4.5: Erkennungsrate und Glaubwürdigkeit für den Kfz-Benchmarkdatensatz bei variablem Kostenverhältnis für die Option vollständige Prämisse.

Der Vergleich mit anderen Ansätzen für den Datensatz zeigt die Vorteile des in der vorliegenden Arbeit vorgestellten Verfahrens. Das in [100] vorgestellte Verfahren bedarf einer Vorverarbeitung der Merkmale. Zusätzlich wird die Zusammensetzung des Lerndatensatzes (Teil C) geändert. Die Zahl der defekten Aggregate wird konstant gehalten und die Zahl der einwandfreien erheblich reduziert. Folgende Werte für Erkennungsrate und Glaubwürdigkeit sind mit diesem Verfahren für Teil C als Lerndatensatz zu erreichen:

Die Ergebnisse über den Lerndaten sind beim entscheidungstheoretischen Verfahren etwas besser. Über den beiden Validierungsdatensätzen ergeben sich deutlichere Unterschiede. Für den Teil A liefert das Verfahren aus [100] die besseren Ergebnisse, für den Teil B das entscheidungstheoretische. Der entscheidungstheoretische Zugang stellt hier die benutzerfreundlichere Variante dar. Bis auf die ebenfalls durchgeführte Reduktion der Merkmale von 304 auf 15 entfallen Manipulationen der Daten.

Ohne die Manipulationen liefert das Verfahren aus [100] keine brauchbaren Ergebnisse

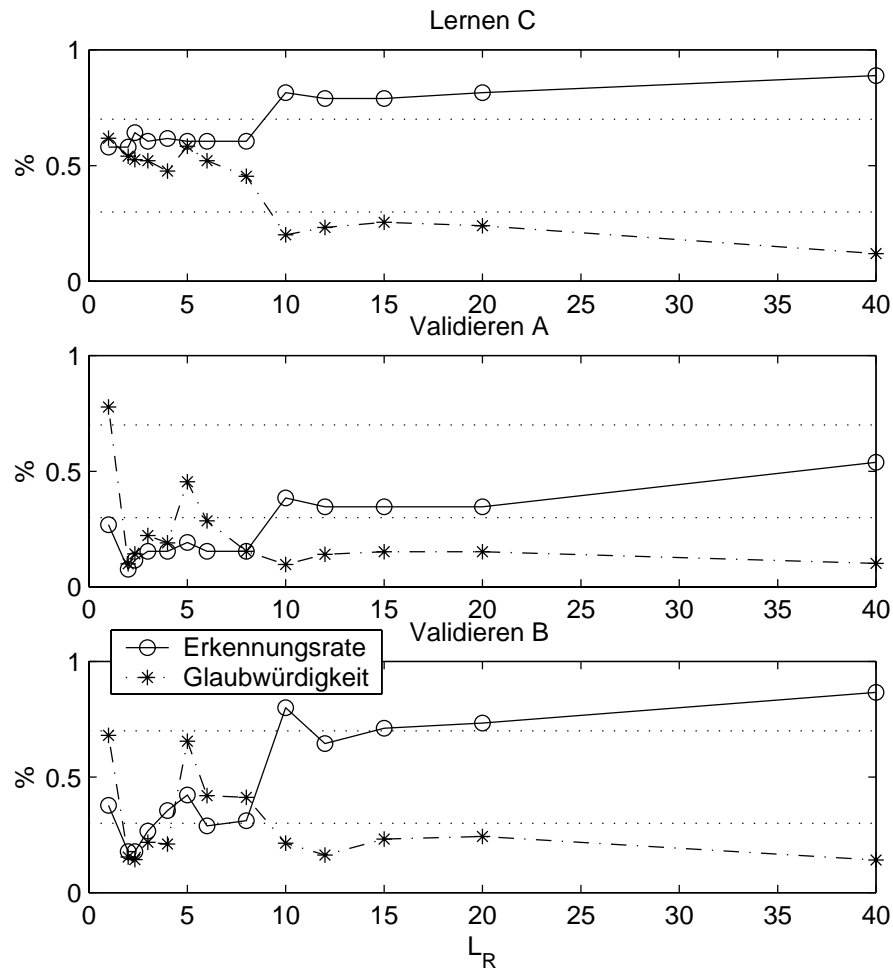


Bild 4.6: Erkennungsrate und Glaubwürdigkeit für den Kfz-Benchmarkdatensatz bei variablem Kostenverhältnis für die Option α -Schnitt.

(Tabelle 4.7). Über den Lerndaten wird eine Erkennungsrate von 26% bei einer Glaubwürdigkeit von 80% erreicht. Über dem Teil A des Datensatzes ergibt sich mit der erzeugten Regelbasis eine Erkennungsrate von 0%. Über dem Teil B des Datensatzes wird eine Erkennungsrate von 4.4% bei einer Glaubwürdigkeit von 33% erreicht. Damit sind die Ergebnisse deutlich schlechter als bei entscheidungstheoretischer Bewertung.

In [65] werden Ergebnisse für das Fuzzy-ROSA¹-Verfahren unter zusätzlicher Berücksichtigung von Expertenwissen bei der Merkmalsauswahl vorgestellt. Es werden mit insgesamt 40 Merkmalen deutlich mehr Merkmale verwendet als beim entscheidungstheoretischen Ansatz in der vorliegenden Arbeit. Die Güte der Ergebnisse ist vergleichbar. Jedoch sind die Regelbasen des entscheidungstheoretischen Ansatzes wesentlich kleiner. Für das Kostenverhältnis von $L_R = 20$ enthalten die Regelbasen 10 Regeln zuzüglich der SONST-

¹Rule Oriented Statistical Analysis

L_R	Erkennungsrate		Glaubwürdigkeit	
	$\mu(P_r)$	α -Schnitt	$\mu(P_r)$	α -Schnitt
1	22.2%	58.0%	90.0%	61.8%
2.3	43.3%	64.2%	70.0%	52.5%
8	63.0%	60.5%	40.8%	45.4%
20	75.3%	81.5%	25.0%	24.0%

Tabelle 4.5: Auswahl der Ergebnisse für den Kfz-Benchmarkdatensatz über den Lern-
daten Teil C bei Variation von L_R .

$L_R = 20$	Erkennungsrate		Glaubwürdigkeit	
	$\mu(P_r)$	α -Schnitt	$\mu(P_r)$	α -Schnitt
Teil A	42.3%	34.6%	19.3%	15.3%
Teil B	73.3%	73.3%	22.3%	24.3%
Teil C	75.3%	81.5%	25.0%	24.0%

Tabelle 4.6: Ergebnisse für den Kfz-Benchmarkdatensatz $L_R = 20$. Lernen Teil C,
Validieren Teil A und Teil B.

Regel. Die in [65] erzeugten Regelbasen haben zwischen 206 und 5200 Regeln.

Zusammenfassend liefert der entscheidungstheoretische Ansatz beim Kfz-Aggregate Benchmark vergleichbare Ergebnisse zu anderen regelbasierten Verfahren [65, 100]. Mit dem in [100] vorgestellten Verfahren sind bisher die besten Ergebnisse erzielt worden, ohne die Anforderungen des Herstellers der Getriebe zu erreichen. Der Vorteil des entscheidungstheoretischen Ansatzes resultiert daraus, dass kein zusätzliches Expertenwissen oder Manipulationen der Zusammensetzung des Lerndatensatzes nötig ist. Allerdings ist es auch mit den entscheidungstheoretischen Ansatz nicht möglich, ein Entscheidungssystem zu entwerfen, das die Anforderungen des Herstellers der Aggregate erfüllt.

		Erkennungsrate	Glaubwürdigkeit
Mit	Teil A	53.8%	26.9%
Vorverarbeitung/ Datenreduktion	Teil B	60.0%	20.6%
	Teil C	72.8%	21.8%
Ohne	Teil A	0%	0%
Vorverarbeitung/ Datenreduktion	Teil B	4.4%	33%
	Teil C	26%	80%

Tabelle 4.7: Ergebnisse für den Kfz-Benchmarkdatensatz aus [100] mit und ohne
Vorverarbeitung beziehungsweise Datenreduktion der Merkmale. Lernen
Teil C, Validieren Teil A und Teil B

4.4 Kontakterkennung einer künstlichen Hand

Künstliche Hände werden beispielsweise für den Einsatz im Weltraum [37, 61, 93], in der Robotik [97, 136] oder als Handprothese [123, 134, 135] entwickelt. Eine umfangreiche Übersicht über künstliche anthropomorphe Hände ist in [107] zu finden.

Eine wichtige Voraussetzung für das sichere Greifen von Objekten mit einer künstlichen Hand ist das exakte Erkennen eines Kontaktes zwischen der Hand und dem Objekt [90, 96]. Die Kontaktinformation kann zum Beispiel zur Strukturumschaltung von Positionsregelung auf Momentenregelung und zur Aktualisierung der Bewegungsplanung der Finger verwendet werden [90]. Dabei sind die Anwendungen von Entscheidungssystemen in der Robotik nicht auf Hände beschränkt [137].

Aktorik und Regelung

Am Forschungszentrum Karlsruhe wird eine künstliche Hand mit flexiblen Fluidaktoren entwickelt [97, 113, 133]. Diese Hand kann sowohl in der Robotik als auch in der Prothetik eingesetzt werden. Die flexiblen Fluidaktoren sind derart an den Gelenken montiert, dass ein steigender Druck im Aktor und die damit einhergehende Expansion des Aktors zu einer Bewegung des Fingergelenkes führt. Über die Elastizität des Aktormaterials bewegt sich das Gelenk bei sinkendem Druck in seine Ausgangslage zurück [132]. Als Positionsregler dient ein adaptiver PID-Regler mit Führungsgrößenaufschaltung und unterlagertem Druckregelkreis [14, 15, 19]. Für die regelungstechnischen Grundlagen sei auf [58, 72, 147] verwiesen. Die Adaption der Reglerparameter Verstärkung K_R und Nachstellzeit T_N basiert auf der Modellierung des Verhaltens des flexiblen Fluidaktors. Die Schätzung des Aktormomentes ist zusätzlich für die Gestaltung einer Kontakterkennung interessant. Für den Zusammenhang zwischen Drehmoment des Aktors M_{Aktor} und dem Druck im Aktor p sowie dem Gelenkwinkel φ stellt das folgende multilineare Modell eine geeignete Approximation dar [14, 15]:

$$\hat{M}_{Aktor} = a_0 + a_1 \cdot \varphi + a_2 \cdot p + a_3 \cdot \varphi \cdot p. \quad (4.8)$$

Bild 4.7 zeigt sowohl die Messwerte als auch die Approximation. Die Messwerte werden über die in [13] vorgestellte Testumgebung aufgezeichnet. Im Bereich zwischen 0 und 70 Grad können die tatsächlichen Momente im Aktor mit den gewählten Parametern mit guter Genauigkeit bestimmt werden [96]. Die Differentialgleichung für die Bewegung eines Einzelgelenkes ergibt sich unter Berücksichtigung externer Störmomente zu:

$$J\ddot{\varphi} + d\dot{\varphi} = M_{Aktor}(p_{Aktor}, \varphi) + M_{Stoer}. \quad (4.9)$$

Über das Aktorkennfeld lässt sich ein oberes Limit für den Druck im Aktor $p_{soll,limit}$ in Abhängigkeit vom Gelenkwinkel berechnen, damit ein maximales Moment M_{max} nicht überschritten wird [14, 18]. Die Struktur der Momentenbegrenzung zeigt Bild 4.8. Die Momentenbegrenzung verhindert bei Objektkontakt große Kräfte auf das Objekt. Ferner schränkt sie die Dynamik des Gelenkes ein, wenn M_{max} kleiner ist, als das Moment, das zur Beschleunigung des Gelenkes notwendig ist.

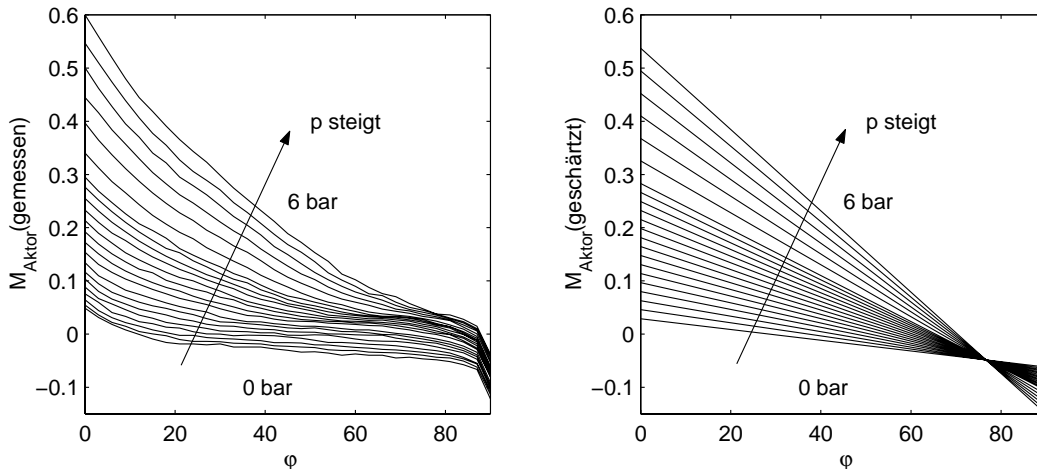


Bild 4.7: Messdaten des Aktormomentes (links) und Approximation (rechts) für einen flexiblen Fluidaktor [18]. Druck zwischen 0 und 6 bar.

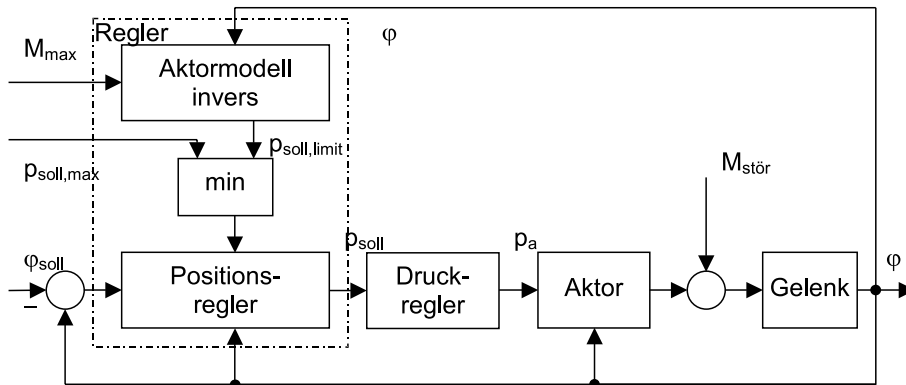


Bild 4.8: Struktur der Momentenbegrenzung [18].

Kontakterkennung

Eine einfache Realisierung einer Kontakterkennung ist über Kraftsensoren an der Oberfläche der Fingerglieder möglich. Der Kraftsensor entspricht dann einem Schalter. Sobald das Signal eines Sensors einen Schwellwert überschreitet, kann von Kontakt an der entsprechenden Stelle des Fingers ausgegangen werden. Die Güte einer solchen Kontakterkennung ist von der Sensitivität des verwendeten Kraftsensors abhängig. In [96] werden FSR²-Sensoren als Kontaktsensoren untersucht. Zusätzliche Kraftsensoren an den Fingern

²Force Sensitive Resistor

bedeuten aber zusätzliche Verkabelung und zusätzliche Kosten. Weiterhin kann der Kontakt zwischen Hand und Umgebung nur dann festgestellt werden, wenn die Kontaktstelle mit einem Sensor versehen ist. Bild 4.9 zeigt eine Situation, in der das Objekt den Sensor trifft. Eine Alternative zu einzelnen Sensoren stellt eine Sensorhaut dar, die über die gesamte Oberfläche der Hand erkennen kann, an welcher Stelle Kontakt zur Umwelt besteht [78].

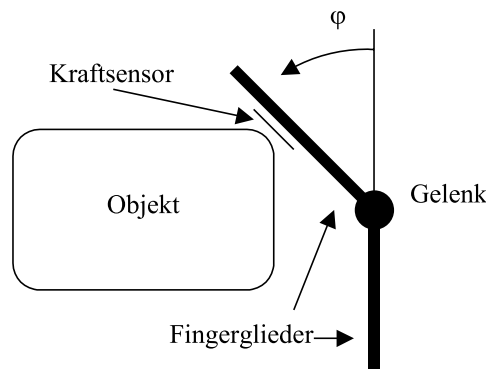


Bild 4.9: Gelenk mit Objekt und Sensor.

Eine Alternative zur Kraftsensorik ist eine modellbasierte Kontakterkennung [18]. Der Objektkontakt wird dabei auf der Basis von gemessenen und geschätzten Informationen bestimmt. Diese Kontakterkennung basiert auf einem manuell erstellten Modell. Als Eingangsgrößen verwendet das Modell das geschätzte Aktormoment \hat{M}_{Aktor} aus (4.8), die Änderung der Führungsgröße φ_{soll} , die Regelabweichung e und die Begrenzung für das Aktormoment M_{max} . Eine Kontakterkennung, die ausschließlich auf dem geschätzten Moment des Aktors basiert, ist nicht sinnvoll. Der Schwellwert ist dann hoch anzusetzen, damit Antriebsmomente des Aktors zur Bewegung des Gelenkes nicht als Kontakt gewertet werden. Dadurch ist der Einsatzbereich der Kontakterkennung eingeschränkt.

Automatischer Entwurf

In diesem Abschnitt wird der automatische Entwurf eines Entscheidungssystems für die Kontakterkennung betrachtet. Für den Entwurfsprozess stehen Datensätze zur Verfügung, die die Betriebsparameter eines Einzelgelenkes beinhalten. Bild 4.10 zeigt den Aufbau für die Generierung der Daten. Das Fingergelenk ist einseitig eingespannt. Ein Objekt kann zwischen einem Gelenkwinkel von 0 und 90° beliebig über den Antrieb des Teststands positioniert werden.

Die Datensätze enthalten sowohl Beispiele für die freie Bewegung des Gelenkes als auch Beispiele für Objektkontakte bei verschiedenen Gelenkwinkeln. Zusätzlich wird das maximal zulässige Moment des Aktors M_{max} variiert. Die Daten haben einen Takt von $t_a = 20$ ms. Die Messwerte aller Größen einer Abtastung entsprechen einem Beispiel. Aus der Kenntnis der Position des Objektes wird über den Winkelsensor während der Aufzeichnung der Daten bestimmt, ob Kontakt vorliegt oder nicht. Die Daten werden entsprechend markiert. Bild 4.11 zeigt einen Ausschnitt aus den Daten für den Entwurf

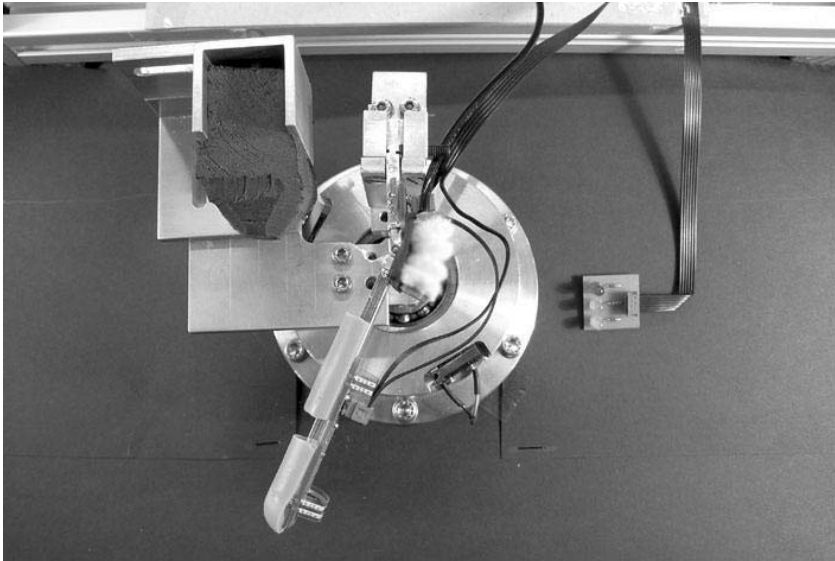


Bild 4.10: Teststand mit eingespanntem Gelenk und Objekt bei 90°.

des Entscheidungssystems. Der Kontaktwinkel des Gelenkes beträgt dabei 50°. Das maximal zulässige Moment des Aktors variiert. Im oberen Teil von Bild 4.11 zeigt die gepunktete Linie die Führungsgröße des Gelenkwinkels. Die durchgezogene Linie markiert den tatsächlichen Verlauf des Gelenkwinkels. Im unteren Teil zeigt die gestrichelte Linie das maximale Moment des Aktors. Die gepunktete Linie entspricht den Messwerten des Drehmomentsensors, der im Teststand eingebaut ist. Die durchgezogene Linie ist die Schätzung des Aktormomentes anhand (4.8) und die strichpunktierte Linie kennzeichnet das Moment auf der Basis der Messwerte eines FSR-Kraftsensors, der am Gelenk befestigt ist. Das Objekt, das das Gelenk berührt, ist nachgiebiger Schaumstoff unterschiedlicher Elastizität. Beim Lerndatensatz (Kontakt A) wird ein steifer Schaumstoff verwendet. Der Validierungsdatsatz (Kontakt B) enthält die Daten eines unabhängigen Experimentes. In Kontakt B besteht das Kontaktobjekt aus einem deutlich weicheren Schaumstoff.

Es sind somit die zwei Umweltzustände 'Kontakt' beziehungsweise 'kein Kontakt' und die gleichnamigen Entscheidungen zu unterscheiden. Ein bestimmtes Kostenverhältnis der Fehlentscheidungen zueinander liegt nicht vor. Tendenziell ist jedoch die Fehlentscheidung kein Kontakt bei vorliegendem Kontakt schwerwiegender als die Fehlentscheidung Kontakt bei tatsächlich freier Bewegung. Bei ersterer kann ein Gegenstand beschädigt werden, bei letzterer wird sich lediglich die Bewegung des Fingers verändern. Es wird daher die selbe Kostenmatrix wie beim Kfz-Aggregate Datensatz verwendet:

$$\mathbf{L} = \begin{pmatrix} 0 & L_R \\ 1 & 0 \end{pmatrix}. \quad (4.10)$$

Bedeutender als das Verhältnis der Kosten der Fehlentscheidungen zueinander ist das Verhältnis der Entscheidungskosten zu den Merkmalskosten. Über die Variation der Merkmalskosten bei konstanten Entscheidungskosten kann zum Beispiel bestimmt werden, bis zu welchen relativen Kosten der Einsatz eines FSR-Sensors sinnvoll ist.

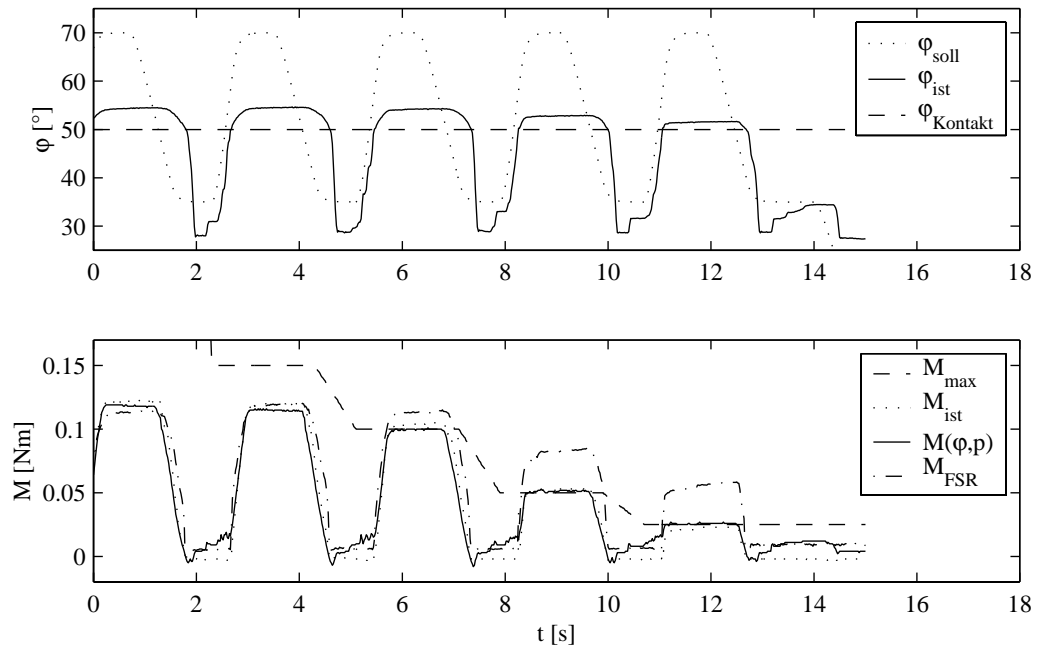


Bild 4.11: Auszug aus den Daten für den automatischen Entwurf der Kontaktenerkennung. Kontaktwinkel bei 50° . Maximales Moment variiert.

Die Entscheidungssysteme werden mit der Option $\mu(P_r)$ für die Aktivierung der Einzelregeln sowie der Option Rückweisung für die Regelbasissuche entworfen. Die Ergebnisse für die Option α -Schnitt für die Aktivierung der Einzelregeln unterscheiden sich nur unwesentlich von denen der Option $\mu(P_r)$. Daher werden nur Ergebnisse für die Option $\mu(P_r)$ vorgestellt. Der Einsatz eines Experten ist bei dieser Aufgabe nicht sinnvoll, da ein Aufschieben der Entscheidung nicht möglich ist. Daher wird die Option Evidenz-Inferenz nicht untersucht.

Entscheidungssystem ohne Merkmalskosten

Werden die Merkmalskosten nicht berücksichtigt, verwendet das entworfene Entscheidungssystem unabhängig vom Kostenverhältnis stets das Merkmal FSR-Sensor. Es sind dann unter anderem die in Tabelle 4.8 angegebenen Regeln in der Regelbasis integriert. Zur Veranschaulichung der Ergebnisse der Entscheidungssysteme bei verschiedenen Kostenverhältnissen werden die Erkennungsrate und die Fehlerrate verwendet. Die Fehlerrate FR ist dabei der Anteil falscher Entscheidungen bezogen auf alle Entscheidungen e_2 'Kontakt':

$$FR = \frac{\hat{P}(e_2 \wedge z_1)}{\hat{P}(e_2 \wedge z_1) + \hat{P}(e_2 \wedge z_2)} = \frac{\hat{P}(e_2 \wedge z_1)}{\hat{P}(e_2)}. \quad (4.11)$$

Die grundlegenden Zusammenhänge zwischen Kostenverhältnis und Erkennungsrate sowie Fehlerrate werden in Bild 4.12 deutlich. Die durchgezogene Linie stellt die Werte für

r	WENN	DANN
1	F2= NICHT $a_{10,1}$	$C_1 = e_2$
2	F2= $a_{10,1}$	$C_2 = e_1$

Tabelle 4.8: Regeln der Kontakterkennung für den FSR Sensor (F2, x_{10}), der mit dem Objekt in Kontakt tritt, $a_{10} = (0, 1, 2, 3, 4)$.

Kontakt A und die strichpunktierte Linie für Kontakt B dar. Mit steigendem Kostenverhältnis steigt die Erkennungsrate. Die Fehlerrate nimmt ebenfalls zu, da zur Steigerung der Erkennungsrate zusätzliche Fehlentscheidungen verursacht werden. Über das Signal des FSR-Sensors allein sind die Beispiele für Kontakt nicht vollständig von denen für keinen Kontakt zu trennen. Die Robustheit der entworfenen Entscheidungssysteme zeigt sich daran, dass für den Testdatensatz Kontakt B ein ähnliches Bild vorliegt, obwohl wegen des weicheren Schaumstoffs die Zusammenhänge von denen im Lerndatensatz Kontakt A abweichen.

Entscheidungssystem mit Merkmalskosten

Für alle Merkmale, die zum Betrieb des Reglers notwendig sind, werden keine Merkmalskosten angesetzt. Für die folgenden Merkmale werden Kosten verrechnet:

- Schätzung des Aktormoment M_{Aktor} , $L_{C,7} = 0.01$,
- FSR-Sensor Signal, $L_{C,10} = 0.1$,
- Ableitung der Führungsgröße $\dot{\varphi}_{soll}$, $L_{C,13} = 0.005$.

Es handelt sich bei den angesetzten Kosten um Schätzungen. Wichtiger als die absolute Höhe der Kosten ist deren Verhältnis. Es wird angenommen, dass für die Integration eines FSR-Sensors 10 mal höhere Kosten anfallen als für die Nutzung der Schätzung des Aktormomentes. Die Berechnung der Ableitung der Führungsgröße erfordert einen geringeren Rechenaufwand als die Berechnung der Schätzung des Aktormomentes.

Die Ergebnisse für die verschiedenen Kostenverhältnisse L_R sind in Bild 4.13 zusammengestellt. Wiederum ist zu erkennen, dass die Erkennungsrate mit steigendem Kostenverhältnis zunimmt. Allerdings steigt die Fehlerrate zunächst, sinkt dann aber für ein Kostenverhältnis von $L_R = 8$ wieder. Die Ursache dafür liegt in den verwendeten Merkmalen. Für $L_R = 6$ wird das Signal des FSR-Sensors für das Entscheidungssystem nicht verwendet. Ab $L_R = 8$ wird das FSR-Signal verwendet. Ohne das FSR-Signal wird das Steigen des Kostenverhältnisses durch stärker generalisierte Regeln kompensiert. Dadurch werden von den Entscheidungssysteme, die das Signal des FSR-Sensors nicht nutzen, höhere Fehlerraten verursacht als von denen, die das Signal des FSR-Sensors nutzen.

Für ein Kostenverhältnis von $L_R = 3$ resultiert ein Entscheidungssystem mit einer Erkennungsrate von 98.3% bei einer Glaubwürdigkeit von 7.2%. Für die Validierungsdaten ergeben sich die Werte 97.4% und 5.8%. Die Regelbasis des Entscheidungssystems ist in Tabelle 4.9 dargestellt. Tabelle 4.10 zeigt die Stützfunktionen der Zugehörigkeitsfunktionen der verwendeten Merkmale. Die wichtigsten Zusammenhänge der Regelbasis werden nun kurz erläutert.

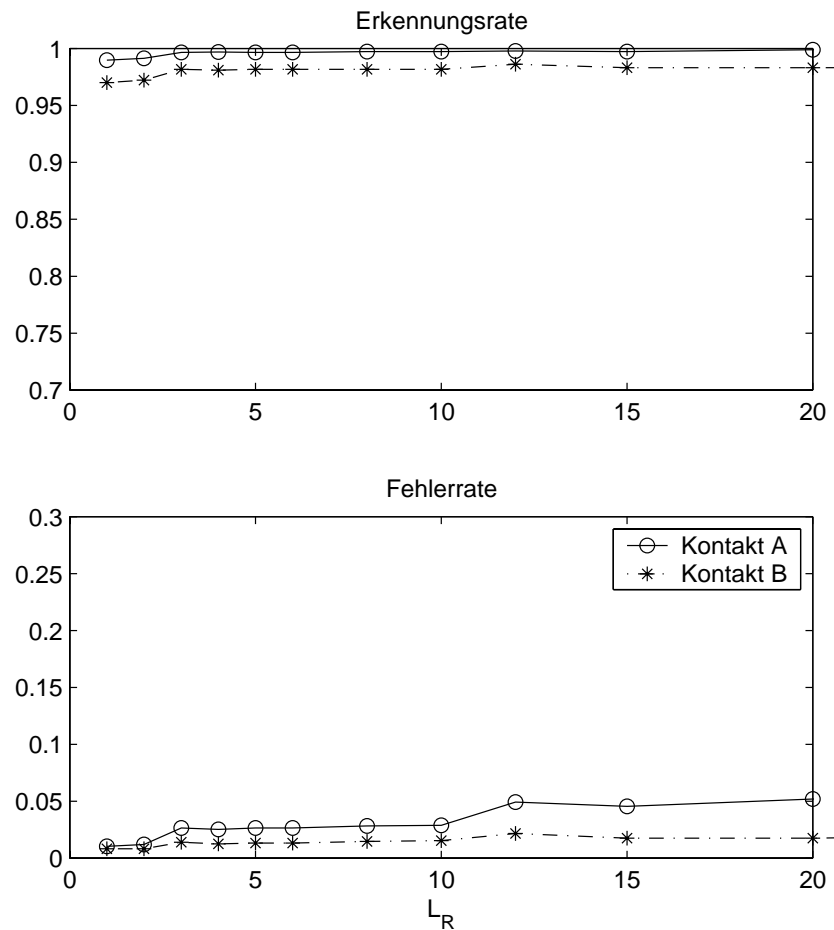


Bild 4.12: Erkennungsrate und Fehlerrate für die Kontakterkennung bei variablem Kostenverhältnis ohne Merkmalskosten.

Die erste Regel sagt aus, dass für geschätzte Drehmomente \hat{M}_{Aktor} größer 0.02 Nm Kontakt vorliegt. Für die freie Bewegung des Aktors sind tatsächlich selten Drehmomente notwendig, die größer sind als der Wert 0.02 Nm. Die dritte Regel bedeutet, dass für mittlere Regelabweichungen e und kleine Drehmomente kein Kontakt vorliegt. Damit wird der Betrieb des Reglers bei freier Bewegung beschrieben. Einen ähnlichen Zusammenhang drückt Regel 7 aus. Ein kleiner Solldruck im Aktor und gleichzeitig kleine Drehmomente lassen auf freie Bewegung schließen. Darüber hinaus nutzt das Entscheidungssystem bei großen Sollwinkeln die Regelabweichung zur Detektion von Kontakt (Regeln 4 und 5). Aus der entworfenen Regelbasis lassen sich somit plausible Zusammenhänge für das betrachtete System ablesen.

In Bild 4.14 sind für die Daten aus Bild 4.11 die Ergebnisse der Kontakterkennung dargestellt. Im unteren Teil stellt die gepunktete Linie den aus der geometrischen Position des Gelenkes abgeleiteten tatsächlichen Umweltzustand dar. Die durchgezogene Linie entspricht dem Ausgang der Kontakterkennung. Tendenziell erkennt das Entscheidungssystem

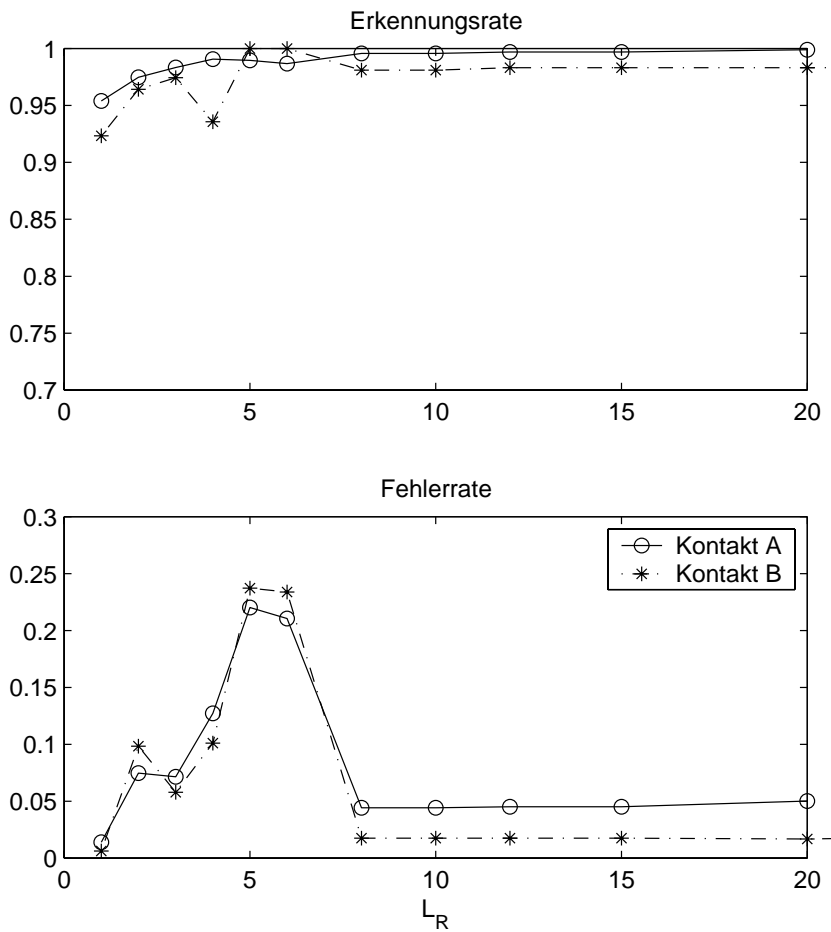


Bild 4.13: Erkennungsrate und Fehlerrate für die Kontakterkennung bei variablem Kostenverhältnis mit Merkmalskosten.

den Kontakt etwas (ein- bis zweifache Zykluszeit des Reglers) zu spät. Ein Grund hierfür ist, dass der Regler diese Zeitspanne benötigt, um auf den Kontakt zu reagieren. Erst durch einen erhöhten Druck, um φ_{soll} zu erreichen, steigt das geschätzte Aktormoment und damit die Aktivierung der ersten Regel aus Tabelle 4.9. Beim Übergang von Kontakt auf keinen Kontakt reagiert das Entscheidungssystem ebenfalls mit einer Verspätung vom ein- bis zweifachen der Zykluszeit.

Es ist allerdings darauf zu achten, dass die Daten das gegebene Problem ausreichend wiedergeben. Ist der Datensatz unvollständig, können irreführende Zusammenhänge erkannt werden. Zur Veranschaulichung werden aus dem Lerndatensatz alle Beispiele für Kontakt bei Gelenkwinkeln kleiner als 65° gelöscht. Anschließend wird ein Entscheidungssystem unter Berücksichtigung von Merkmalskosten für $L_R = 5$ entworfen. Tabelle 4.11 zeigt die resultierende Regelbasis und Tabelle 4.12 die Zugehörigkeitsfunktionen. Regel 1 bedeutet, dass kein Kontakt vorliegt, wenn der Gelenkwinkel φ_{ist} kleiner ist als 65° . Der Zusammenhang stimmt allerdings nur für den modifizierten Datensatz. Für den gesam-

r	WENN	DANN
1	$\hat{M}_{Aktor} = \text{NICHT } A_{7,1}$	$C_1 = e_2$
2	$(\varphi_{ist} = A_{2,2} \cup A_{2,3} \cup A_{2,4}) \cap (\dot{\varphi}_{soll} = A_{13,1} \cup A_{13,2} \cup A_{13,3})$	$C_2 = e_2$
3	$(e = A_{4,2} \cup A_{4,3} \cup A_{4,4}) \cap \hat{M}_{Aktor} = A_{7,1}$	$C_3 = e_1$
4	$\varphi_{soll} = A_{1,5} \cap (e = A_{4,3} \cup A_{4,4} \cup A_{4,5})$	$C_4 = e_2$
5	$\varphi_{soll} = A_{1,4} \cap (e = A_{4,1} \cup A_{4,2})$	$C_5 = e_2$
6	$e = A_{4,1}$	$C_6 = e_2$
7	$(p_{soll} = A_{3,1} \cup A_{3,2}) \cap \hat{M}_{Aktor} = A_{7,1}$	$C_7 = e_1$
	SONST	$C = e_2$

Tabelle 4.9: Regelbasis für die Kontakterkennung bei Berücksichtigung von Merkmalskosten. $L_R = 3$, e_1 : kein Kontakt, e_2 : Kontakt.

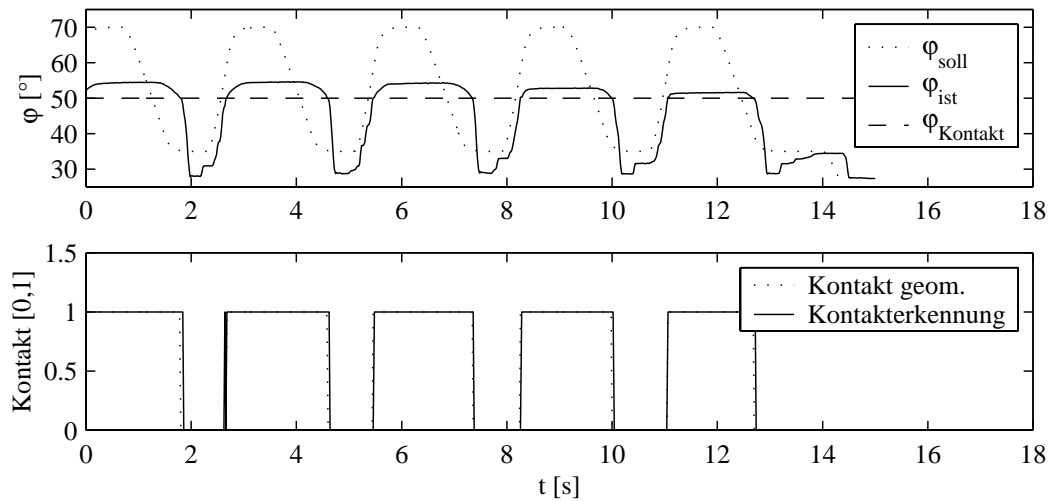


Bild 4.14: Ergebnis der Kontakterkennung für die Daten aus Bild 4.11. $L_R = 3$, Merkmalskosten beim Entwurf berücksichtigt, Regelbasis aus Tabelle 4.9.

ten Datensatz ist die Regelbasis wertlos. Das Beispiel zeigt wiederum die Vorteile der Interpretierbarkeit der Fuzzy-Systeme. Dadurch können entworfene Fuzzy-Systeme vom Anwender auf ihre Plausibilität hin geprüft werden, was das Verwerfen oder Korrigieren nichtplausibler Lösungen in einem nachfolgenden Arbeitsschritt ermöglicht.

Merkmal (l)	$a_{l,1}$	$a_{l,2}$	$a_{l,3}$	$a_{l,4}$	$a_{l,5}$
$\varphi_{soll}(1)$	0.00	28.00	32.00	50.00	70.00
$\varphi_{ist}(2)$	30.00	40.00	50.00	60.00	70.00
$p_{soll}(3)$	0.00	1.00	2.00	4.00	6.00
$e(4)$	-30.00	0.00	5.00	10.00	40.00
$\hat{M}_{Aktor}(7)$	0.00	0.04	0.10	0.16	0.22
$\dot{\varphi}_{soll}(13)$	-200.00	0.00	10.00	40.00	100.00

Tabelle 4.10: Stützpunkte der Zugehörigkeitsfunktionen für die Regelbasis aus Tabelle 4.9.

r	WENN	DANN
1	$\varphi_{ist} = \text{NICHT } A_{1,5}$	$C_1 = e_1$
2	$e = A_{4,3} \cup A_{4,4}$	$C_2 = e_2$
	SONST	$C = e_1$

Tabelle 4.11: Regelbasis für die Kontakterkennung bei Berücksichtigung von Merkmalskosten. $L_R = 5$. Datensatz unvollständig (nur Kontakt bei Gelenkwinkeln größer 65°).

Merkmal (l)	$a_{l,1}$	$a_{l,2}$	$a_{l,3}$	$a_{l,4}$	$a_{l,5}$
$\varphi_{ist}(2)$	30.00	40.00	50.00	60.00	70.00
$e(4)$	-10.00	0.00	3.00	10.00	20.00

Tabelle 4.12: Stützpunkte der Zugehörigkeitsfunktionen für die Regelbasis aus Tabelle 4.11.

4.5 Diskussion der Ergebnisse

Bei den beiden Datensätzen aus dem Statlog-Projekt liefern die entscheidungstheoretischen Fuzzy-Systeme ähnlich gute (Heart-Disease) beziehungsweise nur wenig schlechtere Ergebnisse (German-Credit) als entscheidungstheoretisch ausgewertete statistische Verfahren. Die Fuzzy-Verfahren bieten zusätzlich den Vorteil der Interpretierbarkeit der entworfenen Entscheidungssysteme. Beim Heart-Disease Datensatz sinken die Entscheidungskosten, wenn die Kosten für Merkmale berücksichtigt werden. Die Ursache dafür ist, dass die Spezialisierung der entworfenen Entscheidungssysteme abnimmt und somit das Overfitting an die Lerndaten vermindert wird.

Beim Kfz-Aggregate Datensatz erreicht das vorgestellte Verfahren ähnlich gute Ergebnisse wie das Verfahren aus [100]. Allerdings ist bei entscheidungstheoretischer Bewertung keine aufwändige Vorverarbeitung der Daten und Modifikation der Zusammensetzung des Datensatzes notwendig. Das entscheidungstheoretische Verfahren ist somit für den Benutzer einfacher zu bedienen. Allerdings ist es mit dem entscheidungstheoretischen Verfahren ebenfalls nicht möglich, die Anforderungen des Herstellers der Aggregate zu erfüllen.

Am Beispiel der Kontakterkennung zeigt sich ebenfalls die Leistungsfähigkeit der entscheidungstheoretischen Bewertung. Durch Variation der Entscheidungskosten können zusätzlich Verhältnisse zwischen Merkmalskosten und Entscheidungskosten bestimmt werden, ab denen der Einsatz eines Sensors sinnvoll ist. Die Interpretierbarkeit der entworfenen Entscheidungssysteme zeigt ebenfalls ihre Bedeutung, da unzulässige Zusammenhänge, die aus den Daten abgeleitet werden, vom Anwender schnell erkannt werden können.

Folgende Empfehlungen für das Vorgehen beim Entwurf von Entscheidungssystemen werden gegeben:

- Entwurf einer einstufigen Regelbasis entsprechend der Anforderungen mit oder ohne Merkmalskosten. Für die Risikoparameter empfehlen sich Werte von $\alpha_{ER} = \alpha_{RB} = 1$. Beide Pruning-Optionen α -Schnitt beziehungsweise $\mu(P_r)$ sind zu untersuchen.
- Bei vollständigen Datensätzen kann anschließend eine Optimierung der Zugehörigkeitsfunktionen erfolgen.
- Sollen Merkmalskosten nicht berücksichtigt werden, empfiehlt sich die Ergebnisse der Regelbasissuche mittels des genetischen Algorithmus zu prüfen.
- Der Entwurf einer hierarchischen Regelbasis kann ebenfalls zur Verbesserung der Ergebnisse beitragen. Eine mögliche Verbesserung ist aber stark von der Vollständigkeit des Datensatzes abhängig.

Es steht somit ein leistungsfähiges Verfahren zum Entwurf von Fuzzy-Entscheidungssystemen für Probleme mit unterschiedlichen Konsequenzen verschiedener Fehlentscheidungen zur Verfügung.

5 Zusammenfassung und Ausblick

Für den Entwurf von interpretierbaren Fuzzy-Entscheidungssystemen, die unterschiedliche Konsequenzen für Fehlentscheidungen berücksichtigen, werden neue Bewertungsmaße und modifizierte Entwurfsverfahren vorgestellt. Der wesentliche Unterschied zu Entscheidungssystemen, die die Anzahl realisierter Fehlentscheidungen minimieren, besteht in der Differenzierung der Konsequenzen der einzelnen Kombinationen aus Entscheidung und tatsächlichem Umweltzustand. Die Bewertung dieser Entscheidungssysteme basiert auf einer Risikobetrachtung, die der Entscheidungstheorie entnommen wird. Neben den Konsequenzen der Entscheidungen ist auch die Reduktion der notwendigen Aufwendungen für Merkmale Teil der Optimierung der in der vorliegenden Arbeit betrachteten Entscheidungssysteme.

Die inhaltlichen Schwerpunkte der Arbeit umfassen:

- die Entwicklung neuer entscheidungstheoretischer Bewertungsmaße für Fuzzy-Einzelregeln und Fuzzy-Regelbasen und die Entwicklung eines neuen Inferenzverfahrens für Fuzzy-Regelbasen auf der Basis der Evidenztheorie (Kapitel 2),
- die Modifikation und die Erweiterung bestehender Verfahren zum automatischen Entwurf von Fuzzy-Einzelregeln und Fuzzy-Regelbasen zur Integration der neuen entscheidungstheoretischen Bewertungsmaße (Kapitel 3),
- die Evaluierung der vorgestellten Maße und Verfahren an Benchmarkdatensätzen und einer realitätsnahen Anwendung aus der Robotik (Kapitel 4).

Im Folgenden werden die wichtigsten Ergebnisse der Arbeit zusammengestellt:

1. Entwicklung eines Konzeptes für den automatischen Entwurf von Entscheidungssystemen bei Unsicherheit mittels der Theorie der unscharfen Mengen und der Evidenztheorie.
2. Herleitung von neuen Bewertungsmaßen für Entscheidungsprobleme bei Unsicherheit. Dazu gehören ein Bewertungsmaß für unscharfe Einzelregeln und ein Bewertungsmaß für unscharfe Regelbasen.
3. Erarbeitung eines neuen Dominanzkriteriums zur zweikriteriellen Bewertung von unscharfen Einzelregeln auf der Basis der Entscheidungskosten und der Merkmalskosten während des Prunings.
4. Ableitung eines neuen Inferenzverfahrens für unscharfe Regelbasen mittels Evidenztheorie.

5. Erarbeitung eines Modells, das die Konklusion einer unscharfen Regel als Schätzung der besten Entscheidung für die von der Prämisse abgedeckten Beispiele identifiziert.
6. Aufbau mehrstufiger Regelbasen, die spezieller als einstufige sind, die aber nicht in allen Fällen leistungsfähiger sind.
7. Neues Verfahren zur Visualisierung der Güte von Entscheidungssystemen für Probleme mit zwei Umweltzuständen und direkter Ermittlung der Entscheidungsfunktion.
8. Implementierung und Erprobung aller vorgestellten neuen Maße und Verfahren in der Programmierumgebung Matlab und Integration in das Programmpaket KAFKA.
9. Demonstration der Leistungsfähigkeit der Maße und Verfahren an anerkannten Benchmarkdatensätzen (German-Credit, Heart-Disease, Kfz-Aggregate) und einem realitätsnahen Problem aus der Robotik.
10. Nachweis der Eignung des auf der Evidenztheorie basierenden Inferenzverfahren zur Reduktion der Kosten für Fehlentscheidungen und der Reduktion von Merkmalskosten bei mehrstufigen Regelbasen anhand des Heart-Disease Datensatzes.
11. Nachweis der Vorteile der entscheidungstheoretischen Entwurfsverfahren gegenüber herkömmlichen Entwurfsverfahren für Fuzzy-Systeme am Beispiel des Kfz-Aggregate Benchmarkdatensatzes.
12. Demonstration des Vorteils der Interpretierbarkeit der entworfenen Entscheidungssysteme anhand eines reduzierten Datensatzes für das Beispiel der Kontakterkennung einer künstlichen Hand.

Die vorgestellten Maße und Verfahren sind vollständig implementiert und in ein vorhandenes Programmpaket integriert. Dies erlaubt die zukünftige Anwendung der Maße und Verfahren für beliebige weitere Problemstellungen. Dazu sind ein Lerndatensatz mit den Ausprägungen der Merkmale und der Umweltzustände sowie die Menge der möglichen Entscheidungen und eine Kostenmatrix bereit zu stellen.

Weiterführende Arbeiten sind zum Beispiel bei der Optimierung der Suchstrategien für Einzelregeln und Regelbasen notwendig. Insbesondere bei Problemen mit vielen Merkmalen decken die existierenden Verfahren nicht den kompletten Suchraum ab. Daher werden potentiell geeignete Regeln oder Regelbasen nicht gefunden. Am Beispiel des Kfz-Aggregate Datensatzes konnte zwar gezeigt werden, dass mit den entscheidungstheoretischen Verfahren der Aufwand für die Datenvorverarbeitung reduziert werden kann und damit die Benutzerfreundlichkeit erhöht wird, allerdings erreichen die erzielten Ergebnisse noch nicht die Anforderungen des Herstellers der Aggregate.

Die in der vorliegenden Arbeit entwickelten Entwurfsverfahren für mehrstufige Regelbasen sind weiter zu verbessern. Zu untersuchen ist beispielsweise, ob durch speziell gestaltete Einzelregeln eine Verbesserung der Ergebnisse bei den mehrstufigen Regelbasen erreicht werden kann. Die erzeugten mehrstufigen Regelbasen sind spezieller als einstufige. Deswegen schneidet dieses Verfahren bei einer Kreuzvalidierung schlechter ab, als die Verfahren, die einstufige Regelbasen erzeugen. Zu untersuchen ist, ob ein geeignetes Pruningverfahren für die mehrstufige Regelbasis im Anschluss an den Entwurf das Problem

beheben kann. Um das Konzept der mehrstufigen Regelbasen eindeutig zu bestätigen oder um es zu verwerfen, sind zusätzliche Tests an weiteren Problemstellungen nötig.

Eine Verallgemeinerung der vorgestellten Entwurfsmethodik auf Systeme mit variablen Strukturen ist ebenfalls denkbar. Ein einfaches Beispiel für ein solches System ist das Getriebe eines Kraftfahrzeuges. In Abhängigkeit vom eingelegten Gang liegen unterschiedliche Zusammenhänge zwischen Eingangs- und Ausgangsdrehzahl vor. Für die Diagnose solcher Systeme sind dann Entscheidungssysteme notwendig, die selbst variable Strukturen besitzen.

Als Weiterführung der Anwendung der Kontakterkennung für einen einzelnen Finger einer künstlichen Hand ist der Entwurf eines Entscheidungssystems zur Bewertung der Griffsituation der kompletten Hand denkbar. Industrielle Anwendungen in der Diagnose von Prozessen oder zur Entscheidungsunterstützung bei der Qualitätssicherung sind ebenfalls möglich.

Literaturverzeichnis

- [1] ADAMS, N. M.; HAND, D. J.: Comparing classifiers when the misallocation costs are uncertain. *Pattern Recognition* 32(7) (1999), S. 1139–1147.
- [2] AHLBRANDT, R.; IYER, H. K.; VECCHIA, D. F.: Optimum Design of Serial Measurement Trees. *Journal of Statistical Planning and Inference* 48(3) (1995), S. 379–390.
- [3] AIT KBIR, M.; BENKIRANE, H.; MAALMI, K.; BENSLIMANE, R.: Hierarchical Fuzzy Partition for Pattern Classification with Fuzzy If-Then Rules. *Pattern Recognition Letters* 21 (2000), S. 503–509.
- [4] ALTINÇAY, H.; DEMIREKLER, M.: Speaker identification by combining multiple classifiers using Dempster-Shafer theory of evidence. *Speech Communication* 41(4) (2003), S. 531–547.
- [5] ANGELOV, P. P.; BUSWELL, R. A.: Automatic generation of fuzzy rule-based models from data by genetic algorithms. *Information Sciences* 150(1/2) (2003), S. 17–31.
- [6] BACKHAUS, K.; ERICHSON, B.; PLINKE, W.; WEIBER, R.: *Multivariate Analysemethoden*. Springer, 2000.
- [7] BAMBERG, G.: *Statistische Entscheidungstheorie*. Physica Verlag, 1972.
- [8] BAMBERG, G.; COENENBERG, A. G.: *Betriebswirtschaftliche Entscheidungslehre*. Verlag Franz Vahlen GmbH, 2002.
- [9] BANDEMER, H.: *Ratschläge zum mathematischen Umgang mit Ungewißheit*. B.G. Teubner Verlagsgesellschaft Stuttgart, 1997.
- [10] BANDEMER, H.; GOTTWALD, S.: *Einführung in Fuzzy Methoden*. Akademie-Verlag, Berlin, 1993.
- [11] BANDEMER, H.; NÄTHER, W.: *Fuzzy Data Analysis*. Kluwer Academic Publishers, 1992.
- [12] BAUER, M.: Approximation algorithms and decision making in the Dempster-Shafer theory of evidence - An empirical study. *International Journal of Approximate Reasoning* 17(2/3) (1997), S. 217–237.

- [13] BECK, S.: *Entwicklung, Aufbau und Erprobung einer Testumgebung für die Evaluierung flexibler Fluidaktoren*. Diplomarbeit, Universität Karlsruhe (TH), 2001.
- [14] BECK, S.; LEHMANN, A.; LOTZ, T.; MARTIN, J.; KEPPLER, R.; MIKUT, R.: Modellgestützte adaptive Regelungskonzepte für eine fluidisch betriebene Roboterhand. In: *Proceedings GMA-Kongress*, S. 65–72, VDI Verlag, 2003.
- [15] BECK, S.; LEHMANN, A.; MARTIN, J.; MIKUT, R.: Modellbildung und Fuzzy-Gelenkpositionsregelung für eine 5-Finger-Roboterhand mit flexiblen Fluidaktoren. In: *Proceedings 12. Workshop Fuzzy Systeme* (MIKUT, R.; REISCHL, M., Hg.), S. 177–191, Forschungszentrum Karlsruhe Wissenschaftliche Berichte FZKA 6767, 2002.
- [16] BECK, S.; MIKUT, R.; JÄKEL, J.: A Cost-Sensitive Learning Algorithm for Fuzzy Rule-Based Classifiers. *submitted to: Mathware and Soft Computing* (2004).
- [17] BECK, S.; MIKUT, R.; JÄKEL, J.; BRETTHAUER, G.: Decision-Theoretic Approaches in Fuzzy Rule Generation for Diagnosis and Fault Detection Problems. In: *Proc. Eusflat 2003*, S. 558–563, 2003.
- [18] BECK, S.; MIKUT, R.; LEHMANN, A.; BRETTHAUER, G.: Model-Based Control and Object Contact Detection for a Fluidic Actuated Robotic Hand. In: *42nd IEEE Conference on Decision and Control* (J. JIM ZHU, P. C., Hg.), S. 6369–6374, IEEE, 2003.
- [19] BECK, S.; REISCHL, R.; BRETTHAUER, G.: Steuerungs- und Regelungskonzepte für eine myoelektrische Handprothese. In: *37. Regelungstechnisches Kolloquium in Boppard* (EPPLER, U., Hg.), S. 43, 2003.
- [20] BELLMANN, R.; ZADEH, L.: Decision making in a fuzzy environment. *Management Science* 17(4) (1970), S. 141–164.
- [21] BERARDI, V. L.; ZHANG, G. P.: The Effect of Misclassification Costs on Neural Network Classifiers. *Decision Sciences Journal* 30(3) (1999), S. 659–682.
- [22] BERGER, J. O.: *Statistical Decision Theory and Bayesian Analysis*. Springer Verlag, 1980.
- [23] BERNARD, T.: *Ein Beitrag zur gewichteten multikriteriellen Optimierung von Heizungs- und Lüftungsregelkreisen auf Grundlage des Fuzzy Decision Making*. Dissertation, Universität Karlsruhe (TH), 2000.
- [24] BEYERER, J.: *Verfahren zur quantitativen statistischen Bewertung von Zusatzwissen in der Meßtechnik*. Düsseldorf: VDI Fortschritt-Berichte, Reihe 8, Nr. 783, VDI Verlag, 1999.
- [25] BEZDEK, J.; SPILLMAN, B.; SPILLMAN, R.: A Fuzzy Relation Space for Group Decision Theory. *Fuzzy Sets and Systems* 1 (1978), S. 255–268.

- [26] BINAGHI, E.; GALLO, I.; MADELLA, P.: A Neural Model for Fuzzy Dempster-Shafer Classifiers. *International Journal of Approximate Reasoning* 25 (2000), S. 89–121.
- [27] BINAGHI, E.; MADELLA, P.: Fuzzy Dempster-Shafer Reasoning for Rule-Based Classifiers. *International Journal of Intelligent Systems* 14 (1999), S. 559–583.
- [28] BITZ, M.: *Entscheidungstheorie*. Verlag Vahlen, 1981.
- [29] BOGLER, L. P.: Shafer-Dempster Reasoning with Applications to Multisensor Target Identification Systems. *IEEE Transactions on Systems, Man, and Cybernetics* 17(6) (1987), S. 968–977.
- [30] BÖHME: *Fuzzy-Logik*. Springer, 1993.
- [31] BONARINI, A.: Evolutionary Learning of General Fuzzy Rules with Biased Evaluation Functions: Competition and Cooperation. In: *Proc. 1st IEEE Conf. on Evolutionary Computation*, Bd. 1, S. 51–56, Piscataway, NJ: IEEE Press, 1994.
- [32] BRADFORD, J.; KUNZ, C.; KOHAVI, R.; BRUNK, C.; BRODLEY, C.: Pruning Decision Trees with Misclassification Costs. In: *Proceedings ECML-98*, S. 131–136, 1998.
- [33] BREIMAN, L.; FRIEDMAN, J. H.; OLSHEN, R. A.; STONE, C. J.: *Classification and Regression Trees*. Belmont, CA: Wadsworth, 1984.
- [34] BRETTHAUER, G.: *Anwendung der statistischen Entscheidungstheorie zur Modellsélection in der offenen Wirkungskette und im geschlossenen Regelkreis*. Dissertation, TU Dresden, 1977.
- [35] BRETTHAUER, G.; GAMALEJA, T.; HANDSCHIN, E.; NEUMANN, U.; HOFFMAN, W.: Integrated Maintenance Scheduling System for Electrical Energy Systems. *IEEE Transactions on Power Delivery* 13(2) (1998), S. 655–660.
- [36] BUCKLEY, J. J.; HAYASHI, Y.: Fuzzy Neural Networks: A Survey. *Fuzzy Sets and Systems* 66(1) (1994), S. 1–13.
- [37] BUTTERFASS, J.; GREBENSTEIN, M.; LIU, H.; HIRZINGER, G.: DLR Hand II: Next generation of a dextrous robot hand. In: *International Conference on Robotics and Automation, Seoul, Korea, 2001*.
- [38] CARLSSON, C.; FULLÉR, R.: *Fuzzy Reasoning in Decision Making and Optimization*. Physica-Verlag Heidelberg, 2002.
- [39] CASILLAS, J.; CORDÓN, O.; HERRERA, F.; MAGDALENA, L. (Hg.): *Trade-off between Accuracy and Interpretability in Fuzzy Rule-Based Modelling*. Studies in Fuzziness and Soft Computing, Heidelberg: Physica, 2002.
- [40] CHAN, P. K.; STOLFO, S. J.: Toward Scalable Learning with Non-Uniform Class and Cost Distributions: A Case Study in Credit Card Fraud Detection. In: *Proc. Fourth Intl. Conf. Knowledge Discovery and Data Mining*, S. 164–168, 1998.

- [41] CHEN, M. Y.; LINKENS, D. A.: Rule-base self-generation and simplification for data-driven fuzzy models. *Fuzzy Sets and Systems* 142(2) (2004), S. 243–265.
- [42] COELLO, C. A. C.; MONTES, E.: Constraint-handling in genetic algorithms through the use of dominance-based tournament selection. *Advanced Engineering Informatics* 16(3) (2002), S. 193–203.
- [43] CORDÓN, O.; GOMIDE, F.; HERRERA, F.; HOFFMANN, F.; MAGDALENA, L.: Ten Years of Genetic Fuzzy Systems: Current Framework and New Trends. *Fuzzy Sets and Systems* 141 (2004), S. 5–31.
- [44] CORDÓN, O.; HERRERA, F.; ZWIR, I.: Fuzzy modeling by hierarchically built fuzzy rule bases. *International Journal of Approximate Reasoning* 27(1) (2001), S. 61–93.
- [45] CORDÓN, O.; DEL JESUS, M. J.; HERRERA, F.: Genetic Learning of Fuzzy Rule-Based Classification Systems Cooperating with Fuzzy Reasoning Methods. Techn. Ber. DECSAI-970130, Universidad de Granada, Dept. of Computer Science and Artificial Intelligence, 1997.
- [46] DELGADO, M.; GÓMEZ-SKARMETA, A. F.; MARTÍN, F.: A Methodology to Model Fuzzy Systems Using Fuzzy Clustering in a Rapid-Prototyping Approach. *Fuzzy Sets and Systems* 97 (1998), S. 287–301.
- [47] DEMPSTER, A.: Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics* 38 (1967), S. 325–339.
- [48] DENANT-BOËMONT, L.; PETIOT, R.: Information value and sequential decision-making in a transport setting: an experimental study. *Transportation Research Part B: Methodological* 37(4) (2003), S. 365–386.
- [49] DIN 19226: *Leittechnik, Regelungstechnik und Steuerungstechnik*. Deutsches Institut für Normung e.V., 1994.
- [50] DIN 55350: *Teil 12: Begriffe der Qualitätssicherung und Statistik; Merkmalsbezogene Begriffe*. Deutsches Institut für Normung e.V., 1989.
- [51] DOMINGOS, P.: MetaCost: A General Method for Making Classifiers Cost-Sensitive. In: *Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, S. 155–164, ACM Press, 1999.
- [52] DRUMMOND, C.; HOLTE, R.: Explicitly Representing Expected Cost: An Alternative to ROC Representation. In: *Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2000.
- [53] DUBOIS, D.; PRADE, H.: An Introduction to Fuzzy Systems. *Clinica Chimica Acta* 270(1) (1998), S. 3–29.
- [54] EISENFÜHR, F.; WEBER, M.: *Rationales Entscheiden*. Springer Verlag, 1993.

- [55] ELKAN, C.: The Foundations of Cost-Sensitive Learning. In: *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence (IJCAI01)*, 2001.
- [56] FANDEL, G.: *Produktion I: Produktions- und Kostentheorie*. Springer Verlag, 1996.
- [57] FERGUSON, S. T.: *Mathematical Statistics - A Decision Theoretic Approach*. Academic Press, 1967.
- [58] FÖLLINGER, O.: *Regelungstechnik - Einführung in die Methoden und ihre Anwendung*. Hüthig Buch Verlag Heidelberg, 1994.
- [59] FRANK, M.: Diagnoseverfahren in der Automatisierungstechnik. *at-Automatisierungstechnik* 2/94 (1994), S. 47–64.
- [60] FRITSCH, M.: *Baumorientierte Regel-Induktionsstrategie für das ROSA-Verfahren zur Modellierung komplexer dynamischer Systeme*. Fortschritt-Bericht VDI, Reihe 8, Nr. 565, Düsseldorf: VDI-Verlag, 1996.
- [61] FUKAYA, N.; TOYAMA, S.; ASFOUR, T.; DILLMANN, R.: Design of the TUAT/Karlsruhe Humanoid Hand. *IEEE/RSJ ICIRS* (2000).
- [62] GERKE, W.: *Handwörterbuch des Bank- und Finanzwesens*. Schäffer-Poeschel, 2001.
- [63] GREINER, R.; GROVE, A.; ROTH, D.: Learning cost-sensitive active classifiers. *Artificial Intelligence* 139 (2002), S. 137–174.
- [64] GUAN, J.; BELL, D.: *Evidence Theory and its Applications*. North Holland Amsterdam, 1991.
- [65] HAENDEL, L.: Benchmark: Qualitätskontrolle bei KFZ-Aggregaten. In: *Proc. 12. Workshop Fuzzy Systeme* (MIKUT, R.; REISCHL, M., Hg.), S. 119–126, Forschungszentrum Karlsruhe Wissenschaftliche Berichte FZKA 6767, 2002.
- [66] HOFFMANN, F.: Combining boosting and evolutionary algorithms for learning of fuzzy classification rules. *Fuzzy Sets and Systems* 141(1) (2004), S. 47–58.
- [67] HOLLAND, J. H.: *Adaptation in Natural and Artificial Systems*. The University of Michigan Press, 1975.
- [68] HOLLAND, J. H.; HOLYOAK, K. J.; NISBETT, K. J.; THAGARD, P. R.: *Induction: Processes of Inference*. Cambridge, MA: MIT Press, 1986.
- [69] HOLLMÉN, J.; SKUBACZ, M.; TANIGUCHI, M.: Input Dependent Misclassification Costs for Cost-Sensitive Classifiers. In: *Data Mining II- Proceedings of the Second International Conference on Data Mining 2000* (EBECKEN, N.; BREBBIA, C., Hg.), S. 495–503, WIT Press, 2000.
- [70] HÖPPNER, F.; KLAWONN, F.; KRUSE, R.; RUNKLER, T.: *Fuzzy cluster analysis*. Wiley, 1999.
- [71] HUYGENS, C.: *Ratiociniis in Aleae Ludo*. 1647.

- [72] ISERMANN, R.: *Identifikation dynamischer Systeme 1*. Springer-Verlag, 1992.
- [73] ISERMANN, R.; SCHWARTE, A.; KIMMICH, F.: Model-based fault detection of a Diesel engine with turbo charger - a case study. In: *1st IFAC Symposium on Advances in Automotive Control*, University of Salerno, 2004.
- [74] JANIKOW, C. Z.: Fuzzy Decision Trees: Issues and Methods. *IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics* 28(1) (1998), S. 1–14.
- [75] KAFTANDJIAN, V.; DUPUIS, O.; BABOT, D.; ZHU, Y. M.: Uncertainty Modelling Using Dempster-Shafer Theory for Improving Detection of Weld Defects. *Pattern Recognition Letters* 24 (2003), S. 547–564.
- [76] KAMBEROVA, G.; MANDELBAUM, R.; MINTZ, M.; BAJCSY, R.: Decision-Theoretic Approach to Robust Fusion of Location Data. *Journal of the Franklin Institute* 336 (1999), S. 269–284.
- [77] KEENEY, R.; RAIFFA, H.: *Decisions with Multiple Objective*. John Wiley & Sons, Inc., 1976.
- [78] KERPA, O.; WEISS, K.; WÖRN, H.: Development of a Flexible Tactile Sensor System for a Humanoid Robot. In: *Proceedings IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2003.
- [79] KIENDL, H.: *Fuzzy Control methodenorientiert*. R. Oldenbourg Verlag München, 1997.
- [80] KLAWONN, F.; NAUCK, D.; KRUSE, R.: Generating Rules from Data by Fuzzy and Neuro-Fuzzy Methods. In: *Proc. Fuzzy-Neuro-Systeme '95, Darmstadt*, S. 223–230, 1995.
- [81] KLOSE, A.; NÜRNBERGER, A.: Applying Boolean Transformations to Fuzzy Rule Bases. In: *Proc. 7th Europ. Congr. on Intelligent Techniques and Soft Computing EUFIT'99, Sep. 13-16, 1999*, S. 215–217, Aachen, cD-ROM, 1999.
- [82] KOCH, M.; KUHN, T.; WERNSTEDT, J.: *Fuzzy Control - Optimale Nachbildung und Entwurf optimaler Entscheidungen*. R. Oldenbourg Verlag München, 1996.
- [83] KOFLER, E.; MENGES, G.: *Entscheidungen bei Unvollständiger Information*. Springer Verlag Berlin Heidelberg New York, 1976.
- [84] KRELLE, W.: *Preistheorie*. Mohr, Tübingen, 1961.
- [85] KRONE, A.; KIENDL, H.: An evolutionary concept for generating relevant fuzzy rules from data. *Intern. J. Knowledge-based Intelligent Engineering Systems* 1 (1997) 4, S. 207–213.
- [86] KRUSE, R.; KLAWONN, F.; HÖPPNER, F.: *Fuzzy Cluster Analysis*. Wiley & Sons, 1999.
- [87] KRUSE, R.; SCHECKE, E.; HEINSOHN, J.: *Uncertainty and Vagueness in Knowledge Based Systems*. Springer Verlag, 1991.

- [88] LAUX, H.: *Entscheidungstheorie*. Springer-Verlag, 2003.
- [89] LEE, M.-L.; CHUNG, H.-Y.; YU, F.-M.: Modeling of hierarchical fuzzy systems. *Fuzzy Sets and Systems* 138(2) (2003), S. 343–361.
- [90] LEHMANN, A.; MIKUT, R.; MARTIN, J.; BRETTHAUER, G.: Online-Stabilitätsüberwachung strukturvariabler Roboterregelungen. In: *Robotik 2004 VDI Berichte 1841*, S. 55–62, VDI Verlag, 2004.
- [91] LIN, C.; WEN, U.: A Labeling Algorithm for the Fuzzy Assignment Problem. *Fuzzy Sets and Systems* 142(3) (2004), S. 373–391.
- [92] LOOSE, T.: *Modellgestützte Diagnostik mittels Data Mining am Beispiel der Bewegungsanalyse*. Universitätsverlag Karlsruhe, 2004.
- [93] LOVCHIK, C. S.; DIFTLER, M. A.: The Robonaut Hand: A Dexterous Robot Hand for Space. In: *Proceedings of the IEEE ICRA*, S. 907–913, Detroit, 1999.
- [94] LU, Y.; CHEN, T.; HAMILTON, B.: A Fuzzy System for Automotive Fault Diagnosis: Fast Rule Generation and Self-Tuning. *IEEE Transactions on Vehicular Technology* (49)2 (2000), S. 651–660.
- [95] MAG, W.: *Entscheidung und Information*. Verlag Franz Vahlen München, 1977.
- [96] MARTIN, J.: *Ein Beitrag zur Integration von Sensoren in eine anthropomorphe künstliche Hand mit flexiblen Fluidaktoren*. Dissertation, Universität Karlsruhe (TH), (Im Druck), 2004.
- [97] MARTIN, J.; BECK, S.; LEHMANN, A.; MIKUT, R.; PYLATIUK, C.; SCHULZ, S.; BRETTHAUER, G.: Sensors, Identification, and Low Level Control of a Flexible Anthropomorphic Robot Hand. *International Journal of Humanoid Robotics* 1(3) (2004), S. 517–532.
- [98] MICHIE, D.; SPIEGELHALTER, D.; TAYLER, C.: *Machine Learning, Neural and Statistical Classification*. Ellis Horwood, 1994.
- [99] MIKUT, R.: *Modellgestützte on-line Stabilitätsüberwachung komplexer Systeme auf der Basis unscharfer Ljapunov-Funktionen*. VDI Verlag, 1999.
- [100] MIKUT, R.: Fuzzy-Modellbildung für den Benchmarkdatensatz Kfz-Aggregate. In: *Proc. 12. Workshop Fuzzy Systeme* (MIKUT, R.; REISCHL, M., Hg.), S. 127–134, Forschungszentrum Karlsruhe Wissenschaftliche Berichte FZKA 6767, 2002.
- [101] MIKUT, R.: KAFKA - Modellbildung und Statistik. Techn. Ber., Institut für Angewandte Informatik, Forschungszentrum Karlsruhe, 2002.
- [102] MIKUT, R.; GRÖLL, L.; JÄKEL, J.: Boolesche Algebra und Fuzzy-Operatoren. In: *Proceedings 11. Workshop Fuzzy-Control des GMA-FA 5.22* (REISCHL, M.; MIKUT, R., Hg.), S. 58–72, Forschungszentrum Karlsruhe Wissenschaftliche Berichte FZKA 6660, 2001.

- [103] MIKUT, R.; JÄKEL, J.; GRÖLL, L.: Automatic Design of Interpretable Membership Functions. In: *Proc. 8th Zittau Fuzzy Colloquium, Sep. 6-8, 2000*, S. 103–111, Hochschule Zittau/Görlitz, 2000.
- [104] MIKUT, R.; JÄKEL, J.; GRÖLL, L.: Inference Methods for Partially Redundant Rule Bases. In: *Fuzzy Control: Theory and Practice* (HAMPEL, R.; WAGENKNECHT, M.; CHAKER, N., Hg.), Advances in Soft Computing, S. 177–185, Heidelberg: Physica, 2000.
- [105] MIKUT, R.; JÄKEL, J.; GRÖLL, L.: Interpretability issues in data-based learning of fuzzy systems. *Artikel im Druck: Fuzzy Sets and Systems* (2004).
- [106] MIKUT, R.; JÄKEL, J.; MALBERG, H.; BRETTHAUER, G.: Datenbasierter Entwurf von Fuzzy-Systemen für medizinische Diagnoseaufgaben. *Automatisierungstechnik (at)* 48 (2000) 7, S. 317–326.
- [107] MOUNIER, S.: *Entwicklung einer realitätsnahen Kraftrückkopplung bei fluidisch betriebenen Handprothesen*. Forschungszentrum Karlsruhe Wissenschaftliche Berichte FZKA 7004, 2004.
- [108] NAG, P. K.; GATTIKER, A.; WEI, S.; BLANTON, R. D.; MALY, W.: Modeling the Economics of Testing: A DFT Perspective. *IEEE Design & Test of Computers* 19(1) (2002), S. 29–41.
- [109] NEUMANN, J.; MORGENSTERN, O.: *Theory of Games and Economic Behavior*. Princeton University Press, 1953.
- [110] NEUMANN, U.: *Integrierte Instandhaltungsplanung für elektrische Energieübertragungssysteme*. VDI Verlag GmbH, 1998.
- [111] NYBERG, M.: Automatic Design of Diagnosis Systems with Application to an Automotive Engine. *Control Engineering Practice* (7)8 (1999), S. 993–1005.
- [112] NYBERG, M.; STUTTE, T.: Model Based Diagnosis of the Air Path of an Automotive Diesel Engine. *Control Engineering Practice* (2003 online).
- [113] OSSWALD, D.; MARTIN, J.; BURGHART, C.; MIKUT, R.; WÖRN, H.; BRETTHAUER, G.: Integrating a Flexible Anthropomorphic Robot Hand into the Control System of a Humanoid Robot. *Robotics and Autonomous Systems* 48(4) (2004), S. 213–221.
- [114] PAČLÍK, P.: On Feature Selection with Measurement Cost and Grouped Features. In: *Proceedings of SPR2002 Workshop*, Windsor, Canada, <http://www.ph.tn.tudelft.nl/pavel/files/paclik02:SPR.pdf>, 2002.
- [115] PETIT-RENAUD, S.; DENÈUX, T.: Nonparametric Regression Analysis of Uncertain and Imprecise Data Using Belief Functions. *International Journal of Approximate Reasoning* 35(1) (2004), S. 1–28.
- [116] PRASSER, K.: Unscharfes Schließen in wissensbasierten Systemen auf Basis der Dempster-Shafer-Evidenztheorie. *Informationstechnik-IT* 34(2) (1992), S. 124–131.

- [117] PRIBER, U.: Ergebnisse der Untersuchung von Benchmarkdaten mittels FUCS - Fuzzy Classification System. In: *Proc. 11. Workshop Fuzzy Control des GMA-FA 5.22* (MIKUT, R.; REISCHL, M., Hg.), S. 100–105, Forschungszentrum Karlsruhe Wissenschaftliche Berichte FZKA 6660, 2001.
- [118] PRIBER, U.; KRETZSCHMAR, W.: Inspection and Supervision by Means of Hierarchical Fuzzy Classifiers. *Fuzzy Sets and Systems* 85 (1997), S. 263–274.
- [119] PROVOST, F.; FAWCETT, T.: Analysis and Visualization of Classifier Performance: Comparison under Imprecise Class and Cost Distributions. In: *Proceedings of the Third International Conference on Knowledge Discovery and Data Mining (KDD-97)*, American Association for Artificial Intelligence (www.aaai.org), 1997.
- [120] QUINLAN, J. R.: *C4.5: Programs for Machine Learning*. San Mateo, CA: Morgan Kaufmann, 1993.
- [121] RAGHAVAN, V.; SHAKERI, M.; PATTIPATI, K. R.: Test Sequencing Problems Arising in Test Planning and Design for Testability. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* 29(2) (1999), S. 153–163.
- [122] RAIFFA, H.: *Einführung in die Entscheidungstheorie*. R. Oldenbourg Verlag München Wien, 1973.
- [123] REISCHL, M.; MIKUT, R.; PYLATIUK, C.; SCHULZ, S.; BECK, S.; BRETTHAUER, G.: Steuerungs- und Signalverarbeitungskonzepte für eine multifunktionale Handprothese. *at-Automatisierungstechnik* 50(6) (2002), S. 279–286.
- [124] REUCHER, E.: *Modellbildung bei Sicherheit und Ungewißheit in konditionalen Strukturen*. Logos Verlag Berlin, 2002.
- [125] RIVES, J.: FID3: Fuzzy Induction Decision Tree. In: *Proc. 1st Int. Symp. Uncertainty, Modelling and Analysis*, S. 457–462, Los Alamitos, Calif.: IEEE Computer Soc. Press, 1990.
- [126] ROMMELFANGER, H.: *Fuzzy Decision Support-Systeme*. Springer-Verlag, 1993.
- [127] ROUBOS, J.; SETNES, M.; ABONYI, J.: Learning Fuzzy Classification Rules from Labeled Data. *Information Sciences* 150 (2003), S. 77–93.
- [128] SAVAGE, L. J.: *The foundations of statistics*. Wiley, 1954.
- [129] SBONER, A.; ECCHER, C.; BLANZIERI, E.; BAUER, P.; CRISTOFOLINI, M.; ZUMIANI, G.; FORTI, S.: A Multiple Classifier System for Early Melanoma Diagnosis. *Artificial Intelligence in Medicine* 27 (2003), S. 29–44.
- [130] SCHIFFERS, J.: A Classification Approach Incorporating Misclassification Costs. *Intelligent Data Analysis* 1(1-4) (1997), S. 59–68.
- [131] SCHNEIDER, I. H.: *Die Entwicklung der Wahrscheinlichkeitstheorie von den Anfängen bis 1933*. Darmstadt: Wissenschaftliche Buchgesellschaft, 1988.

- [132] SCHULZ, S.: *Konzept einer neuen Adaptiv-Hand-Prothese auf der Basis flexibler Fluidaktoren*. Dissertation, Universität Karlsruhe (TH), 2003.
- [133] SCHULZ, S.: *Eine neue Adaptiv-Hand-Prothese auf der Basis flexibler Fluidaktoren*. Aachen: Shaker, 2004.
- [134] SCHULZ, S.; PYLATIUK, C.; BRETTHAUER, G.: A New Class of Flexible Fluidic Actuators and their Applications in Medical Engineering. *at-Automatisierungstechnik* 47(8) (1999), S. 390ff.
- [135] SCHULZ, S.; PYLATIUK, C.; BRETTHAUER, G.: A new ultralight anthropomorphic hand. In: *International Conference on Robotics and Automation, Seoul, Korea*, 2001.
- [136] SCHULZ, S.; PYLATIUK, C.; MARTIN, J.; BRETTHAUER, G.: Die Anthropomorphe FZK-Hand. In: *Robotik 2002 Leistungsstand - Anwendungen - Visionen - Trends (VDI-Berichte 1679)* (DILLMANN, R.; SCHRAFT, R.; WÖRN, H., Hg.), S. 531–536, VDI/VDE- Gesellschaft Mess- und Automatisierungstechnik, 2002.
- [137] SEARA, J.; STROBL, K.; MARTÍN, E.; SCHMIDT, G.: Task-Oriented and Situation-Dependent Gaze Control for Vision Guided Humanoid Walking. In: *Proceedings IEEE Humanoids 2003*, IEEE, 2003.
- [138] SHAFER, G.: *A mathematical theory of evidence*. Princeton University Press, 1976.
- [139] SHAYLER, P.; GOOGMAN, M.; MA, T.: The Exploitation of Neural Networks in Automotive Engine Management Systems. *Engineering Applications of Artificial Intelligence* 13 (2000), S. 147–157.
- [140] SLAWINSKI, T.: *Analyse und effiziente Generierung von relevanten Fuzzy-Regeln in hochdimensionalen Suchräumen*. VDI Verlag Düsseldorf, Fortschrittsberichte Reihe 10, Nr. 686, 2001.
- [141] SMARANDACHE, F.; DEZERT, J. (Hg.): *Advancs and Applications of DSmT for Information Fusion*. American Research Press, 2004.
- [142] SMITH, S. F.: *A Learning System Based on Genetic Adaptive Systems*. Dissertation, Univ. of Pittsburgh, 1980.
- [143] SONNENBERG, A.: The End of Diagnosis: When to Stop Testing. *The American Journal of Gastroenterology* 97(10) (2002), S. 2494–2498.
- [144] SRIVASTAVA, R. P.; MOCK, T. (Hg.): *Belief Functions in Business Decisions*. Studies in Fuzziness and Soft Computing, Physica Verlag, 2002.
- [145] TURNEY, P.: Cost-Sensitive Classification: Empirical Evaluation of a Hybrid Genetic Decision Tree Induction Algorithm. *Journal of Artificial Intelligence Research* 2 (1995), S. 369–409.
- [146] TURNEY, P.: Types of Cost in Inductive Concept Learning. In: *Proc. Workshop on Cost-Sensitive Learning at the 17th International Conference on Machine Learning (WCSL at ICML-2000)*, S. 15–21, Stanford University, California, 2000.

- [147] UNBEHAUEN, H.: *Regelungstechnik I*. Verlag Vieweg, 2001.
- [148] VAPNIK, V. N.: *Statistical Learning Theory*. John Wiley & Sons, Inc., 1998.
- [149] VOORBRAAK, F.: On the justification of Dempster's rule of combination. *Artificial Intelligence* 48 (1991), S. 171–197.
- [150] WALLY, P.: *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, 1991.
- [151] WATSON, S. R.; WEISS, J. J.; DONNELL, M. L.: Fuzzy Decision Analysis. *IEEE Transactions on Systems, Man and Cybernetics* 9 (1979), S. 1–9.
- [152] WERNSTEDT, J.: *Experimentelle Prozeßanalyse*. R. Oldenbourg Verlag München Wien, 1989.
- [153] WÖHE, G.: *Einführung in die allgemeine Betriebswirtschaftslehre*. Verlag Vahlen, 2000.
- [154] YAGER, R. R.: Including Probabilistic Uncertainty in Fuzzy Logic Controller Modeling Using Dempster-Shafer Theory. *IEEE Transactions on Systems, Man, and Cybernetics* 25(8) (1995), S. 1221–1230.
- [155] YAGER, R. R.: On the Construction of Hierarchical Fuzzy Systems Models. *IEEE Transactions on Systems, Man, and Cybernetics - Part C: Applications and Reviews* 28(1) (1998), S. 55–66.
- [156] YEH, Z.-M.; LI, K.-H.: A systematic approach for designing multistage fuzzy control systems. *Fuzzy Sets and Systems* 143(2) (2004), S. 251–273.
- [157] YEN, J.: Generalizing the Dempster-Shafer Theory to Fuzzy Sets. *IEEE Transactions on Systems, Man, and Cybernetics* 20(3) (1990), S. 559–570.
- [158] YOSHINARI, Y.; PEDRYCZ, W.; HIROTO, K.: Construction of Fuzzy Models Through Clustering Techniques. *Fuzzy Sets and Systems* 54 (1993) 3, S. 157–165.
- [159] ZADEH, L.: Fuzzy Sets. *Information and Control* 8 (1965), S. 338–353.
- [160] ZADEH, L.: Probability Measures of Fuzzy Events. *Journal of Mathematical Analysis and Applications* 23 (1968), S. 421–427.
- [161] ZADEH, L.: *Fuzzy Sets and Information Granularity*, S. 3–18. North-Holland Publishing Company, 1979.
- [162] ZADEH, L.: A Simple View of the Dempster-Shafer Theory of Evidence and its Implication for the Rule of Combination. *The AI Magazine* 7(2) (1986), S. 85–90.
- [163] ZIMMERMANN, H.-J.: *Methoden und Modelle des Operations Research*. Vieweg Verlag, Braunschweig, 1992.

Bildverzeichnis

1.1	Basiselemente eines Entscheidungsmodells.	3
1.2	Sicherheit, Unsicherheit, Ungewissheit und Risiko. Nach [88] (erweitert).	4
1.3	Kennzahlen beim Glücksspiel	7
1.4	Zusammenhang Massefunktion, Glaubens- und Plausibilitätsgrad.	9
1.5	Datenbasierter überwachter Entwurf von Entscheidungssystemen.	12
1.6	Grundprinzip eines Fuzzy-Systems aus [99].	14
2.1	Beispiel einer Entscheidungsfunktion in \mathbb{R}_2	18
2.2	Gemischte Strategien an einem Beispiel.	23
2.3	Einteilung von Merkmalen vereinfacht nach [50].	29
2.4	Verlauf fixer und variabler Merkmalskosten.	30
2.5	Visualisierung der Linien, an denen die Entscheidungen kippen.	33
2.6	Beispiel: Zugehörigkeitsfunktionen des Merkmals x_l (fünf Terme).	36
2.7	Regelbasis für den Beispieldatensatz.	37
2.8	Verteilungsdichtefunktion und Diskretisierung.	40
2.9	Beispiel für die ODER-Verknüpfung benachbarter Terme.	41
2.10	Beispielproblem mit zwei Umweltzuständen und zwei Prämissen.	42
2.11	Zusammenhang Entscheidungen und Kostenverhältnis.	43
2.12	Überlappung von Regeln.	45
2.13	Visualisierung Dempster's rule of combination.	48
2.14	Beispiel für Massefunktionen.	49
2.15	Auswertung einer Regelbasis über die Evidenz-Inferenz.	51
2.16	Darstellung und Bewertung von Regeln.	54
2.17	Darstellung der Regelbewertung mit Höhenlinien (nur Prämisse).	56
2.18	Bewertung mit den Beispielen der Prämisse und den Fehlerkandidaten.	57
2.19	Mögliche Aufteilungen der Restprämisse.	59
2.20	Darstellung der Bewertung der Regeln aus Tabelle 2.10.	62
2.21	Unvollständige Regelbasis für den Beispieldatensatz.	66
3.1	Ausschnitt aus einem Entscheidungsbaum.	69
3.2	Regeln des Baumes in x_1, x_2	71
3.3	Darstellung der Hypothesen aus Tabelle 3.2. $C_i = e_1$	73
3.4	Darstellung der Hypothesen aus Tabelle 3.2 im Bewertungsgraphen.	74
3.5	Alle Pruningschritte für Regel 5 aus Bild 3.2.	75
3.6	Darstellung der Dominanzsituationen beim Pruning von Einzelregeln.	77
3.7	Pruningergebnis von Regel 5 (Merkmalskosten).	78

3.8	Suche einer Regelbasis mit fester Default-Entscheidung.	80
3.9	Suche einer Regelbasis mit automatischer Default-Entscheidung.	81
3.10	Visualisierung der zweigeteilten Bewertung.	82
3.11	Suche einer Regelbasis mit Rückweisungsentscheidung.	83
3.12	Ergebnisse bei Berücksichtigung von Merkmalskosten.	84
3.13	Suche einer Regelbasis mit Evidenz-Inferenz.	85
3.14	Darstellung eines Genotypen und dem daraus abgeleiteten Phänotyp.	88
3.15	Ergebnis Regelbasissuche Genetischer Algorithmus Evidenz-Inferenz.	90
3.16	Beispielproblem für die Optimierung der Zugehörigkeitsfunktionen.	91
3.17	Optimierung einer Zugehörigkeitsfunktion.	92
3.18	Zu optimierende Terme.	93
3.19	Ausschnitt aus dem Gütegebirge bei der ZGF-Optimierung.	94
3.20	Schematisches Beispiel: Motivation hierarchische Regelbasis (1).	96
3.21	Schematisches Beispiel: Motivation hierarchische Regelbasis (2).	97
3.22	Ablauf einer dreistufigen Entscheidung.	98
3.23	Hypothesenbildung bei der Permutationsstrategie.	99
3.24	Permutationsstrategie: Algorithmus.	100
3.25	Beispiel der Alpha-Strategie.	102
3.26	Hierarchische Regelbasis.	103
3.27	Benutzerschnittstelle von KAFKA.	104
3.28	Menüpunkt von KAFKA.	105
4.1	Schema der Kreuzvalidierung.	109
4.2	Heart-Disease Datensatz.	111
4.3	Variation von α_{ER} und α_{RB}	115
4.4	Kostenverhältnis in Abhängigkeit von G	118
4.5	Kfz-Benchmark: Ergebnisse für verschiedene Kostenverhältnisse (1).	119
4.6	Kfz-Benchmark: Ergebnisse für verschiedene Kostenverhältnisse (2).	120
4.7	Aktorkennfeld Messdaten und Approximation.	123
4.8	Struktur der Momentenbegrenzung.	123
4.9	Gelenk mit Objekt und Sensor.	124
4.10	Teststand mit eingespanntem Gelenk und Objekt bei 90°	125
4.11	Daten für den Entwurf der Kontakterkennung.	126
4.12	Kontakterkennung: Ergebnisse ohne Merkmalskosten.	128
4.13	Kontakterkennung: Ergebnisse mit Merkmalskosten.	129
4.14	Kontakterkennung: Ergebnisse im Zeitverlauf.	130
6.1	Visualisierung von Dempster's Rule of combination.	155

Tabellenverzeichnis

1.1	Ergebnismatrix am Beispiel des Glücksspiels.	3
1.2	Ergebnismatrix am Beispiel des Glücksspiels mit MinMax-Regel.	6
1.3	Beispiel für Dempster's rule of combination.	9
2.1	Kostenmatrix für den 'German Credit' Datensatz aus [98].	20
2.2	Alternative Kostenmatrix für den 'German Credit' Datensatz.	20
2.3	Entscheidungskriterien und Entscheidungsfunktionen.	26
2.4	Einzelregeln zu Bild 2.7.	38
2.5	Regeln zu Bild 2.10.	43
2.6	Zahlenwerte zu Bild 2.13.	48
2.7	Regeln zu Bild 2.14.	50
2.8	Vergleich Fuzzy-Inferenz und Evidenz-Inferenz.	50
2.9	Varianten für die Berücksichtigung der Beispiele in der Restprämisse.	53
2.10	Einzelregeln zu Bild 2.16.	54
2.11	Bewertungen $\hat{L}_{D,P,r,Z}$ für die Prämissen aus Bild 2.18. $L_R = 10$	60
2.12	Regelbewertung gemäß der drei Kriterien.	63
2.13	Bewertung von Regelbasen.	65
3.1	Kosten und Auftrennung des Entscheidungsbaumes aus Bild 3.1.	70
3.2	Hypothesen eines Pruningschrittes.	73
3.3	Pruningergebnis von Regel 5 (Merkmalskosten).	77
3.4	Bewertung der vierten Regel bei Regelbasissuche mit Rückweisung.	84
3.5	Übersicht der Ergebnisse der Regelsuche.	86
3.6	Ergebnisse Optimierung der Zugehörigkeitsfunktionen.	92
3.7	Optimierung der Zugehörigkeitsfunktionen für den Beispieldatensatz.	93
3.8	Optionen des Algorithmus für die Erstellung von Fuzzy-Systemen.	95
3.9	Kosten der hierarchischen Regelbasis.	103
4.1	Zusammenstellung der verwendeten Abkürzungen.	108
4.2	Ergebnisse German-Credit Data Set.	110
4.3	Ergebnisse des Heart-Disease Datensatzes, Fuzzy-Inferenz.	113
4.4	Ergebnisse des Heart-Disease Datensatzes, Evidenz-Inferenz.	114
4.5	Auswahl der Ergebnisse für den Kfz-Benchmarkdatensatz (1).	121
4.6	Auswahl der Ergebnisse für den Kfz-Benchmarkdatensatz (2).	121
4.7	Ergebnisse für den Kfz-Benchmarkdatensatz aus [100].	121
4.8	Regeln für den FSR Sensor (F2).	127

4.9	Regelbasis Kontakterkennung Merkmalskosten $L_R = 3$	130
4.10	Zugehörigkeitsfunktionen Kontakterkennung (2).	131
4.11	Regelbasis Kontakterkennung, unvollständiger Datensatz.	131
4.12	Zugehörigkeitsfunktionen Kontakterkennung (2).	131
6.1	Zahlenwerte zu Bild 6.1.	156

6 Anhang

6.1 Grundlagen der Evidenztheorie

In diesem Abschnitt werden knapp die Grundlagen der Evidenztheorie und Dempster's rule of combination wiedergegeben. Eine ausführliche Darstellung ist in [138] zu finden.

Betrachtet wird stets eine endliche Menge Θ an Ereignissen A_i . Dann ist 2^Θ die Potenzmenge von Θ . Die Potenzmenge entspricht der Menge aller Teilmengen von Θ .

Als Gegenstück zur Verteilungsfunktion der Wahrscheinlichkeitstheorie definiert die Evidenztheorie die Glaubensfunktion *belief function* bel :

$$bel(\emptyset) = 0 \quad (6.1)$$

$$bel(A_i) \geq 0 \quad (6.2)$$

$$bel(\Theta) = 1 \quad (6.3)$$

$$bel(A_1 \cup A_2 \cup \dots \cup A_n) \geq \sum_{I \subseteq \{1,2,\dots,n\}, I \neq \emptyset} (-1)^{|I|+1} bel(\cap_{i \in I} A_i) \quad (6.4)$$

Das Gegenstück zur Verteilungsdichtefunktion der Wahrscheinlichkeitstheorie ist die Massefunktion der Evidenztheorie: Eine Funktion über der Potenzmenge von Θ $m : 2^\Theta \rightarrow [0, 1]$ ist eine Massefunktion, wenn gilt:

$$m(\emptyset) = 0 \quad (6.5)$$

$$\sum_{X \subseteq \Theta} m(X) = 1 \quad (6.6)$$

Jede Teilmenge A von Θ wird fokales Element genannt, wenn $m(A) > 0$ gilt. Die Vereinigung aller fokalen Elemente ist der Kern der Masse- beziehungsweise Glaubensfunktion.

Die Glaubensfunktion bel eines Ereignisses einer beliebigen Teilmenge von Θ ergibt sich aus der Massefunktion:

$$bel(X) = \sum_{B \subseteq X} m(B). \quad (6.7)$$

Der Vergleich mit der Verteilungsdichte der Wahrscheinlichkeitstheorie zeigt, dass in der Definition der Massefunktion und der Glaubensfunktion der formale Unterschied zwischen den beiden Theorien begründet ist. Die Evidenztheorie lässt es im Gegensatz zur Wahrscheinlichkeitstheorie zu, dass den Elementen a_1 und a_2 einer Ereignismenge Θ die Masse $m(a_1) = m(a_2) = 0$, der Menge $a_1 \cup a_2$ jedoch eine Masse $m(a_1 \cup a_2) > 0$ zugewiesen wird. Dadurch wird es ermöglicht, Ungewissheit über das Eintreten der beiden Ereignisse a_1 oder a_2 zu modellieren, indem den Ereignissen selbst eine Masse von Null und der Vereinigung eine Masse größer Null zugewiesen wird.

Daraus folgt, dass eine Massefunktion dann in eine Verteilungsdichtefunktion übergeht, wenn die Massen lediglich auf die Elemente A_i der Menge Θ mit $A_i \in \Theta$ und nicht auf Teilmengen A_j von Θ mit $A_j \subseteq \Theta$ verteilt ist:

$$m(X) = 0 \text{ für } |X| > 1 \quad (6.8)$$

Dabei ist $|X|$ die Kardinalität von X . Dann gilt für eine Funktion:

$$d(x) = m(\{x\}) \text{ für } x \in \Theta \quad (6.9)$$

$$\sum_{x \in \Theta} d(x) = \sum_{x \in \Theta} m(\{x\}) = 1 \quad (6.10)$$

Ein wesentlicher Unterschied zwischen Wahrscheinlichkeit und Evidenz liegt darin, dass die Wahrscheinlichkeit auf das Eintreten eines Ereignisses eines Zufallsexperimentes abstellt und die Evidenz auf den Glaubensgrad, dass ein Ereignis eintritt. Die Bayessche Theorie verwendet ebenfalls den Glaubensgrad. Allerdings wird gefordert, dass die Kolmogorowschen Axiome eingehalten werden.

Im Gegensatz zur Evidenztheorie erlaubt es die Bayessche Theorie nicht, dass ein Ereignis, dem ein Glaubensgrad ungleich Null zugeordnet wird, eine Kombination aus Ereignissen ist, denen selbst ein Glaubensgrad von Null zugewiesen wurde. Dadurch wird es beispielsweise möglich, vollständige Unwissenheit auszudrücken. Eine Glaubensfunktion die vollständige Unwissenheit darstellt ist:

$$bel(A) = \begin{cases} 0 & \text{für } A \neq \Theta \\ 1 & \text{für } A = \Theta \end{cases} \quad (6.11)$$

Dadurch wird allen möglichen Ereignissen als Gesamtheit der Glaubensgrad Eins zugewiesen. Dies bedeutet, dass eines der Ereignisse aus Θ eintreten wird. Welches das sein wird, ist allerdings unbekannt, da jedem einzelnen Element ein Glaubensgrad von Null zugewiesen ist.

Die Wahrscheinlichkeitstheorie ist zu einer solchen Darstellung nicht in der Lage. Unwissenheit ist hier über eine Gleichverteilung der Wahrscheinlichkeit über alle Ereignisse darstellbar.

Als Beispiel werden zwei Urnen mit jeweils 100 schwarzen und roten Kugeln angenommen. Urne 1: bekannte Verteilung der Farben schwarz und rot: 50:50. Daraus folgt eine Wahrscheinlichkeit von 0.5 für das Eintreten der beiden Ereignisse. Hier handelt es sich um sicheres Wissen über ein Zufallsexperiment. Urne 2: Verteilung ist nicht bekannt. Bei Ungewissheit wird in der Wahrscheinlichkeitstheorie häufig die Annahme einer Gleichverteilung getroffen. Es ergeben sich also wieder Wahrscheinlichkeiten von jeweils 0.5 für beide Ereignisse.

Die Differenz zwischen Glaubensgrad und Plausibilität ist die Unsicherheit, der Informationen, die der Massefunktion zu Grunde liegen.

Dempster's rule of combination

Für die Verknüpfungen der Informationen zweier Massefunktionen m_1 und m_2 über der selben Potenzmenge, hat Dempster eine Kombinationsregel vorgeschlagen. Dempster's rule of combination ist bei unabhängigen Quellen anwendbar und ist wie folgt definiert [138]:

$$m(A) = m_1 \oplus m_2 = \frac{\sum_{\substack{i,j \\ A_i \cap B_j = A}} m_1(A_i) \cdot m_2(B_j)}{1 - \sum_{\substack{i,j \\ A_i \cap B_j = \emptyset}} m_1(A_i) \cdot m_2(B_j)} \quad (6.12)$$

Wird W als Maß für die Widersprüchlichkeit der beiden Massefunktionen eingeführt:

$$W = \sum_{\substack{i,j \\ A_i \cap B_j = \emptyset}} m_1(A_i) \cdot m_2(B_j), \quad (6.13)$$

dann vereinfacht sich (6.12) zu:

$$m(A) = \frac{\sum_{\substack{i,j \\ A_i \cap B_j = A}} m_1(A_i) \cdot m_2(B_j)}{1 - W} \quad (6.14)$$

Dempster's rule of combination ist kommutativ und assoziativ [138].

Abbildung 6.1 zeigt die Kombination zweier Massefunktionen über der Potenzmenge $2^\Theta = \{\emptyset, A_1, A_2, A_1 \cup A_2\}$. Die Verknüpfung der Massefunktionen m_1 und m_2 über Dempster's rule of combination wird auch durch das Symbol \oplus gekennzeichnet: $m = m_1 \oplus m_2$.

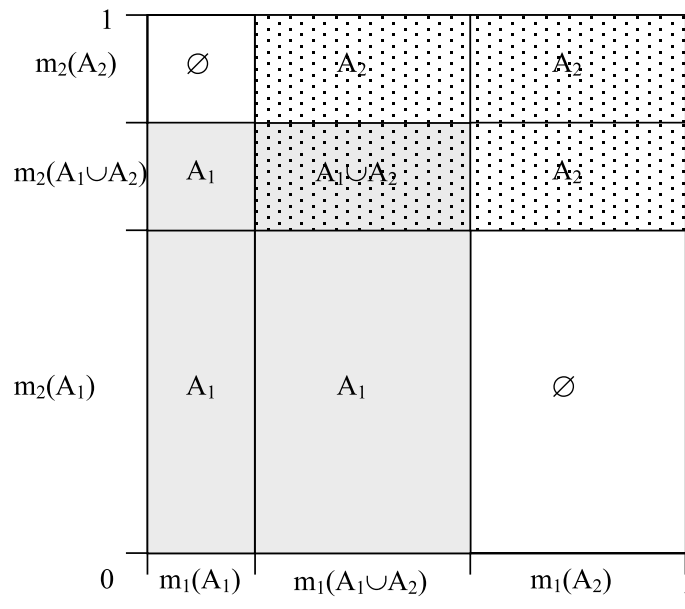


Bild 6.1: Visualisierung von Dempster's Rule of combination.

	A_1	A_2	$A_1 \cup A_2$	W
m_1	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{2}{9}$	—
m_2	$\frac{3}{9}$	$\frac{4}{9}$	$\frac{4}{9}$	—
$m_1 \oplus m_2$	$\frac{4}{9}$	$\frac{6}{9}$	$\frac{6}{9}$	$\frac{7}{25}$

Tabelle 6.1: Zahlenwerte zu Bild 6.1.

Die geometrische Entsprechung von Dempster's rule of combination wird anhand von Bild 6.1 deutlich. Die beiden Massefunktionen m_1 und m_2 sind auf den Achsen aufgetragen. Da die Summe über beide Massefunktionen gleich 1 ist, hat die aufgespannte Fläche den Flächeninhalt 1. Die einzelnen Teilflächen werden mit der Schnittmenge der Elemente A_i markiert, auf die sich die beiden Teile der Massefunktionen beziehen. Der Zähler von (6.12) entspricht dann der Summe der Teilflächen, die mit der entsprechenden Teilmenge von Θ markiert sind. Damit die resultierende Massefunktion die Bedingung (6.5) erfüllt, findet eine Re-Normierung statt. Dazu werden alle Teilflächen aufsummiert, denen die leere Menge \emptyset zugewiesen ist. In Tabelle 6.1 sind die Zahlenwerte einer Beispielrechnung zusammengestellt.

6.2 Datensätze

6.2.1 Beispieldatensatz

Der Datensatz besteht aus:

1. 60 Beispielen des Umweltzustandes 'z₁' und
2. 300 Beispielen des Umweltzustandes 'z₂'.

Für den Entwurf der Entscheidungssysteme wird die folgende Kostenmatrix gewählt:

$$\mathbf{L} = \begin{pmatrix} 0 & 10 \\ 1 & 0 \end{pmatrix}.$$

Jede andere Kostenmatrix ist ebenfalls verwendbar. Die resultierenden Regelbasen und die damit getroffenen Entscheidungen können sich dann ändern.

6.2.2 German-Credit

Der Datensatz besteht aus:

1. 700 Beispielen des Umweltzustandes 'guter Kreditkunde' (z_1) und
2. 300 Beispielen des Umweltzustandes 'schlechter Kreditkunde' (z_2).

Die möglichen Entscheidungen sind:

1. 'Kredit vergeben' (e_1) und
2. 'Kredit ablehnen' (e_2).

Für den Entwurf der Entscheidungssysteme wird die folgende Kostenmatrix verwendet [98]:

$$\mathbf{L} = \begin{pmatrix} 0 & 5 \\ 1 & 0 \end{pmatrix}.$$

Eine Begründung für die Wahl dieser Kostenmatrix wird in [98] nicht gegeben. Es ist jedoch nachvollziehbar, dass ein schlechter Kredit für die Bank höhere Kosten verursacht als ein abgelehnter 'guter Kreditkunde'. Um die Vergleichbarkeit der Ergebnisse mit denen aus [98] zu gewährleisten, wird die selbe Kostenmatrix verwendet.

Für jedes Beispiel stehen die folgenden Merkmale zur Verfügung:

- Merkmal 1 (qualitativ): Status des Girokontos
 1. $\dots < 0DM$
 2. $0 \leq \dots < 200DM$
 3. $\dots \geq 200DM$ / regelmäßiger Gehaltseingang seit mindestens einem Jahr
 4. kein Girokonto vorhanden
- Merkmal 2 (numerisch): Kreditlaufzeit in Monaten (4-72)
- Merkmal 3 (qualitativ): Bonität
 1. bisher keine Kredite / alle pünktlich zurück bezahlt
 2. alle Kredite bei dieser Bank pünktlich zurück bezahlt
 3. alle laufenden Kredite bisher pünktlich bedient
 4. Zahlungsrückstände in der Vergangenheit
 5. unsicheres Konto / weitere Kredite bei anderen Banken
- Merkmal 4 (qualitativ): Verwendung des Kredites
 1. Auto (neu)
 2. Auto (gebraucht)
 3. Möbel / Ausstattung
 4. Haushaltsgeräte
 5. Reparaturen
 6. Ausbildung
 7. Urlaub (kein Beispiel vorhanden)
 8. Umschulung
 9. Geschäft
 10. andere Zwecke
- Merkmal 5 (numerisch): Kreditsumme in DM (250-18.424)
- Merkmal 6 (qualitativ): Sparkonten
 1. $\dots < 100DM$
 2. $100 \leq \dots < 500DM$
 3. $500 \leq \dots < 1000DM$
 4. $\dots \geq 1000DM$
 5. unbekannt / nicht vorhanden

- Merkmal 7 (qualitativ): Status Beschäftigung
 1. arbeitslos
 2. ... < 1 Jahr
 3. $1 \leq \dots < 4$ Jahre
 4. $4 \leq \dots < 7$ Jahre
 5. ... ≥ 7 Jahre
- Merkmal 8 (numerisch): Kreditrate in Prozent des verfügbaren Einkommens
- Merkmal 9 (qualitativ): Geschlecht und Familienstand
 1. männlich, geschieden/getrennt
 2. weiblich, geschieden/getrennt/verheiratet
 3. männlich, single
 4. männlich, verheiratet/verwitwet
 5. weiblich, single
- Merkmal 10 (qualitativ): Weitere Schuldner / Bürgen
 1. keine
 2. weiterer Antragsteller
 3. Bürge
- Merkmal 11 (numerisch): aktueller Wohnsitz seit
- Merkmal 12 (qualitativ): Besitz
 1. Grundbesitz
 2. wenn nicht 1., Bausparvertrag/Lebensversicherung
 3. wenn nicht 1./2., Auto oder anderes nicht in Merkmal 6 gezählt
 4. unbekannt / kein Besitz
- Merkmal 13 (numerisch): Alter in Jahren (19-75)
- Merkmal 14 (qualitativ): Weitere Verbindlichkeiten gegenüber
 1. Bank
 2. Handel
 3. keine
- Merkmal 15 (qualitativ): Wohnung
 1. zur Miete
 2. Eigentum
 3. unentgeltlich
- Merkmal 16 (numerisch): Anzahl Kredite bei dieser Bank (1-4)
- Merkmal 17 (qualitativ): Beruf
 1. arbeitslos/ nicht ausgebildet - obdachlos
 2. nicht ausgebildet - nicht obdachlos

- 3. ausgebildet / öffentlicher Dienst
- 4. leitende Tätigkeit / selbständig / hochqualifiziert / Beamter
- Merkmal 18 (numerisch): Anzahl haftender Personen (1-2)
- Merkmal 19 (qualitativ): Telefon
 - 1. nein
 - 2. ja, unter dem Namen des Kunden
- Merkmal 20 (qualitativ): ausländischer Arbeiter
 - 1. ja
 - 2. nein
- Merkmal 21-24: Hilfsmerkmale für statistische Verfahren

6.2.3 Heart-Disease

Der Datensatz besteht aus:

1. 150 Beispielen des Umweltzustandes 'Herzkrankheit liegt nicht vor' (z_1) und
2. 120 Beispielen des Umweltzustandes 'Herzkrankheit liegt vor' (z_2).

Die möglichen Entscheidungen sind:

1. 'Diagnose positiv' (e_1) und
2. 'Diagnose negativ' (e_2).

Für den Entwurf der Entscheidungssysteme wird die folgende Kostenmatrix verwendet [98]:

$$\mathbf{L} = \begin{pmatrix} 0 & 5 \\ 1 & 0 \end{pmatrix} .$$

Eine Begründung für die Wahl dieser Kostenmatrix wird in [98] nicht gegeben. Es ist jedoch nachvollziehbar, dass eine versäumte positive Diagnose schlimmere Konsequenzen hat, als eine positive Diagnose bei einem tatsächlich gesunden Patienten. Um die Vergleichbarkeit der Ergebnisse mit denen aus [98] zu gewährleisten, wird die selbe Kostenmatrix um den Faktor 1000 multipliziert verwendet.

Folgende Merkmale stehen zur Verfügung:

Nr.	Beschreibung	Kosten
1	Alter in Jahren	1
2	Geschlecht	1
3	Art der Schmerzen	1
4	Blutdruck in Ruhe	1
5	Serum Cholesterin	7.27
6	Blutzucker	5.20
7	Elektrokardiogramm in Ruhe	15.50
8	Analyse der Herzfrequenzspitzen	102.90
9	Ergometrischer Belastungstest	87.30
10	ST-Strecken Senkung (nach ergometrischer Belastung)	87.30
11	ST-Strecken Anstieg (nach ergometrischer Belastung)	87.301
12	Angiographie	100.90
13	3=normal; 6=bleibende Abweichung 7=reversible Abweichung	102.90

6.2.4 Kfz-Aggregate

Der Lerndatensatz C besteht aus:

1. 10532 Beispielen des Umweltzustandes 'in Ordnung' (z_1) und
2. 81 Beispielen des Umweltzustandes 'nicht in Ordnung' (z_2).

Der Testdatensatz A besteht aus:

1. 2586 Beispielen des Umweltzustandes 'in Ordnung' und
2. 26 Beispielen des Umweltzustandes 'nicht in Ordnung'.

Der Testdatensatz B besteht aus:

1. 4328 Beispielen des Umweltzustandes 'in Ordnung' und
2. 45 Beispielen des Umweltzustandes 'nicht in Ordnung'.

Die möglichen Entscheidungen sind:

1. 'in Ordnung' (e_1) und
2. 'nicht in Ordnung' (e_2).

Für den Entwurf der Entscheidungssysteme wird die folgende Kostenmatrix verwendet:

$$\mathbf{L} = \begin{pmatrix} 0 & L_R \\ 1 & 0 \end{pmatrix}.$$

Das Kostenverhältnis L_R wird im Bereich $[1 \dots 40]$ variiert. Dadurch wird die Einschätzung widerspiegelt, dass die Entscheidung, ein Aggregat als in Ordnung einzustufen,

obwohl es einen Defekt hat, zu höheren Kosten führt, als die Entscheidung, ein Aggregat als Defekt einzustufen, obwohl es in Ordnung ist.

Aus den 304 aufgezeichneten Merkmalen werden über das MANOVA¹-Verfahren die 15 Merkmale mit der größten Relevanz bestimmt. Der Entwurf der Entscheidungssysteme erfolgt mittels dieser 15 Merkmale.

6.2.5 Kontakterkennung

Der Lerndatensatz (Kontakt A) besteht aus:

1. 10023 Beispielen des Umweltzustandes 'kein Kontakt' (z_1) und
2. 8732 Beispielen des Umweltzustandes 'Kontakt' (z_2).

Der Testdatensatz (Kontakt B) besteht aus:

1. 3186 Beispielen des Umweltzustandes 'kein Kontakt' und
2. 1369 Beispielen des Umweltzustandes 'Kontakt'.

Die möglichen Entscheidungen sind:

1. 'kein Kontakt' (e_1) und
2. 'Kontakt' (e_2).

Für den Entwurf der Entscheidungssysteme wird die folgende Kostenmatrix verwendet:

$$\mathbf{L} = \begin{pmatrix} 0 & L_R \\ 1 & 0 \end{pmatrix}.$$

Das Kostenverhältnis L_R wird im Bereich $[1 \dots 20]$ variiert. Dadurch wird die Einschätzung widergespiegelt, dass ein nicht erkannter Kontakt schlimmere Konsequenzen hat als ein falscher Alarm.

Die Merkmale für die Kontakterkennung sind:

Nr.	Beschreibung	Kosten
1	φ_{soll}	0
2	φ_{ist}	0
3	p_{soll}	0
4	e	0
5	M_{max}	0
6	C	0
7	\hat{M}_{Aktor}	0.01
8	p_{ist}	0
9	FSR 1	0.1
10	FSR 2	0.1
11	$p_{soll,limit}$	0
12	$p_{soll,unlimit}$	0
13	$\dot{\varphi}_{soll}$	0.005

¹Multivariate analysis of variance

6.3 Implementierung

Folgende Tabelle zeigt die wesentlichen Funktionen des Programmpaketes KAFKA, die neu erstellt bzw. erweitert wurden (N: neu erstellt, E: erweitert):

Funktionsname	Beschreibung	N/E	Zeilen
decision_opt_cost	Entscheidungstheoretische Bewertung	N	480
relemas3	Schätzung Wahrscheinlichkeiten	E	280
genbaum	Entscheidungsbaum	E	217
prun12NM	Pruning Ablauf	E	357
such_rb9	Regelbasissuche	E	360
rulebase_opt_ga	Regelbasissuche mittels Genetischem Algorithmus	N	190
opt_zgf_ent	Optimierung der Zugehörigkeitsfunktionen	N	141
finfer8	Fuzzy-Inferenz	E	209
evidenz_infer2	Evidenz-Inferenz	N	91
klass9	Auswertung und Darstellung der Ergebnisse	E	137
klass_hier_entwerfen2	Entwurf hierarchischer Regelbasen	N	180
klass_hier_anwenden	Auswertung hierarchischer Regelbasen	N	83
klastb18	Hauptprogramm	E	334

Lebenslauf

Name	Beck
Vornamen	Sebastian Mathias
Geburtstag	21. Februar 1976
Geburtsort	Pforzheim
Schulbildung	1982–1986 Grundschule im Arlinger, Pforzheim 1986–1995 Allgemeine Hochschulreife Reuchlin-Gymnasium, Pforzheim
Zivildienst	1995–1996 Arbeiter-Samariter-Bund, Pforzheim
Studium	1996–2001 Allgemeiner Maschinenbau, Universität Karlsruhe (TH)
Aufbaustudium	2002–2004 Wirtschaftsingenieurwesen, FernUniversität Hagen
Berufstätigkeit	2001–2004 Wissenschaftlicher Mitarbeiter, Institut für Angewandte Informatik, Forschungszentrum Karlsruhe GmbH