

**Dissertation**  
**Universität Karlsruhe**

# **Schließen mit unscharfen Begriffen**

---

**Martin Spott**

Martin Spott

**Schließen mit unscharfen Begriffen**

Dissertation, genehmigt von der Fakultät für Informatik  
der Universität Fridericiana zu Karlsruhe, 2000

Erster Gutachter: Prof. Dr. Gerhard Goos

Zweiter Gutachter: Prof. Dr. Rudolf Kruse

Tag der mündlichen Prüfung: 23. November 2000

Diese Ausarbeitung kann elektronisch von der Bibliothek der Universität Karlsruhe  
bezogen werden. Die Adresse lautet <http://www.ubka.uni-karlsruhe.de>.

---

# Kurzfassung

Die Bedeutung automatischer Informationsverarbeitung hat in den letzten Jahren kontinuierlich zugenommen und ein Abreißen dieses Trends ist nicht zu erwarten. Das Anwendungsspektrum reicht von der Regelung und dem Management technischer Systeme über entscheidungsunterstützende Systeme, z. B. für Betrugserkennung, bis hin zur Personalisierung von Dienstleistungen wie maßgeschneiderten Zeitungen im Internet.

Viele solcher Anwendungen verlangen nach Methoden der Informationsverarbeitung, die

1. effizient genug sind, um auf großen Datenmengen operieren zu können (Effizienz hinsichtlich Speicher- und Rechenaufwand)
2. die Ergebnisse der Verarbeitung verständlich darstellen
3. mit vagen und unsicheren Informationen umgehen können

Einen wichtigen Beitrag zu diesem Problemkomplex leisten Methoden unscharfer Informationsverarbeitung. Sie sichern in gewissem Umfang zumindest die letzten beiden Punkte. Für die Repräsentation von Information fassen diese Verfahren Einzelheiten zu größeren Informationseinheiten zusammen, die für Menschen im Allgemeinen leichter verständlich sind, weil sie von unnötigen Details abstrahieren. Größere Informationen sind dabei in vielen Fällen vage oder unsicher; entweder als inhärente Eigenschaft oder aufgrund eines kalkulierten Informationsverlustes, den wir in Kauf nehmen, um großen Datenmengen Herr zu werden. Modelliert werden grobe Informationen als unscharfe Mengen.

Die erste Forderung in obiger Liste verdeutlicht allerdings ein Dilemma üblicher Methoden unscharfer Informationsverarbeitung. Obwohl sie viele Einzelheiten auf wenige grobe Informationseinheiten reduzieren, hängen Speicher- und Rechenaufwand im Allgemeinen trotzdem von der Anzahl der Einzelheiten ab. Durch die Vergrößerung wurde also hinsichtlich des Verarbeitungsaufwandes nichts gewonnen. Zudem wird die Semantik unscharfer Mengen meist nur abstrakt definiert, weshalb die konkrete Bedeutung eines Zugehörigkeitsgrades zu einer unscharfen Menge unklar ist. Anwender wissen daher oft nicht, wie sie Zugehörigkeitsgrade festlegen sollen und insbesondere wird die gemeinsame Verarbeitung unscharfer Informationen unterschiedlicher Herkunft und Qualität zweifelhaft. Die vorliegende Arbeit greift genau diese Probleme, mangelnde Effizienz und Semantik, an.

Wesentlicher Baustein der Arbeit sind *unscharfe Begriffe*: unscharfe Mengen, die von Einzelheiten abstrahieren und in vielen Fällen mit natürlichsprachlichen Begriffen (*linguistischen Werten*) identifiziert werden können. Die Begriffe spiegeln die anschauliche Bedeutung der unscharfen Mengen wider. Dabei greifen wir auf eine Semantik für unscharfe Mengen zurück (*Kontextmodell*), die Zugehörigkeitsgrade mit einer konkreten Bedeutung belegt.

Die grundlegend neue Idee der Arbeit lautet, unscharfe Informationen immer durch eine Kombination vordefinierter unscharfer Begriffe darzustellen. Wie in natürlichsprachlichen Beschreibungen von Phänomenen benutzen wir einen gegebenen Grundwortschatz, um Beobachtungen verständlich zu beschreiben. Die Übersetzung einer unscharfen Information in eine Kombination von Begriffen ist im Allgemeinen mit einem Informationsverlust verbunden. Dies gilt für alle Verfahren im Bereich unscharfen Schließens. In unserem Fall können wir aber sicherstellen, dass die Originalinformation bei der Rückübersetzung rekonstruiert wird, falls kein Informationsverlust vorlag.

Der nächste Schritt zeigt, wie wir unscharfe Informationen auf der Ebene der Begriffe auch verarbeiten können. Anstatt direkt mit den unscharfen Informationen selbst zu rechnen, stellen wir sie durch die unscharfen Begriffe dar und rechnen mit den Begriffen. Auf diese Weise hängt der Rechenaufwand der Verarbeitung nur von der Anzahl der Begriffe und nicht der Anzahl der Einzelheiten ab und das unscharfe Ergebnis der Verarbeitung wird leicht verständlich mit bekannten Begriffen beschrieben. In der Arbeit zeigen wir, dass so der Rechenaufwand für regelbasiertes Schließen im allgemeinsten Fall quadratisch von der Anzahl der Begriffe abhängt, während sich das beste alternative Verfahren quadratisch mit der Anzahl der Einzelheiten verhält. Im günstigsten Fall ist unser Verfahren linear und die beste Alternative quadratisch von der Anzahl der Begriffe abhängig.

Weiterhin stellen wir praktisch relevante Verfahren vor, um unscharfe Begriffe zu definieren (Umfragemodelle, unscharfe Gruppierung) und beschreiben, wie man mit scharfen Informationen umgeht. Dieser Teil der Arbeit zeigt, dass in gewissem Rahmen verschiedene Formen unscharfer Informationen mathematisch fundiert in einem Mechanismus verarbeitet werden können. In diesem Zusammenhang lassen sich der Regler von Takagi/Sugeno und viele Neuro-Fuzzy-Ansätze in unsere Theorie einbetten und auf diese Weise mit einer soliden Semantik versehen. Insbesondere können wir auf diese Weise praktische Verfahren für den Entwurf solcher Systeme angeben.

Die Auswirkungen unserer Theorie auf praktische Probleme sind vielschichtig. Die Steigerung der Effizienz kann die Bearbeitung von Problemen ermöglichen, die bisher als zu komplex erschienen. Darüber hinaus sichert die konkrete Semantik der Theorie die Verständlichkeit der unscharfen Ergebnisse. Wir sind davon überzeugt, dass wir auf diese Weise zur Akzeptanz unscharfer Systeme beitragen können, die bisher wegen der schwächeren Semantik weniger Ansehen genossen haben als die Wahrscheinlichkeitsrechnung.

---

# Danksagung

Diese Arbeit entstand am Institut für Programmstrukturen und Datenorganisation (IPD) in der Gruppe von Herrn PROF. DR. GERHARD GOOS. Ihm danke ich herzlich für die Unterstützung sowie die Freiheit bei der Wahl des Themas und der verfolgten Ideen, die er mir bei der Betreuung der Arbeit gewährte. Dabei habe ich in vielerlei Hinsicht von seinem reichen Erfahrungsschatz profitiert. Herrn PROF. DR. RUDOLF KRUSE danke ich für die Übernahme des Zweitgutachtens und seine andauernde Bereitschaft, mir mit fachlichem Rat und Kontakten zur Seite zu stehen. Der Austausch mit seiner Neuro-Fuzzy-Gruppe in Magdeburg ist nach wie vor sehr hilfreich, anregend und unterhaltsam.

Besonderer Dank gebührt CHRISTIAN BORGELT aus Magdeburg und meinem Kollegen UWE WAGNER, die durch zahlreiche Diskussionen, Hinweise und Kommentare zum Manuskript beigetragen haben. ANDRÉ HALAMA, JUTTA KLEIN und THORSTEN SPOTT danke ich herzlich für sorgfältiges und geduldiges Korrekturlesen.

Ein großes Dankeschön geht an meine Kollegen am IPD und im FZI für jedwede Unterstützung, sei es durch einfaches Zuhören, unterhaltsame Arbeitspausen oder zahllose persönliche, fachliche oder technische Ratschläge. Dies gilt insbesondere für meinen langjährigen Zimmerkollegen THILO GAUL, der sogar Sprünge von abstrakten *Elektronika* zu *Easy Listening* mühelos ertrug und größtenteils genoss, FLORIAN LIEKWEG für seine unermüdliche Unterstützung als Studienarbeiter, wissenschaftliche Hilfskraft, Diplomand und schließlich Kollege, RAINER NEUMANN und MARTIN TRAPP zusammen mit THILO GAUL für viele Hinweise, die das Leben mit Rechnern erleichtern oder sogar versüßen, BERND TRAUB als Hüter der technischen Infrastruktur, MARTIN RIEDMILLER und RALF SCHOKNECHT vom Institut für Logik, Komplexität und Deduktionssysteme für die äußerst nette und produktive Zusammenarbeit im DFG-Projekt FYNESSE und schließlich JOACHIM WEISBROD, der es als Betreuer meiner Diplomarbeit und Kollege stets verstand, mich zu motivieren und zu ermutigen.

Darüber hinaus danke ich den Schöpfern von *Clicks+Cuts*<sup>1</sup> und ihren Geistesverwandten für anregende musikalische Unterhaltung und Ablenkung, sei es live oder aus der Konserve.

---

<sup>1</sup>Diverse: *Clicks+Cuts* (Mille Plateaux), 2000

Danken möchte ich auch meinen Eltern KARIN und CLAUS SPOTT, meinem Bruder THORSTEN und meinen Mitbewohnern – meiner Schwester MAIKE und LOTHAR SCHINDELE – die mit viel Anteilnahme und Einsatz für mein Wohl sorgten.

Karlsruhe, im Oktober 2000

Martin Spott

---

# Inhaltsverzeichnis

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>Einleitung</b>                              | <b>1</b>  |
| 1.1      | Ungenauigkeit . . . . .                        | 2         |
| 1.1.1    | Vergrößerung von Informationen . . . . .       | 2         |
| 1.1.2    | Unschärfe von Informationen . . . . .          | 5         |
| 1.2      | Schließen mit unscharfen Begriffen . . . . .   | 8         |
| 1.3      | Ziele und Anforderungen . . . . .              | 9         |
| 1.4      | Aufbau der Arbeit . . . . .                    | 10        |
| <b>2</b> | <b>Grundlagen</b>                              | <b>13</b> |
| 2.1      | Unschärfe Mengen . . . . .                     | 13        |
| 2.2      | Regelbasiertes Schließen . . . . .             | 18        |
| 2.3      | Modellierung unscharfer Begriffe . . . . .     | 21        |
| 2.3.1    | Modell für unscharfe Begriffe . . . . .        | 22        |
| 2.3.2    | Operationen auf unscharfen Begriffen . . . . . | 25        |
| <b>3</b> | <b>Stand der Technik</b>                       | <b>27</b> |
| <b>4</b> | <b>Modellierung granularen Wissens</b>         | <b>33</b> |
| 4.1      | Partitionierung von Universen . . . . .        | 34        |
| 4.2      | Kombination unscharfer Grundbegriffe . . . . . | 39        |
| 4.2.1    | Begriffswahrscheinlichkeiten . . . . .         | 39        |
| 4.2.2    | Synthese . . . . .                             | 41        |
| 4.2.3    | Analyse . . . . .                              | 48        |
| 4.3      | Zusammenfassung . . . . .                      | 56        |

---

|          |  |            |
|----------|--|------------|
| <b>5</b> | <b>Schließen mit unscharfen Begriffen</b>        | <b>59</b>  |
| 5.1      | Grobgranulare Regeln . . . . .                   | 60         |
| 5.2      | Inferenz . . . . .                               | 62         |
| 5.2.1    | Basisverfahren der Inferenz . . . . .            | 63         |
| 5.2.2    | Mehrstufige Inferenz . . . . .                   | 66         |
| 5.2.3    | Regelnetzwerke . . . . .                         | 72         |
| 5.3      | Allgemeine Regeln . . . . .                      | 74         |
| 5.4      | Zuverlässigkeit von Schlussfolgerungen . . . . . | 77         |
| 5.5      | Rechenaufwand . . . . .                          | 79         |
| 5.6      | Zusammenfassung . . . . .                        | 81         |
| <b>6</b> | <b>Semantik und Repräsentation von Wissen</b>    | <b>85</b>  |
| 6.1      | Unschärfe Gruppen . . . . .                      | 87         |
| 6.2      | Darstellung scharfer Werte . . . . .             | 91         |
| 6.3      | Modellierung unscharfer Regeln . . . . .         | 93         |
| 6.3.1    | Positive und negative Regeln . . . . .           | 94         |
| 6.3.2    | Unschärfe Interpolation . . . . .                | 102        |
| 6.4      | Parametermodell für Grundbegriffe . . . . .      | 106        |
| 6.4.1    | Approximation von Grundbegriffen . . . . .       | 107        |
| 6.4.2    | Speicheraufwand . . . . .                        | 112        |
| 6.5      | Zusammenfassung . . . . .                        | 113        |
| <b>7</b> | <b>Zusammenfassung und Ausblick</b>              | <b>115</b> |
| 7.1      | Zusammenfassung . . . . .                        | 115        |
| 7.2      | Ausblick . . . . .                               | 120        |
| 7.2.1    | Erweiterung der Theorie . . . . .                | 120        |
| 7.2.2    | Auswirkungen auf andere Gebiete . . . . .        | 121        |

---

|                                      |            |
|--------------------------------------|------------|
| <b>A Beispiel: Betrugserkennung</b>  | <b>123</b> |
| A.1 Modellierung . . . . .           | 124        |
| A.2 Ergebnisse . . . . .             | 128        |
| A.3 Bewertung . . . . .              | 135        |
| <b>B Mathematischer Anhang</b>       | <b>137</b> |
| B.1 Maßtheorie . . . . .             | 137        |
| B.2 Funktionsapproximation . . . . . | 139        |
| <b>Literaturverzeichnis</b>          | <b>143</b> |



# Kapitel 1

---

## Einleitung

*„As its name suggests, computing with words (CW) is a methodology in which words are used in place of numbers for computing and reasoning. [...] There are two major imperatives for computing with words. First, computing with words is a necessity when the available information is too imprecise to justify the use of numbers, and second, when there is a tolerance for imprecision which can be exploited to achieve tractability, robustness, low solution costs, and better rapport with reality.“*

LOTFI A. ZADEH<sup>1</sup>

Als Menschen gehen wir ganz selbstverständlich mit ungenauen Informationen um. So kommunizieren wir auf der Basis eines Wortschatzes, der an vielen Stellen nicht scharf definiert ist, und sind trotzdem in der Lage einander zu verstehen. Wir können sinnvolle Entscheidungen fällen, selbst wenn wir Situationen nicht perfekt einschätzen können und die unterliegenden Gesetzmäßigkeiten nur grob bekannt sind. Als Beispiel sei hier nur medizinische Diagnostik genannt. Unsere Leistungsfähigkeit ist hinsichtlich der Problemgröße beschränkt; je mehr Fakten und Hintergrundwissen wir berücksichtigen müssen, desto schlechter arbeiten wir.

Quantitativ große Aufgabenstellungen können nur noch maschinell gelöst werden. Eine maschinelle Verarbeitung hat darüber hinaus Vorteile: Sie läuft automatisiert ab und liefert deterministische Ergebnisse. Allein der letzte Punkt hat sich in praktischen Anwendungen als sehr nützlich erwiesen. Ein Beispiel ist die Beurteilung der Kreditwürdigkeit von Bankkunden. Untersuchungen haben gezeigt, dass sich Bankangestellte erstens nicht vollständig an die Vorgaben der Bank halten und zweitens gerade in kleineren Städten die Entscheidungen stark durch soziokulturelle Faktoren beeinflusst werden [vAJ00]. Ein rechnergestützter Ansatz vermeidet beides.

Das Thema dieser Arbeit ist eine Theorie zur maschinellen Verarbeitung unscharfer Informationen. Wir beschränken uns auf regelbasiertes unscharfes Schließen, an das wir folgende Forderungen stellen: Die Verfahren sollen mehrstufiges unscharfes Schließen ermöglichen, effizient hinsichtlich Speicher- und Rechenaufwand sein, verständliche und exakte unscharfe Ergebnisse berechnen und die Inte-

---

<sup>1</sup>aus [Zad96]

gration von Informationen verschiedener Semantiken erlauben. Im Zentrum der Betrachtungen steht dabei die Verarbeitung unscharfer Begriffe im Sinne von *Computing with Words*: symbolisches „Rechnen“ mit groben, unscharfen Aussagen, die wir mit Begriffen aus unserem Wortschatz beschreiben.

Nach einer allgemeinen Einführung in die Darstellung und Verarbeitung unscharfer Informationen werden wir die Anforderungen näher erläutern und den Aufbau der Arbeit zusammenfassen.

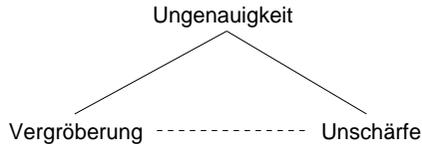


Abbildung 1.1: Aspekte von Ungenauigkeit

## 1.1 Ungenauigkeit

Als Einführung in die Thematik beleuchten wir verschiedene Aspekte der *Ungenauigkeit* von Informationen anhand von Abb. 1.1. Diese bilden die Grundlage für die Verarbeitung unscharfer Informationen. Wir erkennen *Vergrößerung* und *Unschärfe* als wesentliche Aspekte der Ungenauigkeit von Informationen, die aber nicht unabhängig voneinander sind. So geht Unschärfe, wie wir sehen werden, immer mit Vergrößerung einher.

### 1.1.1 Vergrößerung von Informationen

Der „Vater“ der Fuzzy-Techniken, LOTFI A. ZADEH, hat gerade in den letzten Jahren immer wieder auf die Wichtigkeit der Granularisierung von Wissen hingewiesen [Zad97], nicht zuletzt aus dem Gefühl heraus, dass dieser Aspekt lange Jahre vernachlässigt worden ist. *Vergrößerung* oder *Granularisierung* bedeutet, von *Einzelheiten* zu abstrahieren. Einzelheiten sind in unserem Fall die kleinsten Informationseinheiten, die in der Anwendung vorliegen, z. B. physikalische Messwerte. Unter der Vergrößerung oder Granularisierung von Informationen verstehen wir, Einzelheiten zu übergeordneten Informationseinheiten zusammenzufassen. So könnten wir z. B. die Körpergrößen im Bereich 175cm bis 185cm als Einzelheiten in der groben Informationseinheit *normale Körpergröße* eines erwachsenen Mannes vereinen.

In dieser Arbeit unterscheiden wir nur zwei Granularitätsstufen von Informationen: *feingranular* und *grobgranular*. Die feingranulare Ebene ist die der Einzelheiten, in der grobgranularen Ebene befinden sich grobe Informationseinheiten. Diese Beschränkung auf zwei Granularitätsstufen ist künstlich und dient ausschließlich dazu, die Darstellung der mathematischen Theorie einfach zu halten. Selbstverständlich sind Anwendungsszenarien denkbar, die mehrere Hierarchiestufen von Granularität benötigen. Die grundlegenden Lösungsstrategien bleiben aber gleich.

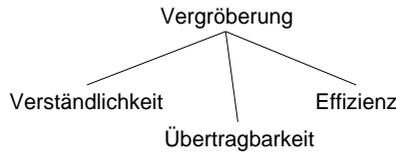


Abbildung 1.2: Aspekte einer Vergrößerung

Als Motivation für die Vergrößerung von Wissen diskutieren wir nun die Aspekte *Verständlichkeit* und *Übertragbarkeit* grober Informationen sowie *Effizienz* der Verarbeitung, s. Abb. 1.2.

**Verständlichkeit** Die Gegensätzlichkeit von Genauigkeit und Verständlichkeit formulierte ZADEH in seinem vielzitierten *Prinzip der Inkompatibilität* folgendermaßen [Zad73]: “[...] as the complexity of a system increases, [...] precision and significance (or relevance) become almost exclusive characteristics.“ Je komplexer ein System ist, desto weniger tragen Einzelheiten zum Verständnis des Systems bei, sie werden unwichtig. Übertragen auf unscharfe Informationsverarbeitung bedeutet dies, dass die Ergebnisse möglichst genau, auf der anderen Seite aber auch verständlich sein sollen.

Zum Beispiel beschreiben wir die Bauteile Rahmen, Scheibe und Griff eines Fensters normalerweise auf diesem abstrakten Niveau. Setzen wir diese Teile zusammen, so erhalten wir ein Fenster unabhängig davon, ob die Teile rot oder blau, klein oder groß, rund oder eckig sind. Alternativ könnten wir die Bauteile auch durch ihre Einzelteile wie Holzleisten, Dübel, Schrauben usw. darstellen. Das zusammengesetzte Fenster wäre dann nur ein Konstrukt aus diesen Einzelteilen. Für das Verständnis sind diese Einzelheiten aber nicht wichtig, ganz im Gegenteil. Es ist u. U. schwierig, in dem Konstrukt von Einzelteilen ein Fenster zu erkennen, wenn wir es nicht mit eigenen Augen sehen. Das offenbart sich besonders in Anwendungen, bei denen grobe Informationseinheiten mit allgemein verständlichen Begriffen wie *Rahmen*, *Scheibe* oder *normale Körpergröße* belegt werden. Hierbei

stellen die Begriffe den Wortschatz der Anwender dar, auf dessen Ebene sich auch die Verständlichkeit bewegt.

In Anwendungen im Bereich des *Data-Mining* ist Verständlichkeit sogar das Hauptziel aller Anstrengungen. Aus den Einzelheiten, den Daten, sollen abstrahierte Zusammenhänge gewonnen werden, die in der Sprache der Anwender ausgedrückt werden können. Erst in dieser vergrößerten, verständlichen Darstellung sind die Daten nutzbar. In dieser Arbeit identifizieren wir daher grobe Informationseinheiten immer mit Begriffen, die deren Bedeutung repräsentieren. Wie der Titel „Schließen mit unscharfen Begriffen“ bereits suggeriert, sind *Begriffe* der zentrale Bestandteil der Untersuchungen.

**Effizienz** Eine effiziente Verarbeitung von Informationen ist bei jenen Anwendungen sehr wichtig, bei denen mit großen Datenmengen operiert wird. In vielen Branchen werden in zunehmendem Maße Daten gesammelt und für verschiedene Zwecke verarbeitet. Da sich das Datenvolumen oft bereits im Bereich von Terabytes bewegt, müssen die Algorithmen für die Verarbeitung sehr effizient sein. Ansonsten ist die Verarbeitung praktisch nicht durchführbar. Aus dieser Sicht ist eine Vergrößerung der Informationen sehr sinnvoll. Nur bei dem einmaligen Vorgang der Vergrößerung müssen wir jedes Datum anfassen, danach arbeiten wir nur noch mit groben Informationseinheiten. Die Größe, von der Rechen- und Speicheraufwand der Verarbeitung abhängen, ist dann die Anzahl der groben Informationseinheiten und nicht die Anzahl der Einzelheiten.

Je nach Anwendung können durchaus mehrere zehntausend Einzelheiten zu einer groben Einheit zusammengefasst werden, wenn wir z. B. Kunden von Telefongesellschaften klassifizieren. Entsprechend reduziert sich der Aufwand. Im Idealfall können wir die Einzelheiten zu groben Einheiten zusammenfassen, die aus der Sicht der Anwendung nicht unterschieden werden können oder müssen. In diesem Fall liegt kein Informationsverlust durch die Vergrößerung vor. Zum Beispiel können wir Pilze in die Klassen der giftigen und der nicht giftigen einordnen. Für die Entscheidung, ob ein Pilz gegessen werden darf, müssen wir nicht wissen, welcher Pilzsorte er angehört; die Kenntnis der Klasse ist vollkommen ausreichend.

In realen Anwendungen müssen wir von einem Informationsverlust durch die Vergrößerung ausgehen. Ein klassisches Beispiel ist die Digitalisierung von Signalen. Wenn wir von der Möglichkeit absehen, das digitale Signal durch Filter wieder zu glätten, müssen wir einen echten Informationsverlust hinnehmen, der sich bei Audiosignalen z. B. in einer Verschlechterung der Klangqualität äußert. In diesem Fall muss man bei der Vergrößerung die Balance zwischen Aufwand und Qualität des Ergebnisses halten. So kann man bei Kompressionsverfahren wie MPEG beispielsweise die Vergrößerung und damit den Aufwand und die Klangqualität steuern.

**Übertragbarkeit** Ein weiterer Aspekt ist die Übertragbarkeit von abstraktem Wissen. Ähnliche Probleme können meist ähnlich gelöst werden. Dabei sind die Lösungen auf grobgranularer Ebene u. U. identisch. Dies wird besonders bei Anwendungen deutlich, bei denen die groben Informationseinheiten mit verständlichen Begriffen belegt sind. Dass Rahmen, Scheibe und Griff ein Fenster ergeben, gilt für die meisten Fenster. Wenn wir diese Regel einmal gewonnen haben, können wir sie auch auf andere Rahmen, Scheiben und Griffe anwenden. Liegt die Regel jedoch nur auf der Ebene der Einzelheiten vor, z. B. den Einzelteilen, so ist das Wissen auf ein spezielles Fenster beschränkt.

Zusammenfassend lässt sich feststellen, dass Vergrößerungen die Verständlichkeit und die Effizienz der Verarbeitung von Informationen erhöhen. Die Genauigkeit dagegen lässt i. A. nach. Man wird in Anwendungen also immer einen Kompromiss anstreben müssen, der den jeweiligen Anforderungen gerecht wird. Die Übertragbarkeit von abstraktem Wissen ist zwar ein wichtiger Punkt, wird aber in dieser Arbeit nicht weiter untersucht.

Bisherige Ansätze zur Verarbeitung unscharfer Informationen werden den Aspekten Verständlichkeit und Effizienz meist nicht gerecht. Die Informationen werden zwar vergrößert, die meisten Verfahren operieren aber weiterhin auf den Einzelheiten, weshalb die Vergrößerung keinen Effizienzgewinn bringt. Die Ergebnisse der Verarbeitung liegen auch meist auf feingranularer Ebene vor. Wie beim Beispiel mit dem Fenster muss man dann versuchen, eine Menge von Einzelheiten im Nachhinein zu interpretieren, was in vielen Fällen sehr schwierig ist. An dieser Stelle greifen unsere Verfahren, die konsequent auf grobgranularer Ebene operieren und damit erstmals Verständlichkeit und Effizienz sicherstellen.

### 1.1.2 Unschärfe von Informationen

Grundsätzlich unterscheiden wir wie in Abb. 1.3 dargestellt zwei Arten von Unschärfe: *Vagheit* und *Unsicherheit*. Sie stellen zwei prinzipiell verschiedene Konzepte dar, können aber auch kombiniert auftreten. In Kap. 2 werden wir darlegen, dass Vagheit auch durch Unsicherheit ausgedrückt werden kann.

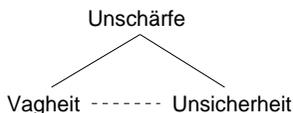


Abbildung 1.3: Konzepte von Unschärfe

- **Vagheit:** Ein Ausdruck ist vage, wenn der Bereich der Situationen oder Objekte, auf die er anwendbar ist, nicht scharf begrenzt werden kann. Natürlichsprachliche Begriffe wie *normale Körpergröße* sind ein typisches Beispiel. Es gibt keine genaue Zentimeterangabe, an der der Bereich der normalen Körpergrößen beginnt und keine, an der er endet.
- **Unsicherheit:** Eine unsichere Information ist genau (scharf). Wir können aber nicht ausschließen, dass sie falsch ist. Ein digitales Thermometer z. B. zeigt einen scharfen Wert an. Wir sind aber unsicher, ob er richtig ist, weil wir wissen, dass das Thermometer nicht immer genau misst. Die Wahrscheinlichkeitsrechnung basiert auf dem Konzept Unsicherheit.

Wie wir in Abschnitt 1.1.1 gesehen haben, werden bei der Vergrößerung von Informationen Einzelheiten zu Gruppen zusammengefasst, d. h. formal, dass Elemente zu Mengen gruppiert werden. Als Beispiel diene die Menge  $[175\text{cm}, 185\text{cm}]$  zur Definition des Begriffs *normale Körpergröße* eines erwachsenen Mannes. Diese Form der Vergrößerung ist in vielerlei Hinsicht nicht zufriedenstellend, wie wir nun aus der Sicht von Vagheit und Unsicherheit beleuchten werden.

Aus der Sicht von Vagheit lassen sich scharfe Grenzen wie  $175\text{cm}$  und  $185\text{cm}$  selten rechtfertigen. Warum soll ein Mann mit  $175\text{cm}$  normal groß sein, einer mit  $174\text{cm}$  jedoch nicht mehr? Ob man einen Begriff wie *normale Körpergröße* mit einer scharfen Menge angemessen beschreibt, ist aber nicht nur eine Frage der intuitiv richtigen Darstellung eines Begriffs, sondern hängt insbesondere davon ab, wozu der Begriff in der Anwendung verwendet wird. So ist eine Definition von *normal groß* durch eine scharfe Menge angemessen, wenn die Folgen der Information ebenso scharf getrennt sind. Das ist in vielen Fällen aber nicht der Fall. Zum Beispiel ist es nicht sinnvoll, die Größe eines Autositzes ausschließlich nach der Körpergröße zu bestimmen. Der Übergang zwischen der Entscheidung für einen kleinen gegenüber einem großen Sitz findet nicht bei einer scharfen Körpergröße statt.

Kommen wir zur Unsicherheit. Für einen Begriff wie *normal groß* existieren i. A. viele Definitionen. Dafür gibt es zwei Gründe. Die Definition eines Begriffs hängt einerseits vom Zusammenhang ab, in dem seine Bedeutung geklärt werden soll. Es ist ein Unterschied, ob wir über normale Körpergrößen von Männern oder Frauen sprechen. Andererseits können selbst bei festem Zusammenhang verschiedene Definitionen vorliegen, weil z. B. Uneinigkeit herrscht oder Messinstrumente Toleranzen unterliegen. Wir sind dann unsicher, welche Definition die richtige ist. Im Zweifelsfall müssen wir alle gleich ernst nehmen. Hier wird deutlich, dass Unsicherheit dazu benutzt werden kann, Vagheit auszudrücken: Unsicherheit resultiert in verschiedenen Definitionen eines vagen Begriffs, deren Gesamtheit wir aber gerade als unsichere Definition des vagen Begriffes interpretieren können. Für den

weiteren Verlauf der Arbeit nehmen wir an, dass der Zusammenhang, in dem ein Begriff definiert wird, fest ist.

Das Nicht-Akzeptieren scharfer Grenzen und das notwendige Berücksichtigen verschiedener Definitionen zeigen, dass eine scharfe Menge in vielen Fällen keine geeignete Formalisierung für einen Begriff ist, der eine vergrößernde Informationseinheit darstellt. Einen solchen Begriff bezeichnen wir als *unscharf*. ZADEH prägte 1965 den Begriff der *unscharfen Menge*<sup>2</sup>, die diesen Mangel scharfer Mengen beseitigt [Zad65]. Während Elemente in einer scharfen Menge entweder enthalten sind oder nicht, können sie einer unscharfen Menge graduell angehören. Eine unscharfe Menge wird durch diese *Zugehörigkeitsgrade* der Elemente definiert, deren Wertebereich meist das Einheitsintervall  $[0, 1]$  ist. Im Gegensatz dazu ist der Wertebereich bei einer scharfen Menge  $\{0, 1\}$ ; ein Element gehört zu einer Menge oder nicht.

Es ist entscheidend für die Verarbeitung unscharfer Informationen zu definieren, was die Zugehörigkeitsgrade einer unscharfen Menge bedeuten. Wir müssen Verfahren angeben können, um Zugehörigkeitsgrade und damit unscharfe Begriffe eindeutig zu definieren. Die Wichtigkeit dieser Forderung untermauern D. DUBOIS und H. PRADE in [DP97]:

*„These difficulties are bypassed by assuming that degrees of membership to fuzzy sets pertaining to unrelated concepts are commensurate. This is done by resorting to a common membership scale (that need not be numerical). This commensurability assumption is often taken for granted and never emphasized in the fuzzy literature. [...] The commensurability assumption is really the one that allows fuzzy set theory to work.“*

Insbesondere wenn Informationen aus verschiedenen Quellen stammen, sind auch die Formen der Unschärfe unterschiedlich. Um die Informationen trotzdem in einem System verarbeiten zu können, müssen wir eine einheitliche Formalisierung finden, welche die Kommensurabilität der Informationen sicherstellt. Wie schon DUBOIS und PRADE schreiben, vernachlässigen die meisten Arbeiten im Bereich der Verarbeitung unscharfer Informationen diese Anforderungen. In theoretischen Arbeiten wird nicht genau erklärt, was ein Zugehörigkeitsgrad ist, und in praktischen Arbeiten sieht man mangels praxistauglicher Theorie nicht so genau hin und bewegt sich auf unsicherem Terrain. Diese Lücke werden wir in dieser Arbeit besonders berücksichtigen. Wir werden theoretisch fundieren, wie praxisrelevante Formen unscharfen Wissens in einem Mechanismus verarbeitet werden können.

---

<sup>2</sup>engl. *fuzzy set*

## 1.2 Schließen mit unscharfen Begriffen

Das Schließen mit unscharfen Begriffen basiert auf zwei Komponenten: einer *Wissensbasis* und einem *Inferenzmechanismus*. In der Wissensbasis werden alle Informationen gesammelt, die im Zusammenhang mit der Lösung der gegebenen Aufgabenstellung wichtig und allgemeingültig sind. Wissen ist dann allgemeingültig, wenn es im Rahmen der Aufgabenstellung immer richtig ist. Eine *Inferenz* ist das Ableiten von Informationen aus der Wissensbasis; der Inferenzmechanismus ist dementsprechend das Verfahren, das die Inferenz durchführt.

Bei medizinischer Diagnose z. B. bilden unsere Kenntnisse über den Zusammenhang von Symptomen und Ursachen die Wissensbasis. Beobachtete Symptome werden dann dazu benutzt, um mithilfe der Wissensbasis auf die Ursachen der Erkrankung zu schließen. Als *Beobachtung* bezeichnen wir dabei eine Information, die nicht in der Wissensbasis enthalten ist, weil sie für die Aufgabenstellung kein allgemeingültiges Wissen darstellt. Im Beispiel sind die Symptome Beobachtungen, weil sie von den Patienten abhängen und daher für eine Diagnose nicht allgemeingültig sind. Inferenz kann also auch bedeuten, eine Beobachtung mit der Wissensbasis zu verknüpfen, um neue Informationen zu gewinnen, die aus der Wissensbasis allein nicht abgeleitet werden können.

Beim Schließen mit *unscharfen Begriffen* wird die Wissensbasis auf grobgranularer Ebene mit unscharfen Begriffen definiert. Das Wissen wird so in leicht verständlicher Form formuliert, denn die unscharfen Begriffe sind Ausdruck unseres Verständnisses. Wir beschränken uns in dieser Arbeit auf regelbasierte Wissensbasen, wobei wir Regeln der Bauart „Wenn Aussage A gilt, dann gilt Aussage B.“ benutzen. Im Gegensatz zu traditionellen Expertensystemen sind die Aussagen aber unscharf, da sie auf unscharfen Begriffen basieren. Das bedeutet insbesondere, dass die Gültigkeit der Aussage A als Antezedens der Regel nicht binär, wahr oder falsch, ist. Vielmehr können Antezedenzen graduell erfüllt sein. Hinzu kommt, dass die Gültigkeit einer Regel unsicher sein kann. Ein Inferenzmechanismus muss aus diesen graduellen Gültigkeiten der Regeln und der Antezedenzen einen unscharfen Schluss berechnen.

Im Gegensatz zu bekannten Verfahren im Bereich regelbasierten unscharfen Schließens liegt bei unseren Verfahren der Schluss sowohl auf grobgranularer Ebene als auch auf feingranularer Ebene vor. Damit sind die Verständlichkeit und die Exaktheit des Schlusses sichergestellt. Darüber hinaus werden wir verschiedene Wege vorstellen, um zu Regeln, also der Wissensbasis, zu gelangen. Wie schon bei der Definition unscharfer Begriffe stellen wir dabei sicher, dass Wissen aus verschiedenen Informationsquellen ohne semantische Bedenken kombiniert werden kann. Damit gewinnen wir nicht nur Praxishöhe, sondern stellen auch einige bisher nur intuitiv motivierte Verfahren unscharfer Informationsverarbeitung auf ein solides, theoretisches Fundament.

## 1.3 Ziele und Anforderungen

Ziel der vorliegenden Arbeit ist eine mathematische Theorie für die maschinelle Verarbeitung grobgranularer, regelbasierter Informationen. Die resultierenden Verfahren sollen folgende Anforderungen erfüllen:

1. Mehrstufiges unscharfes Schließen
2. Effizienz hinsichtlich Speicher- und Rechenaufwand
3. Verständlichkeit und Exaktheit unscharfer Ergebnisse
4. Integration von Informationen verschiedener Semantiken

Zu 1: Die Möglichkeit zu mehrstufigem unscharfen Schließen ist aus zwei Gründen wichtig:

- Zusammenhänge zwischen Größen können oft nicht direkt formuliert werden, sondern nur über Zwischengrößen.
- Die Anzahl der Regeln in einem mehrstufigen, hierarchischen System ist wesentlich kleiner als in einem flachen, einstufigen System.

Ein Beispiel für die mehrstufige Modellierung eines Systems sind praktische Realisierungen von Wettervorhersage. Während in der ersten Stufe nur die Großwetterlage bestimmt wird, geht – vereinfacht dargestellt – zusätzlich zu diesen Informationen Wissen über lokale Gegebenheiten wie die Topologie des Geländes in die zweite Stufe ein. Ein mehrstufiges System ist in diesem Fall die einzige Möglichkeit, trotz der Komplexität der Aufgabe zu sinnvollen Resultaten zu gelangen. Zusammenfassend führt Mehrstufigkeit also zu besserem Verständnis sowie geringerem Rechen- und Speicheraufwand und ist u. U. die einzige Möglichkeit, Wissen zu formulieren.

Zu 2: Effizienz bedeutet, dass die Größe, von der Speicher- und Rechenaufwand abhängen, die Anzahl grobgranularer Informationen sein soll, insbesondere also nicht die Anzahl der Einzelheiten. Für die Rechenverfahren heißt dies, dass sie nur auf der Ebene der grobgranularen Informationen operieren sollen. Denn aus Sicht der Effizienz gewinnen wir nichts, wenn Wissen vergrößert wird, sich die Berechnungen aber auf der feingranularen Ausgangsebene der Einzelheiten bewegen. Schließlich soll Vergrößerung in vielen Fällen die Bearbeitung von Problemen erst ermöglichen, die auf der feingranularen Ausgangsebene zu komplex waren.

Zu 3: Unter der Verständlichkeit unscharfer Ergebnisse verstehen wir, dass die unscharfen Schlüsse des Systems auf der grobgranularen Ebene ausgedrückt werden können. Dies ist wichtig, weil die grobgranulare Ebene die für den Anwender verständliche Sprache darstellt. Trivialerweise gilt das insbesondere für Systeme,

die der Gewinnung von Erkenntnissen über beobachtete Zusammenhänge dienen (Data-Mining). Gleichzeitig muss aber sichergestellt werden, dass die Exaktheit der Schlüsse auf der feingranularen Ebene der Einzelheiten gewährleistet ist. Ein Beispiel sind Werbestrategien für Produkte. Um den Stil einer Fernsehwerbung festzulegen, müssen wir nur grobe Zielgruppen kennen. Einzelne Personen innerhalb dieser Gruppen interessieren nicht. Müssen wir hingegen entscheiden, wer einen Werbebrief bekommt, so müssen wir die Zugehörigkeit jeder Person zu den Zielgruppen betrachten. Wie das Beispiel verdeutlicht, muss ein unscharfes Ergebnis also sowohl auf grobgranularer als auch auf feingranularer Ebene vorliegen und beide Darstellungen müssen durch eine konkrete Semantik verknüpft und in diesem Sinne wohldefiniert sein.

Zu 4: Bei der Verarbeitung unscharfer Informationen wird Wissen aus verschiedenen Quellen gesammelt, das ganz unterschiedliche Qualitäten haben kann und in dieser Hinsicht heterogen ist. Dabei unterscheiden wir die Art der Unschärfe, die Ordnung der Unschärfe und die Art der Beschreibung. Ein unscharfes System muss in der Lage sein, insofern mit dieser Form von Heterogenität umgehen zu können, als es alle Formen von Wissen in einem Mechanismus verarbeitet. Wir müssen es formal mathematisch so entwerfen, dass praxisrelevante Informationstypen auf das formale Modell zurückgeführt werden können. Die wichtigsten Arten von Informationen sind

- scharfe Messungen
- vage Beschreibungen
- unsichere Daten
- positive/negative Informationen

Wir verlangen dabei, dass diesen abstrakten Datentypen eine konkrete Semantik unterlegt wird, die den Anforderungen der Praxis gerecht wird.

## 1.4 Aufbau der Arbeit

In Kapitel 2 stellen wir die Grundlagen unscharfer Informationsverarbeitung zusammen und definieren grundlegende Begrifflichkeiten und Schreibweisen. Weiterhin beschreiben wir mit dem Kontextmodell detailliert die Semantik und Modellierung unscharfer Begriffe, die wir in dieser Arbeit verwenden. Der Stand der Technik in Kapitel 3 beleuchtet verwandte Arbeiten hinsichtlich der in Abschnitt 1.3 gestellten Anforderungen an diese Arbeit. Insbesondere werden die Defizite hinsichtlich der Verständlichkeit und Semantik unscharfer Schlüsse sowie

der Effizienz der Verfahren herausgestellt. Daran schließen sich die Kernkapitel der Arbeit an.

In Kapitel 4 definieren wir die *unscharfe Partitionierung* eines Universums durch *unscharfe Grundbegriffe*, welche die grobgranulare, symbolische Ebene der Informationsverarbeitung bilden. Wir stellen in Abschnitt 4.2 die grundlegende Idee vor, unscharfe Aussagen durch *gewichtete Kombinationen* der Grundbegriffe auszudrücken. Durch die *Synthese* und *Analyse* unscharfer Mengen garantieren wir dabei, dass Aussagen auf grobgranularer und feingranularer Ebene auf der Basis einer konkreten Semantik eindeutig ineinander überführt werden können.

Unscharfe Regeln modellieren wir in Kapitel 5 konsistent zu der gewichteten Kombination als bedingte Wahrscheinlichkeiten auf den Grundbegriffen. Abschnitt 5.2 widmet sich der Inferenz. Sie kann auch im Fall mehrstufiger Regelbasen auf grobgranularer Ebene durchgeführt werden, ohne die Exaktheit der Schlüsse auf feingranularer Ebene zu verlieren. Wir zeigen in Abschnitt 5.4, wie man die Zuverlässigkeit unscharfer Schlüsse misst und in Abschnitt 5.5, dass der Aufwand zur Berechnung unscharfer Schlüsse wesentlich geringer als bei anderen Ansätzen ist.

Kapitel 6 untermauert die Modellierung unscharfer Begriffe und des Inferenzmechanismus insofern, als hier andere Semantiken für unscharfe Begriffe und Regeln angenommen werden, die aber ebenfalls in den vorgestellten Verfahren resultieren. In diesem Rahmen können wir symbolisch schließen, ohne uns um die unterliegende Semantik kümmern zu müssen, und haben mehr Möglichkeiten, um Wissen zu formulieren. Da die Anzahl der Daten zur Speicherung eines Grundbegriffs sehr groß sein kann, wird in Abschnitt 6.4 ein Parametermodell zur Repräsentation der Grundbegriffe vorgestellt, das den Speicheraufwand deutlich reduziert.

Abschließend werden in Kapitel 7 die Ergebnisse zusammengefasst und ein Ausblick gegeben.

Anhang A illustriert unsere Verfahren mit einem System zur Erkennung von Versicherungsbetrug und demonstriert die Vorteile gegenüber anderen Ansätzen zum unscharfen Schließen. Die mathematischen Grundlagen der Maßtheorie und der Funktionsapproximation, die in der Arbeit benötigt werden, fassen wir in Anhang B zusammen.



## Kapitel 2

---

# Grundlagen

Um das Verständnis beim Lesen dieser Arbeit zu erleichtern, fassen wir in diesem Kapitel Grundlagen aus dem Bereich unscharfer Informationsverarbeitung zusammen. Dabei handelt es sich um wesentliche Konzepte zur Definition unscharfer Begriffe, die zur Darstellung vergrößerten Wissens dienen. Nebenbei legen wir die Notation der Begrifflichkeiten fest. Insbesondere geben wir eine Übersicht über mögliche Semantiken unscharfer Mengen, die aus verschiedenen Arten von Unschärfe resultieren. Grundlegende Verfahren für regelbasiertes, unscharfes Schließen werden in Abschnitt 2.2 behandelt.

In Abschnitt 2.3 vertiefen wir die Einführung, indem wir mit dem Kontextmodell eine mathematische Formalisierung unscharfer Begriffe beschreiben. Diese Formalisierung hat sich im Bereich unscharfer Informationsverarbeitung als sehr nützlich erwiesen, weil sie intuitiv, einfach verständlich und anwendungsnah ist. Die vorliegende Arbeit baut auf diesen Ideen auf.

## 2.1 Unscharfe Mengen

Bei der Granularisierung von Wissen gehen wir von der Ebene der **Einzelheiten** aus, welche die feinste Granularisierungsstufe der Modellierung darstellt. Einzelheiten sind die kleinsten Informationseinheiten innerhalb der Anwendung. Wir fassen sie in einer Menge zusammen, die wir als **Universum** bezeichnen und mit kalligraphischen Buchstaben  $\mathcal{X}$ ,  $\mathcal{Y}$  usw. kennzeichnen. Für die zugehörigen Einzelheiten schreiben wir  $u$ ,  $v$  usw. Wir nehmen in dieser Arbeit an, dass die Universen kompakte Mengen sind. Praxisrelevante Beispiele sind endliche Universen und beschränkte, abgeschlossene Teilmengen des  $\mathbb{R}^n$ , insbesondere also abgeschlossene Intervalle  $[a, b]$  auf  $\mathbb{R}$ .

Einem Universum sind eine oder mehrere **Variablen** zugeordnet, bezeichnet mit kleinen Buchstaben  $x$ ,  $y$  usw., die Werte aus dem Universum annehmen können, z. B.  $x = u \in \mathcal{X}$ . Das Universum könnte z. B. die Menge  $\mathcal{X} = [1\text{cm}, 300\text{cm}]$  möglicher Körpergrößen von Menschen<sup>1</sup> sein und die Variable  $x = u \in \mathcal{X}$  die Körpergröße einer ausgewählten Person angeben.

---

<sup>1</sup>Als größter Mensch der Welt gilt ROBERT PERSHING WADLOW (1928–1940) aus Alton, Illinois, mit 2,72m; URL: <http://www.altonweb.com/history/wadlow> (Oktober 2000).

Als das Grundelement unscharfer Informationsverarbeitung definieren wir eine **unscharfe Menge** auf dem Universum  $\mathcal{X}$  durch eine **Zugehörigkeitsfunktion**  $\mu_{\tilde{A}} : \mathcal{X} \rightarrow [0, 1]$ .  $\mu_{\tilde{A}}(u)$  bezeichnet dabei den **Zugehörigkeitsgrad** des Wertes  $u \in \mathcal{X}$  zu der unscharfen Menge  $\tilde{A}$ . Für die Menge aller unscharfen Mengen auf  $\mathcal{X}$  schreiben wir  $\mathcal{UM}(\mathcal{X})$ , die Menge  $\text{supp}(\tilde{A}) := \{u \in \mathcal{X} \mid \mu_{\tilde{A}}(u) > 0\}$  bezeichnen wir als **Träger**<sup>2</sup> oder Trägermenge von  $\tilde{A}$ .

Eine **unscharfe Relation** zwischen Universen  $\mathcal{X}$ ,  $\mathcal{Y}$  ist eine unscharfe Menge auf dem kartesischen Produkt  $\mathcal{X} \times \mathcal{Y}$ . Eine unscharfe Menge  $\tilde{A}$  auf  $\mathcal{X}$  können wir durch **zylindrische Erweiterung** zu einer unscharfen Relation  $\tilde{R} = \tilde{A} \uparrow (\mathcal{X} \times \mathcal{Y})$  auf  $\mathcal{X} \times \mathcal{Y}$  erweitern. Sie wird definiert durch

$$\mu_{\tilde{R}}(u, v) = \mu_{\tilde{A} \uparrow (\mathcal{X} \times \mathcal{Y})}(u, v) := \mu_{\tilde{A}}(u) \quad (2.1)$$

Die Definition spiegelt wider, dass wir nicht wissen, wie die Relation von  $v$  abhängt. Umgekehrt definieren wir die **Projektion**  $\tilde{R} \downarrow \mathcal{X}$  einer Relation  $R$  auf das Universum  $\mathcal{X}$  als

$$\mu_{\tilde{R} \downarrow \mathcal{X}}(u) := \sup_{v \in \mathcal{Y}} \mu_{\tilde{R}}(u, v) \quad (2.2)$$

Wir komprimieren auf diese Weise die Relation zwischen  $\mathcal{X}$  und  $\mathcal{Y}$  zu einer Information auf  $\mathcal{X}$ , indem wir für festes  $u$  die kleinste obere Schranke für den Wert der Relation suchen.

**Schnitt, Vereinigung und Komplement** unscharfer Mengen auf  $\mathcal{X}$  werden punktweise über Abbildungen  $i : [0, 1]^2 \rightarrow [0, 1]$ ,  $u : [0, 1]^2 \rightarrow [0, 1]$  und  $c : [0, 1] \rightarrow [0, 1]$  erklärt:<sup>3</sup>

- $\tilde{A} \cap \tilde{B} : \mu_{\tilde{A} \cap \tilde{B}}(u) := i(\mu_{\tilde{A}}(u), \mu_{\tilde{B}}(u))$
- $\tilde{A} \cup \tilde{B} : \mu_{\tilde{A} \cup \tilde{B}}(u) := u(\mu_{\tilde{A}}(u), \mu_{\tilde{B}}(u))$
- $\tilde{A}^c : \mu_{\tilde{A}^c}(u) := c(\mu_{\tilde{A}}(u))$

Von  $i$  und  $u$  fordern wir axiomatisch folgende Eigenschaften, die  $i$  als sog. **t-Norm** und  $u$  als **co-t-Norm** qualifizieren:

- Grenzbedingungen:  $i(1, a) = a$  und  $u(0, a) = a$
- Symmetrie:  $i(a, b) = i(b, a)$  und  $u(a, b) = u(b, a)$
- Monotonie:  $c > b \Rightarrow (i(a, c) \geq i(a, b) \text{ und } u(a, c) \geq u(a, b))$

<sup>2</sup>engl. *support*

<sup>3</sup> $i$  steht für *intersection*,  $u$  für *union* und  $c$  für *complement*.

Die Grenzbedingungen und die Symmetrie stellen sicher, dass  $i$  und  $u$  im Fall scharfer Mengen den scharfen Operationen für Schnitt und Vereinigung entsprechen. Ausgezeichnete und viel benutzte Normen sind  $i = \min$  als größte t-Norm und  $u = \max$  als kleinste co-t-Norm. Die Komplementfunktion  $c$  soll folgende Axiome erfüllen:

- Grenzbedingungen:  $c(0) = 1$  und  $c(1) = 0$
- Monotonie:  $a < b \Rightarrow c(a) \geq c(b)$

Auch hier wird so die Konsistenz mit der Definition für scharfe Mengen garantiert. Das Standardkomplement ist  $c(a) = 1 - a$ .

Eine unscharfe Menge können wir als **unscharfen Wert** einer Variablen  $x$  auffassen. In dieser Weise unterscheiden wir oft nur eine kleine Anzahl von unscharfen Werten, die  $x$  annehmen kann. Eine Menge  $\mathcal{P}$  unscharfer Werte auf  $\mathcal{X}$  nennen wir **unscharfe Granularisierung** oder **Vergrößerung** des Universums  $\mathcal{X}$ , wobei wir annehmen, dass  $\mathcal{P}$  endlich ist. Im Beispiel der Körpergrößen  $\mathcal{X} = [1\text{cm}, 300\text{cm}]$  können wir uns eine Granularisierung  $\mathcal{P} = \{\text{sehr klein, klein, normal, groß, sehr groß}\}$  vorstellen, wenn wir die Beschreibung eines Menschen geben möchten. *sehr klein, klein* usw. sind dabei die möglichen unscharfen Werte. Die (hier kontinuierliche) Menge  $\mathcal{X}$  von Einzelheiten wird also durch eine kleine Anzahl unscharfer Werte überdeckt. Von einer unscharfen Granularisierung  $\mathcal{P} = \{\tilde{A}_1, \tilde{A}_2, \dots, \tilde{A}_n\}$  erwarten wir, dass jedes  $u \in \mathcal{X}$  mindestens einer unscharfen Menge  $\tilde{A}_i$  mit positivem Grad angehört, d. h.  $\mathcal{P}$  muss  $\mathcal{X}$  überdecken. Da wir unscharfen Werten i. A. einen verständlichen Begriff als Bezeichnung zuordnen, z. B. *klein* als Körpergröße, sprechen wir anstatt von unscharfen Werten meist von **unscharfen Begriffen**. Wir erhalten so die grundlegende

**Definition 1** *Ein unscharfer Begriff auf einem Universum  $\mathcal{X}$  wird durch eine unscharfe Menge  $\tilde{A}$  auf  $\mathcal{X}$  erklärt.  $\tilde{A}$  wird durch die Zugehörigkeitsfunktion  $\mu_{\tilde{A}}$  definiert. Für den unscharfen Begriff schreiben wir kurz  $\tilde{A}$ .*

Im Verlauf der Arbeit werden wir wiederholt von **feingranularer** und **grobgranularer Ebene** sprechen. Ersteres meint die Menge der Einzelheiten, das Universum, Letzteres die Vergrößerung, also die Menge unscharfer Begriffe auf dem Universum. Als Synonym für grobgranular verwenden wir den Begriff **symbolisch**, der verdeutlicht, dass wir auf der grobgranularen Ebene nur mit Symbolen, den unscharfen Begriffen, umgehen, die für die unscharfen Mengen auf der feingranularen Ebene stehen.

Eine Aussage „Die Variable  $x$  besitzt den unscharfen Wert  $\tilde{A}$ .“ oder kurz „ $x = \tilde{A}$ “ nennen wir eine **unscharfe Aussage**. Konjunktion, Disjunktion oder Negation von Aussagen bilden wir wie im scharfen Fall, indem wir die beteiligten unscharfen

Mengen schneiden, vereinigen oder das Komplement bilden. Dazu identifizieren wir:

- Negation  $x = \neg \tilde{A}$  mit  $x = \tilde{A}^c$
- Konjunktion  $x = \tilde{A} \wedge x = \tilde{B}$  mit  $x = (\tilde{A} \cap \tilde{B})$
- Disjunktion  $x = \tilde{A} \vee x = \tilde{B}$  mit  $x = (\tilde{A} \cup \tilde{B})$

Auf diese Weise können verknüpfte Aussagen wieder auf atomare Aussagen zurückgeführt werden. Wir betrachten aus diesem Grund im weiteren Verlauf nur atomare Aussagen „ $x = \tilde{A}$ “.

Eine wichtige Frage bei Anwendungen lautet, wie wir Zugehörigkeitsgrade von unscharfen Mengen definieren und was sie bedeuten. Diese Frage nach der **Semantik** wurde in der Geschichte unscharfer Mengen immer wieder gestellt und diskutiert. Wir begreifen einen unscharfen Begriff, den wir mit einer unscharfen Menge identifizieren, grundsätzlich als Ausdruck von Vagheit. Dabei unterscheiden wir drei Möglichkeiten, um Vagheit zu definieren [DP97]: *Ähnlichkeit*, *Unsicherheit* und *Präferenz*. Während Ähnlichkeit ein Spezialfall von Vagheit ist, sind

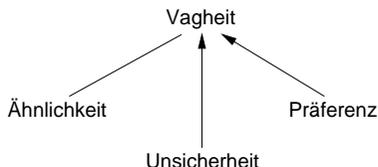


Abbildung 2.1: Ähnlichkeit als spezielle Form von Vagheit sowie Unsicherheit und Präferenz zur Beschreibung von Vagheit

Unsicherheit und Präferenz nur Hilfsmittel, um Vagheit auszudrücken; sie können im Gegensatz zu Ähnlichkeit auch Phänomene beschreiben, die nicht vage sind.

**Ähnlichkeit:** Vagheit wird bei einer Interpretation über Ähnlichkeit durch die Nicht-Unterscheidbarkeit von Einzelheiten gekennzeichnet. Beispielsweise kann man aus größerer Distanz die Körpergröße einer Person nicht zentimetergenau schätzen. Wir sind nicht in der Lage, ähnliche Größen wie 185cm und 187cm zu unterscheiden, sondern können die Größe nur grob abschätzen und mit einem unscharfen Begriff wie *groß* belegen. Dabei werden ähnliche Größen wie 185cm und 187cm in dem unscharfen Begriff subsumiert. Die Variable „Größe der Person“ kann nur unscharfe Werte annehmen.

Formal gehen wir von einer unscharfen *Ähnlichkeitsrelation* (oder Gleichheitsrelation) auf  $X \times X$  aus, also einer unscharfen Relation  $\tilde{S}$  mit  $\mu_{\tilde{S}} : X \times X \rightarrow [0, 1]$  [KF88, KK93].  $\mu_{\tilde{S}}(u_1, u_2)$  misst die Ähnlichkeit von  $u_1$  und

$u_2$ . Eine unscharfe Menge  $\tilde{A}_0$  wird dann durch einen Wert<sup>4</sup>  $u_0 \in \mathcal{X}$  mithilfe der Ähnlichkeitsrelation induziert:  $\mu_{\tilde{A}_0}(u) = \mu_{\tilde{S}}(u_0, u)$ . Die unscharfe Menge  $\tilde{A}_0$  können wir anschaulich mit der Bedeutung „ungefähr  $u_0$ “ belegen.

Nicht unterscheidbare Einzelheiten können zu Gruppen zusammengefasst werden. Auf diese Art und Weise kann das Universum  $\mathcal{X}$  z. B. mithilfe von Gruppierungsverfahren<sup>5</sup> vergrößert werden. Das Ergebnis ist eine Überdeckung von  $\mathcal{X}$  durch mehrere Gruppen, die meist durch einen Prototypen wie  $u_0$  und eine Ähnlichkeitsrelation beschrieben werden. Unscharfe Gruppierungsverfahren [Bez73, GK79, GG89] resultieren in einer Überdeckung von  $\mathcal{X}$  mit unscharfen Gruppen, die durch unscharfe Mengen dargestellt werden. Wenn die Ähnlichkeitswerte der Ähnlichkeitsrelation eine konkrete Bedeutung haben, also z. B. von Metriken wie dem EUKLIDISCHEN Abstand abgeleitet sind, dann gilt dies auch für die Zugehörigkeitswerte der induzierten unscharfen Mengen.

**Unsicherheit:** Nehmen wir an, es lägen verschiedene scharfe Definitionen eines unscharfen Begriffs vor. Wenn wir unsicher sind, welche Definition die richtige ist, müssen wir alle in Betracht ziehen; das Resultat ist eine unscharfe Definition. In einem Experiment könnten wir z. B. die Körpergröße von vielen Personen messen, die von Befragten für groß gehalten wurden. Jede einzelne Messung ist scharf und ein Repräsentant für den unscharfen Begriff *groß*. Zusammengenommen erhalten wir eine Wahrscheinlichkeitsdichte, die als Zugehörigkeitsfunktion den unscharfen Begriff *groß* definiert.

In der Literatur gibt es verschiedene formale Ansätze, um unscharfe Begriffe durch Unsicherheit zu modellieren. Die meisten verallgemeinern die Wahrscheinlichkeitsrechnung dadurch, dass Axiome gelockert oder scharfe Größen durch unscharfe ersetzt werden.

Wir werden in dieser Arbeit zwei Ansätze benutzen. Im *Kontextmodell* werden scharfe Realisierungen einer Zufallsvariablen durch mengenwertige ersetzt, was in *Plausibilitätsdichten* statt Wahrscheinlichkeitsdichten resultiert [Geb92, GK93a].  $\mu_{\tilde{A}}(u)$  misst dann die Plausibilität dafür, dass  $u$  zu  $\tilde{A}$  gehört.

Eng verwandt ist die Interpretation einer unscharfen Menge als *Likelihood-Funktion* [His88, DMP93, Tho95]. Wenn z. B. die unscharfe Menge  $\tilde{A}$  den Begriff *groß* modelliert, so wird  $\mu_{\tilde{A}}(u)$  als *Likelihood*  $L(\tilde{A}; u)$  oder die bedingte Wahrscheinlichkeit  $P(\tilde{A}|u)$  interpretiert, dass  $u$  Element von  $\tilde{A}$  ist. Kontextmodell und Likelihood werden ausführlich in Abschnitt 2.3 besprochen.

---

<sup>4</sup>oder auch eine Teilmenge von  $\mathcal{X}$

<sup>5</sup>engl. *cluster algorithms*

**Präferenz:** Die Zugehörigkeitswerte drücken hier Präferenzen aus. Sie sind auf einer Ordinalskala angeordnet, der absolute Wert hat im Gegensatz zu den Interpretationen als Ähnlichkeit oder Unsicherheit meist keine Bedeutung. In gewisser Hinsicht kann man Präferenzen auch als Ähnlichkeiten deuten, denn ähnliche Einzelheiten werden mit ähnlichen Präferenzwerten belegt. Präferenzen werden aber normalerweise nur auf einer endlichen Skala gemessen, die in vielen Fällen nicht numerisch ist. Zum Beispiel könnten Bewerber um eine Arbeitsstelle hinsichtlich verschiedener Kriterien wie Intelligenz, Kreativität oder Flexibilität über eine Skala *wenig*, *normal* und *stark ausgeprägt* bewertet werden. *Intelligent*, *kreativ* und *flexibel* wären dann unscharfe Mengen auf den Bewerbern.

Viele Anwendungen zielen darauf ab, eine Reihenfolge zu berechnen, nach der besser eingestufte Einzelheiten schlechter bewerteten vorzuziehen sind. Es geht hier weniger um die Vergrößerung von Wissen als um den unscharfen Vergleich von Einzelheiten.

In dieser Arbeit beschränken wir uns auf die Modellierung unscharfer Begriffe über Ähnlichkeit und Unsicherheit, da diese die wesentlichen Konzepte der unscharfen Vergrößerung von Wissen darstellen. Dabei werden wir die speziellen Ausprägungen von Ähnlichkeit und Unsicherheit mittels der Verfahren festlegen, mit denen wir die unscharfen Begriffe definieren.

## 2.2 Regelbasiertes Schließen

Schließen bedeutet, aus einer Wissensbasis mittels Inferenz Informationen abzuleiten, wie wir in Abschnitt 1.2 gesehen haben. Dazu nehmen wir an, das Universum der Anwendung sei durch Variablen beschrieben, deren Werte den Zustand des Universums wiedergeben.<sup>6</sup> In der **Wissensbasis** fassen wir alle allgemeingültigen Informationen über das Universum zusammen. Dabei handelt es sich erstens um Informationen über mögliche Zustände von Teilen des Universums, d. h. die Werte einiger Variablen. Zweitens enthält die Wissensbasis Zusammenhänge zwischen den Variablen, die man bei regelbasiertem Schließen als Regeln formuliert.

Neben der Wissensbasis liegen bei Anwendungen **Beobachtungen** vor, die den Zustand eines Teiluniversums beschreiben. Beobachtungen unterscheiden wir dadurch von Informationen in der Wissensbasis, wie in Abschnitt 1.2 dargestellt, dass sie keine allgemeingültigen Informationen über das Universum repräsentieren. Beobachtungen hängen von einer Größe ab, die außerhalb des Universums liegt. Eine **Inferenz** schließt aus den Informationen in der Wissensbasis und, falls vorhanden,

---

<sup>6</sup>vgl. Zustandsraumtheorie in der Physik

zusätzlichen Beobachtungen auf den Zustand der Teile des Universums, an denen wir interessiert sind.

Für die Darstellung der Zusammenhänge zwischen Variablen benutzen wir Regeln der Form „Wenn  $x = A$ , dann  $y = B$ “, die wir aus der Logik kennen;  $A$  und  $B$  sind in der Logik scharfe Mengen. Die Regel schränkt die Menge möglicher Werte der Variablen  $y$  ein, falls die Aussage „ $x = A$ “ erfüllt ist. In diesem Fall hängt also der Wert von  $y$  vom Wert von  $x$  ab. Wenn wir nun einen Wert  $A'$  der Variablen  $x$  wahrnehmen, dann können wir daraus einen Wert für  $y$  folgern. Dazu müssen wir feststellen, ob  $A'$  das **Antezedens** „ $x = A$ “ der Regel erfüllt, um daraus auf die Gültigkeit des **Konsequens** „ $y = B$ “ zu schließen. Dieses Ableiten neuer Informationen durch eine Verknüpfung von Beobachtungen mit der Wissensbasis stellt die Inferenz dar.

Als Basis für die Inferenz benutzen wir in dieser Arbeit den **Modus Ponens** (Tab. 2.1). Aus der Gültigkeit des Antezedens „ $x = A$ “ leiten wir die Gültig-

$$\begin{array}{l} \text{Wenn } x = A, \text{ dann } y = B. \\ \hline x = A \\ \hline y = B \end{array}$$

Tabelle 2.1: Modus Ponens.

keit des Konsequens „ $y = B$ “ ab. Diese Anforderung lässt immer noch Freiheitsgrade für die genaue Bedeutung der Regel „Wenn  $x = A$ , dann  $y = B$ “. Insbesondere trifft der Modus Ponens keine Aussage für den Fall, dass das Antezedens „ $x = A$ “ nicht erfüllt ist. In der Logik kann man die Regel als *Implikation*  $x = A \rightarrow y = B$  oder *Konjunktion*  $x = A \wedge y = B$  interpretieren, je nachdem, wie die Wahrheitswerte der Regel definiert werden, wenn das Antezedens nicht erfüllt ist. Darüber hinaus kann man eine Regel als partielle Abbildungsvorschrift  $A \mapsto B$  einer Funktion  $f: \mathcal{X} \rightarrow \mathcal{Y}$  auffassen oder als sichere bedingte Wahrscheinlichkeit  $P(y = B|x = A) = 1$  verstehen. Die Wahl einer Interpretation hängt stark von der Anwendung ab, also der Frage, in welcher Form Wissen über das Universum vorhanden ist.

Im Bereich des unscharfen, regelbasierten Schließens verarbeiten wir unscharfe Informationen. Die Regeln besitzen die Form „Wenn  $x = \tilde{A}$ , dann  $y = \tilde{B}$ “, wobei  $x$  und  $y$  Variablen sind, die unscharfe Werte wie  $\tilde{A}$  oder  $\tilde{B}$  annehmen können. Eine Inferenz wird im Fall unscharfer Werte schwieriger, weil die Antezedenzen u. U. nur partiell erfüllt sind. Wir können nicht eindeutig sagen, ob ein Antezedens gilt oder nicht, sondern bestimmen einen Grad der Gültigkeit. Entsprechend schwierig gestaltet sich auch die Berechnung der Inferenz, also die Bestimmung des Wertes von  $y$ , der i. A. ebenfalls unscharf ist.

Das meistbenutzte Verfahren für unscharfes Schließen ist das sogenannte **Prinzip von Kombination/Projektion**<sup>7</sup> auf Basis der *sup-min-Komposition* [Zad75, Zad79]. Hierbei wird eine Regel durch eine unscharfe Relation  $\tilde{R}$  auf  $\mathcal{X} \times \mathcal{Y}$  modelliert.  $\tilde{R}$  wird aus dem Antezedens  $\tilde{A}$  und dem Konsequens  $\tilde{B}$  berechnet. Die Berechnungsvorschrift hängt von der Semantik der Regel ab. Für den Schließvorgang wird die Information „ $x$  ist  $\tilde{A}$ “ zylindrisch auf  $\mathcal{X} \times \mathcal{Y}$  erweitert, mit der Regelrelation  $\tilde{R}$  geschnitten und auf den Ergebnisraum  $\mathcal{Y}$  projiziert, was im unscharfen Schluss  $\tilde{B}'$  resultiert:

$$\mu_{\tilde{B}'}(v) = \sup_{u \in \mathcal{X}} \min\{\mu_{\tilde{A}' \uparrow (\mathcal{X} \times \mathcal{Y})}(u, v), \mu_{\tilde{R}}(u, v)\} \quad (2.3)$$

$$= \sup_{u \in \mathcal{X}} \min\{\mu_{\tilde{A}'}(u), \mu_{\tilde{R}}(u, v)\} \quad (2.4)$$

Auf symbolischer Ebene schreiben wir für die Verknüpfung  $\tilde{B}' = \tilde{A}' \circ \tilde{R}$ . Wenn mehrere Regeln vorliegen, werden abhängig von ihrer Semantik alle Regeln zu einer Metaregel zusammengefasst und diese eine wird ausgewertet, oder es wird jede Regel einzeln ausgewertet und die Schlüsse aller Regeln werden aggregiert. Statt der Minimumbildung werden in (2.3) teilweise auch andere t-Normen als Schnittoperatoren benutzt.

Der meistbenutzte Spezialfall dieses Vorgehens ist der Regler von MAMDANI [MA75]. Eine Regel „Wenn  $x = \tilde{A}$ , dann  $y = \tilde{B}$ “ wird als Konjunktion der Aussagen interpretiert, was in der Regelrelation  $\mu_{\tilde{R}}(u, v) = \min\{\mu_{\tilde{A}}(u), \mu_{\tilde{B}}(v)\}$  resultiert. Wiederum wird hier die Minimumbildung als Schnittoperation benutzt. Im Fall von mehreren Regeln wird jede Regel einzeln nach dem Prinzip von Kombination/Projektion ausgewertet und die Schlüsse werden durch Vereinigung mit Maximumbildung zusammengefasst. Alternativ können wir alle Regelrelationen durch Maximumbildung zu einer Metaregel aggregieren und diese eine zur Inferenz benutzen. In der Praxis wird normalerweise die erste Vorgehensweise gewählt, weil der Rechenaufwand geringer ist und in MAMDANIS Fall die Schlüsse identisch sind.

Der Schluss einer Inferenz liegt beim Prinzip von Kombination/Projektion als unscharfe Menge  $\tilde{B}'$  vor, die den unscharfen Wert der Variablen  $y$  repräsentiert. In einigen Anwendungen benötigen wir aber ein scharfes Ergebnis. Wenn wir z. B. ein technisches System mit einem unscharfen Regelwerk regeln möchten, dann erwarten wir einen scharfen Wert für  $y$ , den Stellwert, den wir an das technische System anlegen. Zu diesem Zweck müssen wir den Schluss  $\tilde{B}'$  zu einem Wert  $y' \in \mathcal{Y}$  schärfen. Die Wahl des Verfahrens für die **Schärfung** ist stark anwendungsabhängig, da sie das funktionale Verhalten des unscharfen Systems beeinflusst. Die meistbenutzten Verfahren sind:

<sup>7</sup>engl. *principle of combination/projection* oder *compositional rule of inference*

- *Schwerpunktmethode*:  $y' = \frac{\int_{\mathcal{Y}} \mu_{\tilde{B}}(v) v dv}{\int_{\mathcal{Y}} \mu_{\tilde{B}}(v) dv}$
- *Maximummethode*:  $y' \in \{v \mid \mu_{\tilde{B}}(v) = \max_{v' \in \mathcal{Y}} \mu_{\tilde{B}}(v')\}$

Im Fall eines endlichen Universums ersetzen wir die Integrale bei der Schwerpunktmethode durch Summen. Bei der Maximummethode gehen wir davon aus, dass das Maximum existiert. Falls  $y'$  nicht eindeutig definiert ist, wird meist der Mittelwert der möglichen Werte gewählt.

Speziell im Bereich unscharfer Regelung ist der Regler von TAKAGI/SUGENO [TS84] weit verbreitet. Er besteht aus Regeln der Form „Wenn  $x = \tilde{A}_i$ , dann  $y = f_i(x)$ “. Diese Regeln stellen direkt einen scharfen Zusammenhang zwischen  $x$  und  $y$  her und machen so eine Schärfung überflüssig: Wenn  $x$  im unscharfen Bereich  $\tilde{A}_i$  liegt, dann wird  $y$  durch die scharfe Abbildung  $f_i$  festgelegt. In technischen Anwendungen haben sich solche Regeln als sinnvoll erwiesen, weil hier oft Wissen über den funktionalen Zusammenhang von  $x$  und  $y$  vorliegt, das in den Regler von MAMDANI nicht integriert werden kann. Das Verfahren setzt einen scharfen Wert für  $x$  voraus; das Ergebnis der Inferenz ist ein ebenfalls scharfer Wert  $y'$ :

$$y' = f(x) = \frac{\sum_i \mu_{\tilde{A}_i}(x) \cdot f_i(x)}{\sum_i \mu_{\tilde{A}_i}(x)} \quad (2.5)$$

Die Abbildung  $f$  wird aus den lokalen Abbildungen  $f_i$  zusammengesetzt, wobei die Lokalität durch  $\tilde{A}_i$  festgelegt wird. In den Bereichen, in denen sich die unscharfen Mengen  $\tilde{A}_i$  überlappen, berechnen wir  $f$  als gewichteten Mittelwert der  $f_i$ . In der Praxis wird meist der Spezialfall konstanter Funktionen  $f_i(x) = v_i$  verwendet, auf den wir in Abschnitt 6.3.2 im Zusammenhang mit unscharfer Interpolation zurückkommen.

## 2.3 Modellierung unscharfer Begriffe

Unscharfen Begriffen können verschiedene Semantiken unterlegt werden. In Abschnitt 2.1 wurden mit *Ähnlichkeit*, *Unsicherheit* und *Präferenz* die wichtigsten erläutert. Wie dort erwähnt, beziehen wir uns auf die ersten beiden Interpretationen von Vagheit, weil sie die wesentlichen Konzepte zur Vergrößerung von Wissen darstellen. Die in dieser Arbeit vorgestellte Theorie für das Schließen mit unscharfen Begriffen basiert auf der Interpretation von Vagheit als Unsicherheit. Im Nachhinein zeigen wir in Abschnitt 6.1 die Verwandtschaft von Unsicherheit und Ähnlichkeit, um die Flexibilität der Theorie zu untermauern.

### 2.3.1 Modell für unscharfe Begriffe

Die Modellierung unscharfer Begriffe stützen wir in dieser Arbeit auf das *Kontextmodell* [GK91, Geb92, GK93a, Bor00], das man als eine Verallgemeinerung der Wahrscheinlichkeitsrechnung verstehen kann. Das Kontextmodell erklärt, was die Zugehörigkeitsfunktion  $\mu_{\bar{A}}$  der unscharfen Menge bedeutet, die einen unscharfen Begriff formalisiert. Daraus ergeben sich konkrete Verfahren, um in Anwendungen die Zugehörigkeitsfunktionen festzulegen. Die Grundelemente aus der Wahrscheinlichkeitstheorie, die zum Verständnis benötigt werden, befinden sich in Anhang B.1.

Mathematische Grundlage unserer Betrachtungen ist ein Universum  $X$  bzw. dessen Potenzmenge  $2^X$ . Als Beispiel für ein Universum betrachten wir wieder die Menge  $X = [1cm, 300cm]$  möglicher Körpergrößen eines Mannes. Ein unscharfer Begriff wie *kleine Körpergröße* ist eine unscharfe Teilmenge von  $X$ . In unterschiedlichen Kontexten wird ein Begriff i. A. verschieden definiert. Einen **Kontext** interpretieren wir dabei als Quelle einer Definition. Insbesondere meinen wir damit nicht den Zusammenhang, in dem wir die Bedeutung eines Begriffs klären möchten. In Abschnitt 1.1.2 haben wir bei der Beschreibung von Unsicherheit bereits vorausgesetzt, dass der Zusammenhang fest ist. Im Beispiel der *kleinen Körpergröße* eines Mannes könnte man sich als Kontexte verschiedene Länder, Zeitalter oder verschiedene Personen vorstellen, die den Begriff *klein* definieren. Wenn wir die Bedeutung eines Begriffs herausfinden möchten, müssen wir uns alle Kontexte aus der jeweiligen Anwendung ansehen, in denen Wissen über den Begriff vorhanden ist. Eine allgemeingültige Definition des Begriffs sollte möglichst allen Kontexten gerecht werden.

Die Menge der Kontexte bezeichnen wir mit  $C$ . In jedem Kontext  $c \in C$  wird der gesuchte Begriff als scharfe Teilmenge von  $X$  definiert. Das geschieht über eine Abbildung  $\Gamma: C \rightarrow 2^X$ , die jedem  $c \in C$  eine Menge von Elementen aus  $X$  zuordnet, die im Kontext  $c$  als mögliche Definition des Begriffs angesehen werden. Die Mengen  $\Gamma(c)$  heißen **vage Charakteristika**; vage deshalb, weil mangels genauem Wissen der Begriff nicht als scharfer Wert  $u \in X$  sondern nur als Teilmenge von  $X$  definiert wird. Wenn wir als Beispiel für Kontexte eine Menge von Personen nehmen, dann wird jede Person  $c \in C$  eine Menge  $\Gamma(c)$  von Körpergrößen festlegen, die sie als *klein* einstuft. Die Abbildung  $\Gamma$  fasst also verschiedene Definitionen des gesuchten Begriffs zusammen.

In einem zweiten Schritt werden den Kontexten und damit den verschiedenen Bedeutungen des Begriffs Gewichte zugeordnet. Die Gewichte werden als Wahrscheinlichkeiten in einem **Kontextmaßraum**  $(C, 2^C, P_C)$  formalisiert. Ein Gewicht  $P_C(c)$  wird als das Zutrauen in die Korrektheit der Aussage von Kontext  $c$  oder als dessen Wichtigkeit angesehen [GK93a]. Eine frequentistische Interpre-

tation von  $P_C$  befindet sich in [Mat75]. Die vagen Charakteristika werden dort *zufällige Mengen* genannt, weil sie durch die Abbildung  $\Gamma$  an den Ausgang  $c$  eines Zufallsexperiments gekoppelt sind. In unserem Beispiel mit Personen als Kontexten werden wir z. B. eine Gleichverteilung für  $P_C$  ansetzen, wenn wir allen Personen in gleichem Maß vertrauen oder keine Informationen über das Vertrauen vorhanden sind<sup>8</sup>. Bei  $n$  befragten Personen erhält dann jede Person  $c$  das Gewicht  $P_C(c) = 1/n$ .

Über die Inverse der Funktion  $\Gamma$  übertragen wir die Wahrscheinlichkeiten  $P_C$  auf die zugehörigen vagen Charakteristika:

$$m(A) := P_C(\Gamma^{-1}(A)) \text{ für alle } A \in 2^X \quad (2.6)$$

mit  $\Gamma^{-1}(A) := \emptyset$ , falls  $A \notin \text{Bild}(\Gamma)$ . Die Abbildung  $m : 2^X \rightarrow [0, 1]$  heißt **Maßbasis**.  $m(A)$  misst das Zutrauen, dass die Menge  $A$  den Begriff beschreibt. Eine Maßbasis erfüllt folgende Bedingungen [Sha76]:

$$\sum_{A \in 2^X} m(A) = 1 \text{ und } m(\emptyset) = 0 \quad (2.7)$$

Die Mengen  $A \in 2^X$  mit positivem Zutrauen heißen **fokale Elemente**. Aus der Maßbasis konstruiert man ein **Plausibilitätsmaß** durch die Vorschrift

$$Pl(A) := \sum_{\substack{B \in 2^X: \\ B \cap A \neq \emptyset}} m(B) = \sum_{\substack{B \in \text{Bild}(\Gamma): \\ B \cap A \neq \emptyset}} P_C(\Gamma^{-1}(B)) \quad (2.8)$$

Ein Plausibilitätsmaß  $Pl$  ist ein unscharfes Maß, vgl. Anhang B.1, und misst die Plausibilität der Aussage  $u \in A$ . Damit definieren wir eine **Plausibilitätsdichte**

$$pl(u) := Pl(\{u\}) = P_C(\{c \mid u \in \Gamma(c)\}), \quad (2.9)$$

welche die Plausibilität eines einzelnen Wertes  $u$  misst. Plausibilitätsmaß und Plausibilitätsdichte abstrahieren Informationen über die Kontexte. Man kann an der Plausibilität  $pl(u)$  nur ablesen, wie groß die Wahrscheinlichkeit der Kontexte ist, in denen  $x = u$  plausibel ist. Qualitative Informationen gehen verloren; so kann man nicht mehr rekonstruieren, welcher Kontext für welche Information verantwortlich ist. Das bedeutet insbesondere, dass die Abbildung der vagen Charakteristika  $\Gamma(c)$  auf die Plausibilitätsdichte nicht injektiv ist: Verschiedene Abbildungen  $\Gamma$  können dieselbe Plausibilitätsdichte erzeugen. In dieser Hinsicht komprimiert eine Plausibilitätsdichte Informationen.

Falls die vagen Charakteristika scharfe Werte sind, d. h. für alle  $c \in C$  gilt  $\Gamma(c) = \{u_c\}$  mit  $u_c \in X$ , dann ist die Plausibilitätsdichte eine Wahrscheinlichkeitsdichte.

<sup>8</sup>*Inufficient reason principle*: Bei Unwissen wird eine Gleichverteilung angenommen.

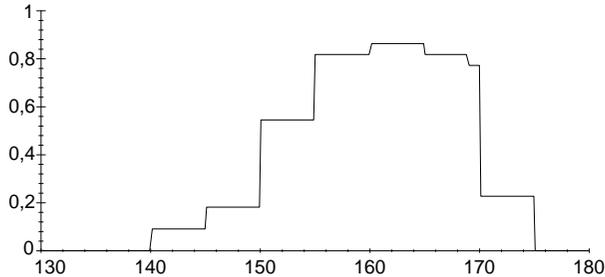


Abbildung 2.2: Plausibilitätsdichte für *kleine* Körpergrößen erwachsener Männer

Je kleiner also die Mengen  $\Gamma(c)$  werden, desto präziser wird das Wissen und desto mehr nähern wir uns einer Wahrscheinlichkeitsdichte an. Insbesondere kann man beweisen, dass eine Plausibilitätsdichte eine obere Schranke für eine Wahrscheinlichkeitsdichte ist.

Wie bereits angedeutet, verwenden wir in großen Teilen der Arbeit das Kontextmodell als konkrete Semantik unscharfer Mengen. Insbesondere wird damit ein unscharfer Begriff  $\tilde{B}$  durch eine Plausibilitätsdichte  $pl_{\tilde{B}}$  definiert. In der Possibilitätstheorie deutet man diese Identifikation folgendermaßen: Unter der Annahme, dass die Variable  $x$  den unscharfen Wert  $\tilde{B}$  annimmt, gibt  $pl_{\tilde{B}}(u)$  die Plausibilität von  $x = u$  an.

Abbildung 2.2 zeigt die Plausibilitätsdichte der Variablen *Körpergröße* unter der Annahme, dass ein *kleiner* erwachsener Mann beschrieben wird. Das Ergebnis geht aus einer Umfrage mit 22 Personen hervor. Dabei definiert jede Person als Kontext eine Menge von Körpergrößen, die sie mit *klein* bezeichnet. Zum Verständnis der Abbildung muss erwähnt werden, dass die Personen gleichzeitig noch die Begriffe *sehr klein*, *normal*, *groß* und *sehr groß* definieren sollten. Die größte Plausibilität liegt mit etwa 0,86 im Intervall  $[160\text{cm}, 164\text{cm}]$ . Dass an keiner Stelle eine Plausibilität von eins vorliegt, bedeutet, dass es keine Körpergröße gibt, die alle Befragten als *klein* bezeichnen. Lediglich 86% der Befragten konnten sich auf obiges Intervall einigen. Die Unschärfe des Begriffs *klein* entsteht also durch eine Abstraktion über der Menge von Kontexten, in diesem Fall der befragten Personen.

Alternativ zu der Interpretation eines unscharfen Begriffs als Plausibilitätsdichte verwenden wir sogenannte **Likelihood-Funktionen**. Eine Likelihood-Funktion  $L(\tilde{B}; u) = P(\tilde{B}|u)$  misst die Wahrscheinlichkeit des unscharfen Begriffs  $\tilde{B}$  unter der Annahme, dass  $x = u$ . Wir werden nun zeigen, dass wir Plausibilitätsdichten und Likelihood-Funktionen miteinander identifizieren können:  $pl_{\tilde{B}}(u) = P(\tilde{B}|u)$ . Dazu gehen wir wiederum von einer Kontextmenge  $C$  aus. Zu einem gegebenen  $u \in X$

suchen wir alle Kontexte  $c \in \mathcal{C}$ , in denen  $u$  als definierender Wert des Begriffs gilt. Anstatt die Personen aus unserem Beispiel zu bitten, eine scharfe Menge für kleine Körpergrößen zu definieren, könnten wir sie bei jeder möglichen Körpergröße fragen, ob sie diese als *klein* bezeichnen würden oder nicht. Wir halten also  $u$  fest und erhalten für jeden Kontext  $c$  die Information, ob  $u \in \Gamma(c)$ . In dieser Weise konstruieren wir die Abbildung  $\Gamma$ . Für die Wahrscheinlichkeit oder Likelihood gilt dann  $P(\tilde{B}|u) = P_{\mathcal{C}}(\{c \mid u \in \Gamma(c)\})$ , was nach (2.9) der Definition einer Plausibilitätsdichte entspricht.

Wenn man davon absieht, dass diese Konstruktion von  $\Gamma$  nicht immer praktisch durchführbar ist, sehen wir doch die Verwandtschaft der Konzepte Kontextmodell und Likelihood. So könnte man bei kontinuierlichen Universen wie den möglichen Körpergrößen zwar praktisch nicht nach jeder Körpergröße fragen, die Äquivalenz beider Modelle ist aber trotzdem unmittelbar klar. Im größten Teil dieser Arbeit verwenden wir Plausibilitätsdichten, um unscharfe Begriffe zu beschreiben. An einigen Stellen ist jedoch die Darstellung durch Likelihood-Funktionen um einiges einfacher. In der Literatur findet man die Interpretation unscharfer Mengen als Likelihood-Funktionen beispielsweise in [His88, DMP93, Tho95].

### 2.3.2 Operationen auf unscharfen Begriffen

Wichtig für den weiteren Verlauf der Arbeit sind Operationen auf unscharfen Begriffen, also Schnitt, Vereinigung und Komplementbildung der definierenden unscharfen Mengen. In Abschnitt 2.1 haben wir mit t-Norm, co-t-Norm und Komplementfunktion bereits Klassen von Operatoren kennengelernt, mit denen unscharfe Mengen verknüpft werden können. Da diese Arbeit auf der Interpretation unscharfer Mengen als Plausibilitätsdichten aufbaut, definieren wir die Operationen hier für Plausibilitätsdichten.

Für diese Betrachtung gehen wir davon aus, dass die Plausibilitätsdichten aller Begriffe auf demselben Kontextmaßraum beruhen. Wir folgen hier den Definitionen in [KKG95], bei denen die drei Operationen für jeden Kontext  $c$  auf Durchschnitt, Vereinigung und Komplementbildung der vagen Charakteristika  $\Gamma(c)$  zurückgeführt werden. Beim Komplement beispielsweise wird ein Kontext  $c$  nicht mehr mit der Menge  $A = \Gamma(c)$  verknüpft, sondern mit ihrem Komplement  $A^c = \Gamma(c)^c = \mathcal{X} \setminus \Gamma(c)$ . Wenn z. B. eine Person das Intervall  $[155cm, 170cm]$  für *kleine* Körpergrößen angibt, so liegen *nicht kleine* Körpergrößen aus ihrer Sicht in der Menge  $[1cm, 300cm] \setminus [155cm, 170cm] = [1cm, 155cm) \cup (170cm, 300cm]$ .

**Definition 2** Seien  $\tilde{A}$  und  $\tilde{B}$  unscharfe Begriffe über dem Kontextmaßraum  $(\mathcal{C}, 2^{\mathcal{C}}, P_{\mathcal{C}})$ . Dann werden Schnitt, Vereinigung und Komplement wie folgt definiert:

- **Schnitt**  $\tilde{A} \cap \tilde{B}$ :  $\Gamma_{\tilde{A} \cap \tilde{B}}(c) := \Gamma_{\tilde{A}}(c) \cap \Gamma_{\tilde{B}}(c)$
- **Vereinigung**  $\tilde{A} \cup \tilde{B}$ :  $\Gamma_{\tilde{A} \cup \tilde{B}}(c) := \Gamma_{\tilde{A}}(c) \cup \Gamma_{\tilde{B}}(c)$
- **Komplement**  $\tilde{A}^c$ :  $\Gamma_{\tilde{A}^c}(c) := \Gamma_{\tilde{A}}(c)^c$

Für die Verknüpfung der zugehörigen Plausibilitätsdichten gilt damit

**Satz 1** Für Schnitt, Vereinigung und Komplement von Begriffen gelten

- $\max\{0, pl_{\tilde{A}}(u) + pl_{\tilde{B}}(u) - 1\} \leq pl_{\tilde{A} \cap \tilde{B}}(u) \leq \min\{pl_{\tilde{A}}(u), pl_{\tilde{B}}(u)\}$
- $\min\{1, pl_{\tilde{A}}(u) + pl_{\tilde{B}}(u)\} \leq pl_{\tilde{A} \cup \tilde{B}}(u) \leq \max\{pl_{\tilde{A}}(u), pl_{\tilde{B}}(u)\}$
- $pl_{\tilde{A}^c}(u) = 1 - pl_{\tilde{A}}(u)$

Die Beweise findet man z. B. in [KKG95]. Direkt aus Definition 2 folgt ein Spezialfall von Satz 1, der im weiteren Verlauf benötigt wird.

**Korollar 2** Seien  $\tilde{A}$  und  $\tilde{B}$  unscharfe Begriffe über dem Kontextmaßraum  $(C, 2^C, P_C)$ . Falls die Mengen  $\Gamma_{\tilde{A}}(c)$  und  $\Gamma_{\tilde{B}}(c)$  für alle Kontexte  $c \in C$  disjunkt sind, gelten:

- $pl_{\tilde{A} \cap \tilde{B}} \equiv 0$
- $pl_{\tilde{A} \cup \tilde{B}} \equiv pl_{\tilde{A}} + pl_{\tilde{B}}$

Insbesondere die leere Schnittmenge erscheint zunächst nicht intuitiv, wenn man Plausibilitätsdichten z. B. von  $\tilde{A}$  und  $\tilde{A}^c$  betrachtet. Die beiden Funktionen überlappen sich offensichtlich, weshalb man einem  $u \in X$  im Überlappungsbereich mit Plausibilitätsgrad  $pl_{\tilde{A}}(u)$  den Begriff  $\tilde{A}$  und entsprechend mit Grad  $pl_{\tilde{A}^c}(u) = 1 - pl_{\tilde{A}}(u)$  das Komplement  $\tilde{A}^c$  zuordnet. Aus dieser Sicht erwartet man, dass zumindest mit gewissem Plausibilitätsgrad eine unscharfe Schnittmenge existiert, so wie das in der klassischen Theorie unscharfer Mengen der Fall ist (Minimumbildung der Zugehörigkeitsfunktionen). Auf der anderen Seite jedoch erklärt jeder Kontext  $c$  die Begriffe  $\tilde{A}$  und  $\tilde{A}^c$  für disjunkt; innerhalb des Kontextmodells existiert keine Evidenz für eine nichtleere Schnittmenge. Wir geben daher

**Definition 3** Seien  $\tilde{A}$  und  $\tilde{B}$  unscharfe Begriffe über dem Kontextmaßraum  $(C, 2^C, P_C)$ .  $\tilde{A}$  und  $\tilde{B}$  heißen **disjunkt** genau dann, wenn  $\Gamma_{\tilde{A}}(c)$  und  $\Gamma_{\tilde{B}}(c)$  für alle  $c \in C$  disjunkt sind.

Zwei disjunkte Begriffe  $\tilde{A}$  und  $\tilde{B}$  werden also nach Korollar 2 verknüpft. Man sieht leicht, dass Komplementbildung, Schnitt- und Vereinigungsoperation disjunkter Begriffe den Axiomen unscharfer Mengenoperationen gehorchen; insbesondere ist die Schnittoperation eine t-Norm und die Vereinigungsoperation eine co-t-Norm.

## Kapitel 3

---

# Stand der Technik

Wir geben hier einen Überblick über Ansätze für regelbasiertes Schließen im Bereich unscharfer Informationsverarbeitung und beurteilen sie bezüglich der gestellten Ziele und Anforderungen, die in Abschnitt 1.3 erläutert wurden:

1. Mehrstufiges unscharfes Schließen
2. Effizienz hinsichtlich Speicher- und Rechenaufwand
3. Verständlichkeit und Exaktheit unscharfer Ergebnisse
4. Integration von Informationen verschiedener Semantiken

Bis auf die Ansätze aus dem Bereich *spezieller Funktionsapproximatoren* beschränken wir uns von vornherein auf Verfahren, die mehrstufiges unscharfes Schließen erlauben, also die erste Anforderung erfüllen.

Wir werden die Ansätze in drei Klassen einteilen, die beschreiben, wie eine mehrstufige Inferenz berechnet wird, und so direkt Bezug auf Anforderung 2 nehmen.

V1: Mehrstufiges Schließen auf grobgranularer Ebene

V2: Einstufiges Schließen auf grobgranularer Ebene, vor Einsatz der nächsten Stufe Schritt auf feingranularer Ebene

V3: Mehrstufiges Schließen auf feingranularer Ebene

Ein Ziel der Arbeit ist, zu einem V1-Verfahren zu gelangen. Ein Verfahren, das mehrstufiges Schließen auf grobgranularer Ebene ermöglicht, ist sehr effizient, weil das Wissen durch wenige, grobgranulare Informationseinheiten repräsentiert und auf dieser Ebene auch verarbeitet wird. Rechen- und Speicheraufwand eines V1-Verfahrens hängen damit nur von der Anzahl der grobgranularen Informationen ab. Bei V2-Verfahren ist vor dem Einsatz jeder Regelstufe ein teurer Schritt auf feingranularer Ebene nötig und V3-Verfahren operieren ausschließlich auf feingranularer Ebene, weshalb sie extrem aufwändig sind.

Für eine Beurteilung der Ansätze bzgl. der Anforderungen 3 und 4 benötigen wir noch eine Klassifikation verschiedener Arten der Semantik unscharfer Begriffe bzw. Mengen. In den meisten theoretischen Arbeiten wird nicht definiert, was der Zugehörigkeitsgrad zu einer unscharfen Menge konkret bedeutet;

als eher philosophischer Ansatz werden vage Bedeutungen wie *Wahrheitsgrad* [Zad75, Haj99, DP91b] oder *Possibilität* [Zad78, DP88, DP91a] angenommen. Auf Basis dieser vagen Bedeutungen werden Axiome definiert, die für die Verknüpfung von Zugehörigkeitsgraden gelten sollen. Was aber z. B. die Zahl 0,7 als Zugehörigkeitsgrad bedeutet und wo sie herkommt, bleibt offen. Wir bezeichnen solche Erklärungen von Zugehörigkeitsgraden als **abstrakte Semantik**. Im Gegensatz dazu stehen Ansätze wie das Kontextmodell aus Abschnitt 2.3, welche die Bedeutung der Zugehörigkeitsgrade konkret festlegen; wir nennen dies im weiteren Verlauf **konkrete Semantik** unscharfer Mengen.

Für Anforderung 3 prüfen wir, ob die unscharfen Schlüsse der Verfahren gleichzeitig auf grobgranularer Ebene (Verständlichkeit) und feingranularer Ebene (Exaktheit) dargestellt werden. Dabei verlangen wir, dass die unterliegenden Theorien auf einer konkreten Semantik beruhen und die Beziehung zwischen grob- und feingranularer Ebene wohldefiniert ist. Schließlich untersuchen wir für Anforderung 4, ob die Ansätze Informationen verarbeiten können, die sich hinsichtlich der Art und Ordnung der Unschärfe oder der Form der Beschreibung unterscheiden. Dabei müssen diese verschiedenen Qualitäten von Informationen konkrete Semantiken aufweisen, damit wir deren korrekte Integration sicherstellen können.

In der Literatur<sup>1</sup> über regelbasiertes unscharfes Schließen wird zwischen *unscharfen Produktionsregeln*<sup>2</sup> und Verfahren unterschieden, die auf dem *Prinzip von Kombination/Projektion* beruhen. Das Prinzip von Kombination/Projektion haben wir bereits in Abschnitt 2.2 als klassisches Verfahren für unscharfes Schließen besprochen. Insbesondere die Effizienzprobleme werden mit unscharfen Produktionsregeln gemindert, wie wir unten sehen werden. Darüber hinaus betrachten wir den Regler von MAMDANI als Sonderfall der beiden genannten Ansätze sowie *interpolierendes unscharfes Schließen* und *spezielle Funktionsapproximatoren*, die insbesondere im Bereich unscharfer Regelung benutzt werden.

**Prinzip von Kombination/Projektion (PKP)** [Zad75, Zad79] Regeln werden auf grobgranularer Ebene formuliert, aber auf feingranularer Ebene als unscharfe Relationen auf dem kartesischen Produkt der zugrunde liegenden Universen formalisiert. Die Modellierung bewegt sich damit auf feingranularer Ebene. Für die Berechnung des unscharfen Schlusses mit der sup-min-Komposition (2.3) bildet man das Supremum über den Einzelheiten  $u \in \mathcal{X}$ . Dieser Inferenzvorgang spielt sich also ebenfalls auf feingranularer Ebene ab. Damit gehören diese Verfahren zur Klasse V3. Das unscharfe Ergebnis einer Inferenz ist außerdem nicht auf grobgranularer Ebene beschreibbar. Es liegt als unscharfe Menge auf feingranularer Ebene vor, die von den unscharfen Begriffen der Regelkonsequenzen losgelöst

<sup>1</sup>siehe z. B. [DP91a, YT97]

<sup>2</sup>engl. *fuzzy production rules*

ist; Anforderung 3 wird nicht erfüllt. Die Verfahren des PKP waren zunächst nur mit einer abstrakten Semantik ausgestattet. So lässt sich in der Possibilitätstheorie z. B. die sup-min-Komposition mit GÖDELS Inferenzrelation [Got89] auf abstrakter Ebene eindeutig als Inferenzmechanismus herleiten [Spo94, Spo99]. Unter Annahme der sup-min-Komposition als Inferenzmechanismus gelingt mit dem *Kontextmodell* die Herleitung der Gödelrelation auf der Basis einer konkreten Semantik [Geb92, GK93b]. Insgesamt ist uns aber keine vernünftige Erklärung des Gesamtmechanismus auf der Basis einer konkreten Semantik bekannt, womit Anforderung 4 fällt.

**Unschärfe Produktionsregeln (UPR)** Eine Inferenz auf der Basis unscharfer Produktionsregeln besteht aus vier Schritten:

1. Mustervergleich<sup>3</sup>: Vergleich einer unscharfen Eingabe mit den Antezedenzen der Regeln
2. Regelauswahl: Wahl der Regeln, deren Antezedens „genügend“ mit der Eingabe übereinstimmt
3. Feuern: partielles Aktivieren der Konsequenzen der gewählten Regeln
4. Adaption: Die Konsequenzen der Regeln werden abhängig vom Aktivierungsgrad modifiziert und zusammengefasst.

Der Mustervergleich einer unscharfen Eingabe  $\tilde{A}'$  mit einem Antezedens  $\tilde{A}$  resultiert in einem Skalar. Dieser kann die Übereinstimmung von  $\tilde{A}'$  und  $\tilde{A}$  hinsichtlich verschiedener Kriterien messen; gebräuchlich sind *Konsistenz*, z. B. eine Form von Möglichkeit  $\Pi(\tilde{A}|\tilde{A}')$  [CFP82], und *Ähnlichkeit*, wie etwa eine unscharfe Version des Hausdorff-Abstands [DP83]. Die Berechnung des Skalars wird auf feingranularer Ebene durchgeführt. Anschließend werden alle Regeln herausgesucht, bei denen die Übereinstimmung des Antezedens mit der unscharfen Eingabe einen Schwellenwert, im Zweifelsfall null, überschreitet. Die zugehörigen Konsequenzen werden über den Grad der Übereinstimmung modifiziert und zusammengefasst (Konsistenzverfahren [MS89], Ähnlichkeit [DP83]). Dieser Inferenzvorgang spielt sich auf grobgranularer Ebene ab, da pro Regel nur ein Skalar von dem Antezedens zum Konsequens gereicht wird. Am Eingang der nächsten Regelschicht muss aber wieder eine aufwändige Messung der Übereinstimmungen auf feingranularer Ebene durchgeführt werden. Damit gehören die meisten Verfahren auf der Basis von Produktionsregeln zur Klasse der V2-Verfahren. Eine Ausnahme ist der Regler von MAMDANI, bei dem der Übergang von einer zur nächsten Schicht auf grobgranularer Ebene berechnet werden kann [CP96]. Bei Anforderung 3 unterscheiden sich verschiedene Methoden von UPR. Bei dem Ansatz in [YT97] wird nur eine

---

<sup>3</sup>engl. *pattern matching*

heuristische Interpretation der Schlüsse auf grobgranularer Ebene ohne jegliche Semantik angegeben; das gilt ebenso für die Berechnung des Schlusses auf feingranularer Ebene. Beim Regler von MAMDANI ergeben sich exakte Schlüsse auf feingranularer Ebene, die auf grobgranularer Ebene nicht sinnvoll gedeutet werden können. Sehr interessant ist der Ansatz in [MS89], der Anforderung 3 nahezu erfüllt. Er basiert aber auf einer axiomatisch definierten, abstrakten Semantik, deren Relevanz für praktische Anwendungen unklar ist. Daher wird Anforderung 4 nicht erfüllt.

**Regler von MAMDANI [MA75]** Dieser Ansatz nimmt insofern eine Sonderstellung ein, als er sowohl dem Prinzip von Kombination/Projektion als auch den Produktionsregeln zugeordnet werden kann. Zudem gehört er zu den V1-Verfahren, wie die Ergebnisse in [CP96] zeigen. Die unscharfen Schlüsse sind nicht auf grobgranularer Ebene interpretierbar, wie bereits bei den UPR erwähnt wurde. Der Ansatz von MAMDANI lässt sich mit abstrakter Semantik herleiten [Wei95a, Wei95b] und zumindest approximativ mit konkreter Semantik belegen (*Gleichheitsrelationen*) [KK93, KKG95].

**Interpolierendes unscharfes Schließen** Ein V1-Verfahren ist der Ansatz in [SH98], der unscharfes Schließen auf der Basis von Interpolation<sup>4</sup> [KH93] und Analogien<sup>5</sup> realisiert. Dabei wird ein Abstand zwischen der unscharfen Eingabe  $\tilde{A}'$  und dem Antezedens  $\tilde{A}$  gemessen und auf die Seite des Konsequens übertragen. Die Berechnung nimmt auch Rücksicht auf die Unschärfe der benutzten Informationen. Unklar bleibt, wie eine Inferenz mit mehreren Regeln berechnet wird. Die Autoren deuten lediglich für zwei Regeln eine Art linearer Interpolation an. Unschärfe Schlüsse liegen nur auf feingranularer Ebene vor, Anforderung 3 wird nicht erfüllt. Weiterhin wird den unscharfen Mengen keinerlei Semantik unterlegt. Die Definition des Abstandsmaßes weist zwar auf die implizite Annahme einer Semantik auf Basis von Ähnlichkeiten hin; ob und wie sich auf diese Weise aber der Inferenzmechanismus herleiten lässt, bleibt offen. Damit wird Anforderung 4 verletzt.

**Spezielle Funktionsapproximatoren** Zu dieser Klasse gehören der Regler von TAKAGI/SUGENO [TS84], radiale Basisfunktionen<sup>6</sup> [Pow87, PG90], *Fuzzy Basis Functions* [WM92, ZS94, KM95], ANFIS [Jan92, Jan93], die Sugeno-Variante von NEFPROX [NK99] und Spline-basierte Systeme [BH94, BH95, ZK96, ZK99]. Wir bezeichnen sie als Pseudo-V1-Verfahren, da sie vom Rechenaufwand her in

---

<sup>4</sup>engl. *interpolative reasoning*

<sup>5</sup>engl. *analogical reasoning*

<sup>6</sup>engl. *radial basis functions (RBFs)*

die Klasse V1 eingeordnet werden können; sie realisieren allerdings kein unscharfes Schließen. Es handelt sich um spezielle Approximatoren scharfer Funktionen, die zwar als regelbasiert interpretiert werden können, aber über keine konkrete Semantik verfügen (rein funktionale Ansätze). Im Bereich unscharfer Regelung verdrängen die Funktionsapproximatoren in immer stärkerem Maße den Regler von MAMDANI, weil dessen Schließverhalten einige Anomalien aufweist [Mik99, LH95]. Weiterhin ist MAMDANIS Ansatz, wie auch die anderen Ansätze, die auf dem Prinzip von Kombination/Projektion basieren oder zur Klasse der Produktionsregeln gehören, ungeeignet, um gegebene Daten zu approximieren. Die Bedeutung der Funktionsapproximatoren im Bereich unscharfer Regelung qualifiziert sie trotz mangelhafter Semantik als Ansatz unscharfer Informationsverarbeitung. Wir werden sehen, dass der Ansatz dieser Arbeit als eine Verallgemeinerung dieser Verfahren verstanden werden kann.

|                                   | PKP | UPR | MAMDANI | Interpol. | Approx. |
|-----------------------------------|-----|-----|---------|-----------|---------|
| mehrstufiges Schließen            | +   | +   | +       | +         | –       |
| Effizienz                         | –   | o   | +       | +         | +       |
| Verständlichkeit<br>und Exaktheit | o   | o   | o       | –         | –       |
| Semantik                          | o   | o   | o       | –         | –       |

Tabelle 3.1: Leistungen verschiedener Ansätze im Vergleich

Tabelle 3.1 fasst noch einmal die vorgestellten Ansätze hinsichtlich der Erfüllung der gestellten Anforderungen zusammen. „+“ bedeutet *erfüllt*, „o“ *teilweise erfüllt* und „–“ heißt *nicht erfüllt*. Bei Effizienz steht „+“ für ein V1-Verfahren, „o“ für V2 und „–“ für V3. Eine konkrete Semantik erhält in der vierten Zeile ein „+“, eine abstrakte Semantik „o“ und keine Semantik „–“. Der Ansatz von MAMDANI wird insgesamt am besten bewertet. Auch dies verdeutlicht, warum das Verfahren so populär ist.

Insgesamt ist festzustellen, dass keiner der bekannten Ansätze mehr als zwei der vier Anforderungen befriedigend erfüllt. An dieser Stelle setzt unsere Theorie an, die effizientes mehrstufiges Schließen erlaubt, auf einer konkreten, praxisrelevanten Semantik basiert und gleichermaßen verständliche wie exakte Ergebnisse liefert.



## Kapitel 4

---

# Modellierung granularen Wissens

Das Ziel dieses Abschnitts ist, die Granularisierung oder Vergrößerung von Wissen zu formalisieren. Dabei unterscheiden wir zwei Aspekte: die einfache Grobheit von Informationen und die Partitionierung von Universen. Zunächst einmal können wir einer Variablen einen unscharfen Wert in Form einer unscharfen Menge zuordnen. Wir geben uns mit dem groben, unscharfen Wert zufrieden, wenn wir keine genauere Aussage treffen können oder eine präzisere Information in der Anwendung irrelevant ist. Diese Aspekte haben wir in Kapitel 2 diskutiert. Wenn wir aber feststellen, dass wir in einer Anwendung nur wenige unscharfe Werte einer Variablen unterscheiden können oder möchten, dann vergrößern oder partitionieren wir das Universum, also die Menge möglicher Werte der Variablen. So kann die Körpergröße nicht mehr beliebige unscharfe Werte als unscharfe Mengen über  $[1\text{cm}, 300\text{cm}]$  annehmen, sondern nur noch in einer Menge unscharfer Begriffe wie  $\{\textit{sehr klein}, \textit{klein}, \textit{normal}, \textit{groß}, \textit{sehr groß}\}$  liegen. Wissen über den Wert der Variablen formulieren wir dann ausschließlich mit den unscharfen Begriffen.

Für die Verständlichkeit und Exaktheit unscharfer Informationen ist es entscheidend, die Beziehung zwischen den Darstellungen auf grobgranularer und feingranularer Ebene klar zu definieren. Die unscharfen Grundbegriffe einer Partition spiegeln gewissermaßen den Wortschatz des Entwicklers wider, wenn er über Werte von Variablen spricht. Da ein unscharfer Begriff formal durch eine unscharfe Menge dargestellt wird, liegt der unscharfe Wert der Variablen gleichzeitig auf grobgranularer Ebene als unscharfer Begriff und auf feingranularer Ebene als unscharfe Menge vor. Eine Aussage der Form „Variable  $x$  hat den unscharfen Wert  $\tilde{A}$ .“ ist einerseits verständlich, da wir  $\tilde{A}$  einen Begriff aus unserem Wortschatz zuordnen können, und andererseits wohldefiniert, da wir seine Bedeutung auf feingranularer Ebene kennen.

Liegt aber eine beliebige unscharfe Beobachtung vor, so kennen wir nur die unscharfe Menge auf feingranularer Ebene. Für diesen Fall müssen wir eine Interpretation auf grobgranularer Ebene finden, die unscharfe Begriffe aus unserem Wortschatz benutzt.

In diesem Kapitel widmen wir uns dem Problem, wie unscharfe Begriffe und unscharfe Mengen bzw. unscharfe Aussagen auf grobgranularer und feingranularer Ebene semantisch zusammenhängen. In der Literatur wird die Bidirektionalität des

Zusammenhangs zwischen grobgranularer und feingranularer Darstellung bisher nicht betrachtet. Im Allgemeinen wird Wissen auf grobgranularer Ebene entworfen und dann auf feingranularer Ebene verarbeitet. So wird zwar der Entwurf auf der verständlichen Ebene unscharfer Begriffe durchgeführt, doch geht die Verständlichkeit während der Verarbeitung verloren. Der Schritt zurück auf die grobgranulare Ebene ist nicht vorgesehen.

Die Grundidee aller weiteren Betrachtungen ist die *Kombination von Begriffen*, die wir auch in der alltäglichen Sprache benutzen. Dort treffen wir nichteindeutige Zuordnungen wie „Der Mann war *groß* bis *sehr groß*.“ oder „Die Haarfarbe ist weniger *hellblond*, eher *mittelblond*.“ Wir beschreiben Beobachtungen also durch die Kombination von unscharfen Begriffen, welche in dieser Weise unseren *Basiswortschatz* bilden. Dabei versehen wir unscharfe Begriffe aus dem Basiswortschatz mit Gewichtungen wie *eher* oder *weniger* und kombinieren sie zu neuen unscharfen Werten. Formal gesehen bildet der Basiswortschatz auf diese Weise ein Erzeugendensystem für unscharfe Werte.

Wir werden in diesem Kapitel klären, wie wir Gewichtungen und die Kombinationsoperatoren mathematisch formalisieren. Als Ergebnis erhalten wir ein semantisch fundiertes Verfahren, das aus einer verständlichen Kombination unscharfer Begriffe einen unscharfen Wert in Form einer unscharfen Menge erzeugt. Wir bezeichnen diesen Vorgang als *Synthese* einer unscharfen Menge. Umgekehrt geben wir Verfahren an, um aus einer unscharfen Menge eine Kombination von Begriffen abzuleiten, was dann der Interpretation einer unscharfen Beobachtung entspricht. Wir nennen dies *Analyse* einer unscharfen Menge.

Synthese und Analyse sind grundlegend neue Ansätze für die Darstellung unscharfer Informationen. Wir werden im weiteren Verlauf der Arbeit sehen, dass sie den Grundstein für die Verständlichkeit unscharfer Schlüsse und die Effizienz unscharfer Informationsverarbeitung legen.

In Abschnitt 4.1 zeigen wir, wie man Universen partitioniert. Wir bauen das Verfahren auf den Anforderungen *Vollständigkeit* und *Konsistenz* der Partition auf, die wir vom scharfen auf den unscharfen Fall übertragen. Daraus resultiert die elementare Definition eines *unscharfen Grundbegriffs*. Anschließend formalisieren wir in Abschnitt 4.2 die Kombination von Grundbegriffen, insbesondere definieren wir die Synthese und Analyse unscharfer Mengen.

## 4.1 Partitionierung von Universen

Einer der zentralen Punkte dieser Arbeit ist die Partitionierung von Universen durch unscharfe Begriffe. Als Beispiel betrachten wir wiederum mögliche Körper-

größen erwachsener Männer im Bereich  $[1cm, 300cm]$ . Für viele Entscheidungen reicht es aus, die Körpergröße nur grob zu kennen. Beim Autofahren werden wir z. B. *kleine* Personen auf der Rückbank und *große* Personen auf den Vordersitzen platzieren. Wir teilen also die Grundmenge  $[1cm, 300cm]$  in Bereiche auf, die in unserer Anwendung mit Bedeutungen belegt sind. Für Personenbeschreibungen werden wir etwa fünf Bereiche erwarten, die wir mit den unscharfen Begriffen *sehr klein*, *klein*, *normal*, *groß* und *sehr groß* bezeichnen können. Diese Begriffe bilden die Partition des Universums und gleichzeitig den Grundwortschatz, das Vokabular, auf dessen Basis Schlüsse gezogen werden. Grundsätzlich nehmen wir an, dass die Feinheit der Partition der Anwendung angemessen ist, wobei wir nur endliche Partitionen betrachten.

Wir stellen an eine Partition folgende Anforderungen. Sie soll ein Universum so vergrößern, dass

1. das ganze Universum erfasst wird: Jede mögliche feingranulare Aussage muss von der Partition eingefangen werden.
2. maximale Differenzierung gewährleistet wird: Aussagen, die auf grobgranularer Ebene unterscheidbar sind, müssen durch die Partition getrennt werden.

Die Anforderungen lassen sich formal durch zwei Eigenschaften beschreiben: *Vollständigkeit* und *Konsistenz*. Im Fall einer scharfen Einteilung eines Universums in Teilbereiche bedeutet das:

- Eine Einteilung ist **vollständig**, wenn sie den gesamten Wertebereich überdeckt, d. h. wenn die Vereinigung der Teilbereiche die Grundmenge ergibt.
- Eine Einteilung ist **konsistent**, wenn sich die Teilbereiche nicht überlappen.

Im Beispiel der Körpergrößen ist eine Einteilung des Bereiches  $[1cm, 300cm]$  vollständig, wenn jede scharfe Körpergröße mindestens einem Teilbereich zugeordnet werden kann. Die Einteilung ist konsistent, wenn jede scharfe Körpergröße höchstens einem Teilbereich angehört. Eine vollständige und konsistente Einteilung nennen wir **Partitionierung**, das Resultat eine **Partition**.

Um zu einer unscharfen Partition zu gelangen, verallgemeinern wir den Begriff einer scharfen Partition im Sinne des Kontextmodells aus Abschnitt 2.3.1. Das Ergebnis ist eine Einteilung des Universums in unscharfe statt in scharfe Mengen. Im Kontextmodell wird eine unscharfe Menge als Abstraktion über einer Menge von Kontexten definiert. Jeder Kontext bestimmt eine scharfe Menge, die mit der Wahrscheinlichkeit des Kontextes gewichtet wird. Die gewichteten Mengen werden schließlich zusammengefasst, wodurch die unscharfe Menge als Plausibilitätsdichte determiniert wird. Wir verallgemeinern scharfe Partitionen auf die gleiche Weise zu einer unscharfen Partition, wie wir die scharfen Mengen

zu einer unscharfen Menge kombiniert haben. Jeder Kontext liefert eine scharfe Partition des Universums. Wenn wir  $k$  Teilbereiche annehmen, erhalten wir auf diese Weise  $k$  Plausibilitätsdichten. Wir bitten z. B.  $n$  Personen, den Bereich  $X = [1cm, 300cm]$  in Bezug auf die Körpergröße von Männern in fünf scharfe, disjunkte Mengen aufzuteilen, also zu partitionieren. Die Intervalle sollen die Begriffe  $\tilde{B} \in \mathcal{P} = \{\text{sehr klein, klein, normal, groß, sehr groß}\}$  modellieren. Wir erhalten von der Person  $c$  für jeden Begriff  $\tilde{B}$  eine Menge  $\Gamma_{\tilde{B}}(c)$ . Für jeden Begriff bilden wir über den Kontextmaßraum die Plausibilitätsdichte. Abbildung 4.1 zeigt das Ergebnis einer Befragung von 22 Personen.

Wir verlangen bei dieser Vorgehensweise, dass jeder Kontext eine scharfe Partition der Grundmenge generieren muss, die also vollständig und konsistent ist. Diese Forderung scheint sehr stark, was wir uns anhand des Beispiels der Körpergrößen veranschaulichen können: Eine Person hat u. U. Schwierigkeiten, erstens die Grenze zwischen zwei Bereichen genau anzugeben und zweitens bestimmte Körpergrößen eindeutig einem Begriff zuzuordnen. Diese Unschärfe modellieren wir aber mit unserem Ansatz ebenfalls, wenn auch indirekt: Wir erzwingen zwar in jedem Kontext eine scharfe Aussage, dies geschieht aber vor dem Hintergrund, dass die Unsicherheit jedes Kontextes durch die Variation in den Aussagen aller Kontexte ausgedrückt wird. Damit geben wir

**Definition 4** Eine endliche Menge  $\mathcal{P}$  unscharfer Begriffe über einem Kontextmaßraum  $(C, 2^C, P_C)$  heißt **unscharfe Partition** des Universums  $X$ , wenn die vagen Charakteristika  $\Gamma_{\tilde{B}}$  der unscharfen Begriffe  $\tilde{B} \in \mathcal{P}$  folgende Bedingungen erfüllen: Für alle  $c \in C$  gilt

$$\text{Überdeckung: } \bigcup_{\tilde{B} \in \mathcal{P}} \Gamma_{\tilde{B}}(c) = X \quad (4.1)$$

$$\text{paarweise Disjunktheit: } \forall \tilde{A} \neq \tilde{B} \in \mathcal{P} : \Gamma_{\tilde{A}}(c) \cap \Gamma_{\tilde{B}}(c) = \emptyset \quad (4.2)$$

Ein Element einer unscharfen Partition heißt **unscharfer Grundbegriff**.

Jeder Kontext  $c$  bringt so eine scharfe Partition des Universums  $X$  hervor. Alle Bereiche  $\Gamma_{\tilde{B}}(c)$  dieser scharfen Partition werden dadurch mit der gleichen Wahrscheinlichkeit  $P_C(c)$  belegt. Die Definition könnte dahingehend erweitert werden, dass jeder Grundbegriff durch eine eigene Kontextmenge gebildet wird. Beispielsweise könnten für jeden Grundbegriff andere Personengruppen befragt werden; die Begriffe werden dann unabhängig voneinander definiert. Da die Elemente einer Partitionierung aber wegen der Vollständigkeit und Konsistenz inhärent abhängig voneinander sind, erscheint eine Erweiterung auf mehrere Kontextmengen nicht sinnvoll.

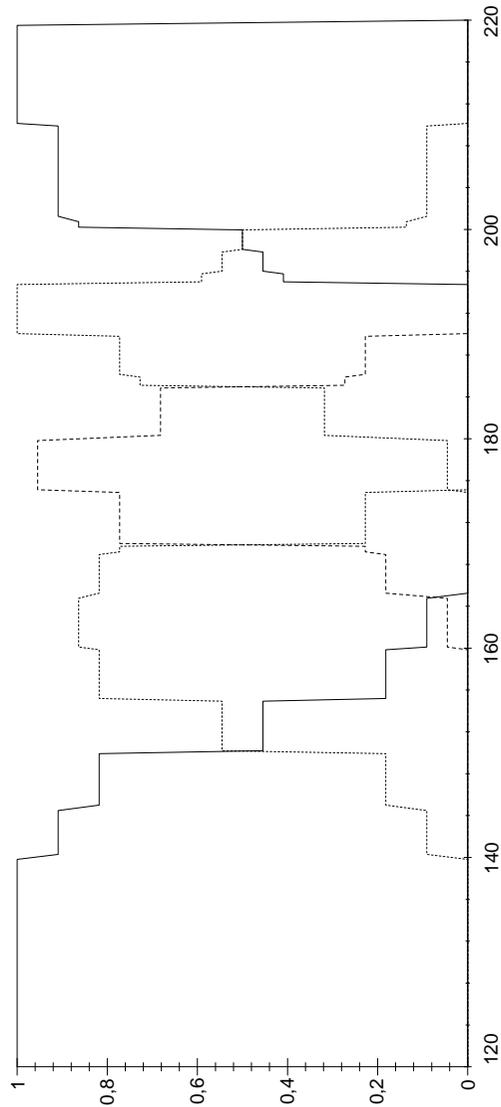


Abbildung 4.1: Unschärfe Körpergrößen erwachsener Männer: die Plausibilitätsdichten der Grundbegriffe *sehr klein*, *klein*, *normal*, *groß* und *sehr groß* (von links nach rechts)

Die Frage nach Konsistenz und Vollständigkeit der unscharfen Partition wird mit folgender *Überdeckungseigenschaft* beantwortet.

**Satz 3 (Überdeckungseigenschaft)** *Für die Plausibilitätsdichten der Begriffe einer unscharfen Partition  $\mathcal{P}$  gilt:*

$$\forall u \in \mathcal{X} : \sum_{\bar{B} \in \mathcal{P}} pl_{\bar{B}}(u) = 1 \quad (4.3)$$

*Insbesondere sind Grundbegriffe disjunkt.*

*Beweis:*

$$\begin{aligned} \sum_{\bar{B} \in \mathcal{P}} pl_{\bar{B}}(u) &= \sum_{\bar{B} \in \mathcal{P}} \sum_{c \in \mathcal{C}} \mathbb{1}_{u \in \Gamma_{\bar{B}}(c)} \cdot P_{\mathcal{C}}(c) \\ &= \sum_{c \in \mathcal{C}} P_{\mathcal{C}}(c) \sum_{\bar{B} \in \mathcal{P}} \mathbb{1}_{u \in \Gamma_{\bar{B}}(c)} \\ &= \sum_{c \in \mathcal{C}} P_{\mathcal{C}}(c) \\ &= 1 \end{aligned}$$

mit der Indikatorfunktion

$$\mathbb{1}_{\text{Bedingung}} = \begin{cases} 1, & \text{falls Bedingung wahr} \\ 0, & \text{sonst} \end{cases} \quad (4.4)$$

Die Disjunktheit der Grundbegriffe folgt unmittelbar aus den Definitionen 4 und 3 von Grundbegriff und Disjunktheit.  $\square$

Gleichung (4.3) beschreibt gleichzeitig die Vollständigkeit und die Konsistenz einer unscharfen Partition. Einerseits verallgemeinert die Gleichung scharfe Vollständigkeit, denn die Summe der Plausibilitäten muss im scharfen wie im unscharfen Fall größer oder gleich eins sein. Die Verallgemeinerung liegt darin, dass im Fall scharfer Mengen  $pl_{\bar{B}}(u) \in \{0, 1\}$ , hier aber  $pl_{\bar{B}}(u) \in [0, 1]$  gilt. Analog fordern wir andererseits für die Disjunktheit, dass die Summe der Plausibilitäten kleiner oder gleich eins ist. Zusammen ergibt sich damit die Überdeckungseigenschaft (4.3).

In der Literatur findet man die Überdeckungseigenschaft auch unter dem Stichwort *partition of unity*. Sie wird oft als Faustregel für den Entwurf unscharfer Granularisierungen benutzt [DHR93], i. A. aber ohne präzise Rechtfertigung. In unserem Fall wie auch bei der Interpretation von unscharfen Mengen als Likelihood-Funktionen [DMP93] ergibt sich die Überdeckungseigenschaft konstruktionsbedingt aus der Theorie. Bei Likelihood-Funktionen steht allerdings nicht der Gedanke im Vordergrund, ein Universum vollständig und konsistent aufzuteilen. Diese Eigenschaften sind nur implizit in den Axiomen der Wahrscheinlichkeitstheorie

enthalten. Insbesondere spielt aber die Disjunktheit der unscharfen Grundbegriffe, die aus der Konsistenz folgt, eine wesentliche Rolle bei der Kombination von Grundbegriffen im folgenden Abschnitt.

## 4.2 Kombination unscharfer Grundbegriffe

Wie in der alltäglichen Sprache werden wir unscharfe Grundbegriffe dazu benutzen, unscharfe Beobachtungen zu beschreiben oder Aussagen zu treffen. Wir drücken eine gegebene Situation durch den Wortschatz aus, der in der Anwendung definiert ist. Dafür benutzen wir gewichtete Kombinationen von Grundbegriffen; weiter oben haben wir mit der Aussage „Die Haarfarbe ist weniger *hellblond*, eher *mittelblond*.“ schon ein Beispiel gesehen. Während hier Begriffe mit linguistischen Gewichten wie *weniger* und *eher* versehen werden, greifen wir in der mathematischen Formalisierung auf numerische Gewichte zurück: Wir modellieren die Gewichte als Wahrscheinlichkeiten.

Unter der gewichteten *Kombination unscharfer Begriffe* verstehen wir die Vereinigung gewichteter unscharfer Begriffe. Als formale Aussage bedeutet dies „ $x = \tilde{A}$  mit Wahrscheinlichkeit  $P_{\mathcal{P}}(\tilde{A})$ , oder  $x = \tilde{B}$  mit Wahrscheinlichkeit  $P_{\mathcal{P}}(\tilde{B})$ , oder ...“ Man beachte an dieser Stelle, dass man umgangssprachlich die Disjunktion „oder“ gern durch „und“ ersetzt: „Der Täter war zu 70% *groß* und zu 30% *sehr groß*.“ Lediglich bei gleicher Gewichtung wird „oder“ benutzt: „Der Täter war *groß* oder *sehr groß*.“ Formal handelt es sich aber um eine Disjunktion.

Kombinierte unscharfe Aussagen sind auf grobgranularer Ebene definiert. Wie ein einzelner unscharfer Begriff wird auch eine Kombination unscharfer Begriffe durch eine unscharfe Menge auf den Einzelheiten erklärt. Die Übersetzung der Kombination von der grobgranularen Ebene der Begriffe auf die feingranulare Ebene der unscharfen Mengen nennen wir *Synthese* einer unscharfen Menge. Den Rückweg, also die Interpretation einer unscharfen Menge durch eine Kombination von Begriffen auf grobgranularer Ebene bezeichnen wir als *Analyse* der unscharfen Menge. Abbildung 4.2 illustriert noch einmal die beiden Aspekte einer Kombination: Die verständliche unscharfe Aussage und die unterliegende unscharfe Menge sind durch Synthese und Analyse miteinander verknüpft.

### 4.2.1 Begriffswahrscheinlichkeiten

Für alle weiteren Betrachtungen gehen wir davon aus, dass eine unscharfe Partition  $\mathcal{P}$  des Universums  $\mathcal{X}$  durch unscharfe Grundbegriffe gegeben ist. Auf  $\mathcal{P}$  definieren wir einen **Begriffsmaßraum**  $(\mathcal{P}, 2^{\mathcal{P}}, P_{\mathcal{P}})$ , mit dem wir die **Begriffswahrscheinlichkeiten**  $P_{\mathcal{P}}(\{\tilde{B}\})$  für  $\tilde{B} \in \mathcal{P}$  messen. Der Lesbarkeit halber schreiben wir  $P_{\mathcal{P}}(\tilde{B})$  statt  $P_{\mathcal{P}}(\{\tilde{B}\})$ .

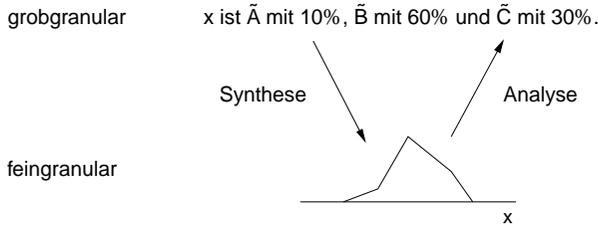


Abbildung 4.2: Synthese und Analyse einer unscharfen Menge

Es ist nicht trivial, dass die Konstruktion des Begriffsmaßraums vernünftig ist. Das Problem liegt darin, dass wir mit  $P_{\mathcal{P}}(\tilde{B})$  nicht nur die Wahrscheinlichkeit eines unscharfen Begriffs als eines Elements von  $\mathcal{P}$  messen, sondern auch die Wahrscheinlichkeit der unscharfen Menge  $\tilde{B} \in \mathcal{UM}(X)$ , die den Begriff definiert. Die Wahrscheinlichkeit der unscharfen Menge könnten wir bei Kenntnis der Wahrscheinlichkeitsdichte  $f$  der Variablen  $x$  direkt bestimmen<sup>1</sup>:

$$P(\tilde{B}) = \int_x P(\tilde{B}|u) \cdot f(u) du \quad (4.5)$$

wobei wir  $\tilde{B}$  durch die Likelihood-Funktion  $P(\tilde{B}|u)$  dargestellt haben. Mit unserem Begriffsmaßraum reduzieren wir also ein Wahrscheinlichkeitsmaß über  $\mathcal{UM}(X)$  auf ein Maß über  $\mathcal{P}$ . Unsere Definition des Begriffsmaßraums ist damit nur dann sinnvoll, wenn die unscharfen Mengen der Grundbegriffe im Sinne der Wahrscheinlichkeitsrechnung disjunkt sind, also  $P(\tilde{A} \cap \tilde{B}) = 0$  für  $\tilde{A} \neq \tilde{B} \in \mathcal{P}$ . Dann können wir die Grundbegriffe im Sinne der Wahrscheinlichkeitsrechnung als Elementarereignisse deuten und durch die einfachen Elemente der Menge  $\mathcal{P}$  substituieren. Folgender Satz beweist die gewünschte Eigenschaft.

**Satz 4** Seien  $\tilde{A}, \tilde{B} \in \mathcal{UM}(X)$  disjunkte, unscharfe Mengen,  $f$  die Wahrscheinlichkeitsdichte der Variablen  $x$  und  $P$  das durch Gleichung (4.5) induzierte Wahrscheinlichkeitsmaß über  $(\mathcal{UM}(X), 2^{\mathcal{UM}(X)})$ . Dann gelten

$$P(\tilde{A} \cup \tilde{B}) = P(\tilde{A}) + P(\tilde{B}) \quad (4.6)$$

$$P(\tilde{A} \cap \tilde{B}) = 0 \quad (4.7)$$

*Beweis:* Wir stellen die unscharfen Begriffe  $\tilde{A}, \tilde{B}$  durch die Likelihood-Funktionen

<sup>1</sup>gemäß der Formel von der totalen Wahrscheinlichkeit

$P(\tilde{A}|x)$ ,  $P(\tilde{B}|x)$  dar. Mit Gleichung (4.5) und Korollar 2 gilt

$$\begin{aligned} P(\tilde{A} \cup \tilde{B}) &= \int_{\mathcal{X}} P(\tilde{A} \cup \tilde{B}|u) \cdot f(u) \, du \\ &= \int_{\mathcal{X}} (P(\tilde{A}|u) + P(\tilde{B}|u)) \cdot f(u) \, du \\ &= \int_{\mathcal{X}} P(\tilde{A}|u) \cdot f(u) \, du + \int_{\mathcal{X}} P(\tilde{B}|u) \cdot f(u) \, du \\ &= P(\tilde{A}) + P(\tilde{B}) \end{aligned}$$

Die zweite Behauptung folgt unmittelbar aus Definition 3, Korollar 2 und der Definition eines Wahrscheinlichkeitsmaßes:  $P(\tilde{A} \cap \tilde{B}) = P(\emptyset) = 0$ .  $\square$

Da die Elemente einer Partition  $\mathcal{P}$  Grundbegriffe und damit nach Satz 3 disjunkt sind, können wir den Satz anwenden. Statt die Begriffswahrscheinlichkeiten auf der feingranularen Ebene von  $\mathcal{UM}(\mathcal{X})$  zu definieren, dürfen wir deshalb auf der grobgranularen Partition  $\mathcal{P}$  arbeiten. Wir werden daher bei allen weiteren Betrachtungen den Begriffsmaßraum  $(\mathcal{P}, 2^{\mathcal{P}}, P_{\mathcal{P}})$  für die Gewichtung von Grundbegriffen verwenden.

Trotz dieser formalen Vereinfachung dürfen wir nicht vergessen, dass wir mit der Wahrscheinlichkeit eines unscharfen Begriffes nicht nur das Symbol aus  $\mathcal{P}$  gewichten, sondern auch die unscharfe Menge, die den Begriff definiert. Dies ist insbesondere dann wichtig, wenn wir die Wahrscheinlichkeit eines Begriffes bestimmen möchten, denn ein Begriff hat in unterschiedlichen Kontexten verschiedene Bedeutungen. In unserem Fall ist die Bedeutung immer durch die unterliegende unscharfe Menge gegeben.

### 4.2.2 Synthese

Eine unscharfe Aussage „ $x = \tilde{A}$ “ können wir auf zwei Arten interpretieren. Einerseits ordnen wir der Variablen  $x$  den unscharfen Wert  $\tilde{A}$  in Form einer unscharfen Menge zu. Dabei haben wir den unscharfen Wert normalerweise mit einem unscharfen Begriff belegt, der dessen Bedeutung in der Sprache des Anwenders festlegt. Wir bewegen uns hier auf der grobgranularen Ebene. Auf der anderen Seite wird  $\tilde{A}$  durch eine unscharfe Menge auf dem Universum  $\mathcal{X}$  der Einzelheiten definiert, in unserem Fall durch eine Plausibilitätsdichte  $pl_{\tilde{A}}$ . Mit dem Wissen, dass  $x$  den unscharfen Wert  $\tilde{A}$  besitzt, können wir mit  $pl_{\tilde{A}}(u)$  die Plausibilität messen, dass  $x$  den scharfen Wert  $u$  annimmt. Eine gewichtete Kombination unscharfer Begriffe repräsentiert ebenfalls einen unscharfen Wert. In diesem Abschnitt werden wir herleiten, wie die Plausibilitätsdichte berechnet wird, die diesen unscharfen Wert darstellt. Als Beispiel betrachten wir wiederum die Aussage „Der Täter war mit

70% Wahrscheinlichkeit *groß* und mit 30% Wahrscheinlichkeit *sehr groß*.“ Aus dieser grobgranularen Aussage wollen wir schließen, wie groß die Plausibilität der feingranularen Aussage „Der Täter war *u cm* groß.“ für beliebige  $u \in \mathcal{X}$  ist.

Formal gehen wir von einem Begriffsmaßraum  $(\mathcal{P}, 2^{\mathcal{P}}, P_{\mathcal{P}})$  aus. Jeder Grundbegriff  $\tilde{B} \in \mathcal{P}$  ist durch den Kontextmaßraum  $(\mathcal{C}, 2^{\mathcal{C}}, P_{\mathcal{C}})$  und die vagen Charakteristika  $\Gamma_{\tilde{B}}(c)$  für  $c \in \mathcal{C}$  definiert. Daraus berechnen wir die Plausibilitätsdichten der Grundbegriffe. Wir suchen nun den Kontextmaßraum und die vagen Charakteristika, welche die gewichtete Kombination der Grundbegriffe darstellen. Dazu rufen wir uns wieder ins Gedächtnis, wie die Plausibilitätsdichte eines Begriffs  $\tilde{B}$  berechnet wird. Jeder Kontext  $c$  legt eine scharfe Menge, das vage Charakteristikum  $\Gamma_{\tilde{B}}(c)$ , als Definition des Begriffs fest. Das Vertrauen in diese Definition wird durch die Kontextwahrscheinlichkeit  $P_{\mathcal{C}}(c)$  gemessen. Die vagen Charakteristika werden schließlich mit  $P_{\mathcal{C}}(c)$  gewichtet und zusammengefasst. Liegt eine gewichtete Kombination unscharfer Begriffe vor, so sind die vagen Charakteristika aller beteiligten Begriffe mögliche Definitionen des unscharfen Wertes, den die Kombination darstellt. Die Menge der vagen Charakteristika der Kombination ist damit die Vereinigung der Charakteristika aller benutzten Grundbegriffe.

Als nächstes müssen wir das Vertrauen in die vagen Charakteristika bestimmen. Jeder Grundbegriff  $\tilde{B}$  wird mit der Wahrscheinlichkeit  $P_{\mathcal{P}}(\tilde{B})$  gewichtet. Damit verändern wir auch das Vertrauen in die Kontexte, die  $\tilde{B}$  definieren: Wir gewichten  $P_{\mathcal{C}}(c)$  multiplikativ mit  $P_{\mathcal{P}}(\tilde{B})$ . Dies ist das übliche Vorgehen in der Wahrscheinlichkeitsrechnung, wenn ein zweistufiges Experiment durchgeführt wird, bei dem die beiden Stufen unabhängig voneinander sind. In der ersten Stufe wird die Wahrscheinlichkeit des Begriffs festgelegt und in der zweiten Stufe die Wahrscheinlichkeiten der definierenden Kontexte. Das Gesamtvertrauen in den Kontext ist dann das Produkt von Begriffs- und Kontextwahrscheinlichkeit.

Die Gewichtung der Kontextwahrscheinlichkeiten durch die Begriffswahrscheinlichkeiten veranschaulichen wir an folgendem Beispiel. Wir lassen über die Größe einer Person abstimmen. Jeder Befragte muss sich für eine unscharfe Größe aus  $\{\textit{sehr klein}, \textit{klein}, \textit{normal}, \textit{groß}, \textit{sehr groß}\}$  entscheiden. Nehmen wir einmal  $P_{\mathcal{P}}(\textit{normal}) = P_{\mathcal{P}}(\textit{groß}) = 1/2$  als Ergebnis an. Das können wir so interpretieren, als wenn zwei andere Umfragen gemacht wurden: Zwei Personengruppen (Kontextmengen) sollen die Größe der Person mit scharfen Mengen (vage Charakteristika) definieren. Wir nehmen an, dass eine Gruppe die vagen Charakteristika von *normal*, die andere Gruppe die von *groß* liefert. Diese zwei Umfragen können wir verschmelzen, indem wir vorgeben, es wäre nur eine Umfrage gewesen. Dazu vereinigen wir die Mengen der befragten Personen (die Kontexte) und normieren die Kontextwahrscheinlichkeiten  $P_{\mathcal{C}}(c)$  entsprechend der neuen Anzahl von Kontexten. Wenn die Begriffe mit unterschiedlichen Wahrscheinlichkeiten belegt werden

( $P_{\mathcal{P}}(\text{normal}) \neq P_{\mathcal{P}}(\text{groß})$ ), dann gewichten wir die Umfragen und damit die Kontextwahrscheinlichkeiten entsprechend.

Um die gewichtete Kombination von Grundbegriffen in das Kontextmodell einzubetten, müssen wir nach den vorangehenden Überlegungen einen neuen Kontextmaßraum konstruieren. Bei einer unscharfen Partitionierung beruhen alle Grundbegriffe auf derselben Kontextmenge  $\mathcal{C}$ . Wie geschildert, werden bei der gewichteten Kombination die Kontextwahrscheinlichkeiten mit den Begriffswahrscheinlichkeiten modifiziert. Da jeder Grundbegriff seine eigene Wahrscheinlichkeit hat, ist diese Modifikation für jeden Grundbegriff verschieden. Wir müssen daher die Kontextmenge  $\mathcal{C}$  für jeden Grundbegriff duplizieren, womit wir die Kontextmenge  $\mathcal{C}'$  der Kombination bilden. Insgesamt gilt dann für den Kontextmaßraum  $(\mathcal{C}', 2^{\mathcal{C}'}, P_{\mathcal{C}'})$  der Kombination:

- Dupliziere die Menge  $\mathcal{C}$  von Kontexten für jeden Grundbegriff  $\tilde{B} \in \mathcal{P}$  durch Umbenennung in  $\mathcal{C}(\tilde{B})$ :

$$\mathcal{C}' := \bigcup_{\tilde{B} \in \mathcal{P}} \mathcal{C}(\tilde{B}) \quad (4.8)$$

- Bilde die neue Kontextwahrscheinlichkeit  $P_{\mathcal{C}'}$  durch Gewichtung der alten Kontextwahrscheinlichkeit  $P_{\mathcal{C}}$ .  $c(\tilde{B})$  bezeichne das Duplikat des Originalkontextes  $c \in \mathcal{C}$  für  $\tilde{B}$ .

$$P_{\mathcal{C}'}(c(\tilde{B})) := P_{\mathcal{P}}(\tilde{B}) \cdot P_{\mathcal{C}}(c) \quad (4.9)$$

$$\text{und } \forall A \subseteq \mathcal{C}' : P_{\mathcal{C}'}(A) := \sum_{c \in A} P_{\mathcal{C}'}(c) \quad (4.10)$$

- Definiere die vagen Charakteristika  $\Gamma^l(c)$  als triviale Fortsetzung der partiellen Abbildungen

$$\Gamma^l|_{\mathcal{C}(\tilde{B})}(c(\tilde{B})) := \Gamma_{\tilde{B}}(c) \quad (4.11)$$

auf  $\mathcal{C}'$ .

Wir müssen noch beweisen, dass das so definierte  $P_{\mathcal{C}'}$  tatsächlich ein Wahrscheinlichkeitsmaß ist.

**Satz 5** Die durch die Gleichungen (4.9) und (4.10) definierte Abbildung  $P_{\mathcal{C}'}$  ist ein Wahrscheinlichkeitsmaß im Messraum  $(\mathcal{C}', 2^{\mathcal{C}'})$ .

*Beweis:* Wir zeigen die Gültigkeit der drei Axiome von Kolmogoroff.

1. Aus (4.9) folgt unmittelbar  $0 \leq P_{C'}(A) \leq 1$  für alle  $A \in 2^{C'}$ .
2.  $P_{C'}(\bigcup_i A_i) = \sum_i P_{C'}(A_i)$  für disjunkte Mengen  $A_i \in 2^{C'}$  folgt direkt aus Gleichung (4.10).
3. Die Wahrscheinlichkeit der Grundmenge  $C'$  muss eins sein:

$$\begin{aligned}
 P_{C'}(C') &= \sum_{\tilde{B} \in \mathcal{P}} \sum_{c \in C(\tilde{B})} P_{C'}(c(\tilde{B})) \\
 &= \sum_{\tilde{B} \in \mathcal{P}} P_{\mathcal{P}}(\tilde{B}) \sum_{c \in C} P_C(c) \\
 &= \sum_{\tilde{B} \in \mathcal{P}} P_{\mathcal{P}}(\tilde{B}) \\
 &= 1
 \end{aligned}$$

□

Damit können wir formal die gewichtete Kombination von Grundbegriffen und die Synthese einer unscharfen Menge definieren:

**Definition 5** Sei  $\mathcal{P}$  eine unscharfe Partition des Universums  $X$  mit Grundbegriffen im Kontextmaßraum  $(C, 2^C, P_C)$ , die durch das Wahrscheinlichkeitsmaß  $P_{\mathcal{P}}$  im Begriffsmaßraum  $(\mathcal{P}, 2^{\mathcal{P}}, P_{\mathcal{P}})$  gewichtet werden. Die **gewichtete Kombination der Grundbegriffe** wird durch eine unscharfe Menge auf  $X$  erklärt. Sie wird mit der Plausibilitätsdichte identifiziert, die wir mithilfe der Gleichungen (4.8)–(4.11) auf dem Kontextmaßraum  $(C', 2^{C'}, P_{C'})$  berechnen. Diese Berechnung nennen wir **Synthese der unscharfen Menge**.

Für die Plausibilitätsdichte einer Kombination von Grundbegriffen gilt

**Satz 6 (Kombination von Plausibilitätsdichten)** Werden die Grundbegriffe einer Partition  $\mathcal{P}$  über einen Begriffsmaßraum  $(\mathcal{P}, 2^{\mathcal{P}}, P_{\mathcal{P}})$  kombiniert, so gilt für die Plausibilitätsdichte  $pl_{\tilde{A}}$  der Kombination:

$$pl_{\tilde{A}}(u) = \sum_{\tilde{B} \in \mathcal{P}} P_{\mathcal{P}}(\tilde{B}) \cdot pl_{\tilde{B}}(u) \quad (4.12)$$

*Beweis:* Die Grundbegriffe  $\tilde{B} \in \mathcal{P}$  seien wiederum durch vage Charakteristika  $\Gamma_{\tilde{B}}$

mit dem Kontextmaßraum  $(C, 2^C, P_C)$  erklärt. Dann gilt

$$\begin{aligned}
 pl_{\tilde{A}}(u) &= P_{C'}(\{c \in C' \mid u \in \Gamma'(c)\}) \\
 &= P_{C'}\left(\bigcup_{\tilde{B} \in \mathcal{P}} \{c \in C(\tilde{B}) \mid u \in \Gamma_{\tilde{B}}(c)\}\right) \\
 &= \sum_{\tilde{B} \in \mathcal{P}} P_{C'}(\{c \in C(\tilde{B}) \mid u \in \Gamma_{\tilde{B}}(c)\}) \\
 &= \sum_{\tilde{B} \in \mathcal{P}} P_{\mathcal{P}}(\tilde{B}) \cdot P_C(\{c \in C \mid u \in \Gamma_{\tilde{B}}(c)\}) \\
 &= \sum_{\tilde{B} \in \mathcal{P}} P_{\mathcal{P}}(\tilde{B}) \cdot pl_{\tilde{B}}(u)
 \end{aligned}$$

□

Dieser Satz zeigt, dass wir die Plausibilitätsdichte einer Kombination direkt aus den Plausibilitätsdichten der Grundbegriffe bestimmen können. Wir müssen nicht den Umweg über die vagen Charakteristika gehen, wie Definition 5 suggeriert. Dies ist insbesondere für Anwendungen wichtig, bei denen die Grundbegriffe direkt über Plausibilitätsdichten und nicht über vage Charakteristika definiert werden, vgl. die Abschnitte 6.1 und 6.4.

Die Synthese haben wir benutzt, um die Größe von Herrn G. zu bestimmen. Die Gruppe von 22 Personen wurde zu diesem Zweck gebeten, über die unscharfen Mengen der Grundbegriffe  $\{\textit{sehr klein}, \textit{klein}, \textit{normal}, \textit{groß}, \textit{sehr groß}\}$  als mögliche unscharfe Größen abzustimmen. 41% der Befragten entschieden sich für *normal*, 59% für *groß*, was der grobgranularen Aussage „Herr G. ist mit Wahrscheinlichkeit 0,41 *normal* groß und mit Wahrscheinlichkeit 0,59 *groß*.“ entspricht. Nach Satz 6 gilt für die Plausibilitätsdichte der Kombination  $pl_{\tilde{B}}(u) = 0,41 \cdot pl_{\textit{normal}}(u) + 0,59 \cdot pl_{\textit{groß}}(u)$ . Abbildung 4.3 zeigt die Plausibilitätsdichten von *normal*, von *groß* und die der gewichteten Kombination. Die größte Plausibilität liegt mit 0,59 im Intervall  $[190\textit{cm}, 194\textit{cm}]$ .

Wie bereits angedeutet, ist das vorgestellte Verfahren der Kombination nur dann richtig, wenn die Definition der unscharfen Begriffe unabhängig von der Bestimmung der Begriffswahrscheinlichkeiten ist. Wenn wie im Beispiel dieselben Personen über die unscharfe Größe des Mannes abstimmen, die vorher die unscharfen Größen *normal* und *groß* definiert haben, dann müssen wir beachten, welche Definition der Grundbegriffe sie zugrunde legen. In Abschnitt 4.2.1 haben wir erklärt, dass die Grundbegriffe immer mit den unscharfen Mengen identifiziert werden, die sie definieren. Würden die Befragten ihre Entscheidungen für *normal* oder *groß* nicht auf der Basis der unscharfen Mengen sondern auf der Basis ihrer eigenen Definitionen von *normal* bzw. *groß* fällen, dann müssten wir die Kombination der

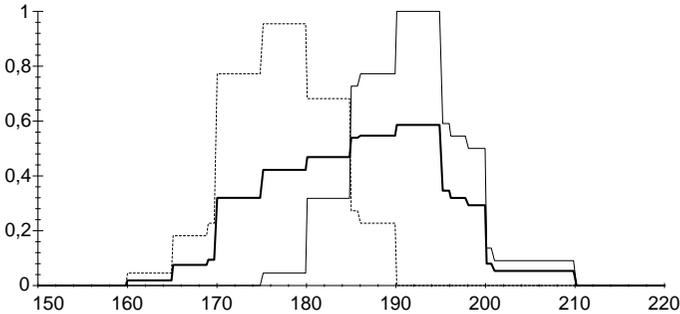


Abbildung 4.3: Die Körpergröße von Herrn G. als Kombination (fett) von *normaler* (gestrichelt) und *großer* Körpergröße (dünn)

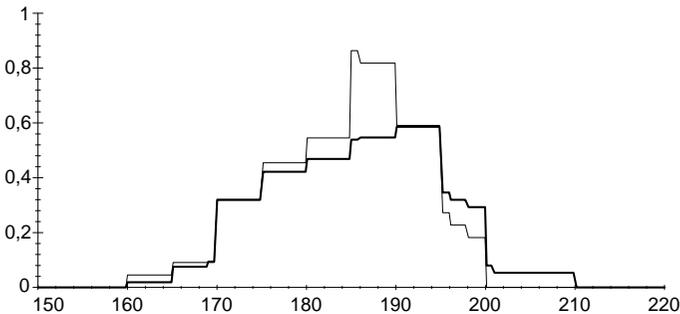


Abbildung 4.4: Die Körpergröße von Herrn G. als Kombination von *normaler* und *großer* Körpergröße; Abstimmung über *normal* oder *groß* unabhängig (fett) und abhängig (dünn) von der Definition von *normal*, *groß*

Begriffe anders definieren. Jeder Befragte würde die Größe der betrachteten Person seiner selbst definierten scharfen Menge zuordnen. Wir erhielten scharfe Mengen als vage Charakteristika für die Größe der Person und bildeten daraus wie in Abschnitt 2.3.1 die Plausibilitätsdichte. Das Ergebnis wäre offensichtlich ein anderes, als wenn wir aus derselben Abstimmung eine Plausibilitätsdichte nach Gleichung (4.12) bestimmen. Abbildung 4.4 zeigt die Körpergröße von Herrn G. als Kombination sowie unter Berücksichtigung der Abhängigkeit der Umfragen. Letzteres liefert i. A. genauere Aussagen. Die höchste Plausibilität liegt bei  $[185\text{cm}, 186\text{cm}]$  mit 0,86, gefolgt vom Intervall  $[186\text{cm}, 190\text{cm}]$  mit Plausibilität 0,82.

Für alle folgenden Betrachtungen setzen wir die Unabhängigkeit der Definitionen von Grundbegriffen und der Bestimmung der Begriffswahrscheinlichkeiten voraus. Wir gehen davon aus, dass die Partition immer als Erstes generiert wird und sich damit alle weiteren Aussagen auf die so definierten Grundbegriffe beziehen.

Dieses Vorgehen ist insbesondere dann unausweichlich, wenn die Partition aus Daten gewonnen wird, z. B. durch Gruppierungsverfahren wie in Abschnitt 6.1. Die Partition wird in diesem Fall als gegeben hingenommen, und die Bestimmung der Wahrscheinlichkeit eines Grundbegriffs ist damit unabhängig von seiner Definition.

Die Grundbegriffe bilden über ihre gewichtete Kombination eine Art Erzeugendensystem für unscharfe Aussagen. Satz 6 zeigt, dass die gewichtete Kombination eine Linearkombination der Plausibilitätsdichten ist. Wir kennen dieses Vorgehen aus der Mathematik von Verfahren zur Funktionsapproximation. Eine Menge von Basisfunktionen wird durch Linearkombination zu einer Zielfunktion zusammengesetzt. Bei der Kombination von Grundbegriffen ist die Zielfunktion die Plausibilitätsdichte der unscharfen Menge, welche die Kombination auf feingranularer Ebene definiert. Diesen Zusammenhang zwischen Funktionsapproximation und der Kombination von Grundbegriffen werden wir bei der Analyse unscharfer Mengen in Abschnitt 4.2.3 nutzen.

Wir erkennen nun auch den Raum möglicher unscharfer Werte oder Aussagen, den die gewichteten Kombinationen von Grundbegriffen aufspannen. Offenbar ist dies ein linearer Unterraum von  $\mathcal{UM}(\mathcal{X})$ . Wir können mit einer Kombination also nicht alle unscharfen Werte in  $\mathcal{UM}(\mathcal{X})$  exakt beschreiben. Dies ist der Preis für die Granularisierung von  $\mathcal{X}$ , die aber die Verständlichkeit, und wie wir später sehen werden, auch die Effizienz beim unscharfen Schließen sicherstellt.

Für den weiteren Verlauf der Arbeit nehmen wir an, dass die Plausibilitätsdichten der Grundbegriffe linear unabhängige Funktionen sind. Falls die Plausibilitätsdichten linear abhängig sein sollten, reduzieren wir sie auf einen Satz linear unabhängiger Dichten, indem wir einige Dichten streichen. Wir verlieren dadurch nichts, denn die gestrichenen Plausibilitätsdichten können durch eine Linearkombination der nicht gestrichenen dargestellt werden. Insbesondere spannt der reduzierte Satz von Plausibilitätsdichten denselben Unterraum von  $\mathcal{UM}(\mathcal{X})$  auf wie der vollständige Satz.

Der Grund für den Wunsch linearer Unabhängigkeit liegt darin, dass unter dieser Annahme jedes Element des linearen Unterraums eindeutig durch die Plausibilitätsdichten der Grundbegriffe beschrieben wird, denn die Dichten bilden dann eine Basis des Unterraums. Im Hinblick auf eine unscharfe Partition von  $\mathcal{X}$  bedeutet das, dass jede Information, die durch eine unscharfe Menge aus dem linearen Unterraum von  $\mathcal{UM}(\mathcal{X})$  dargestellt wird, eindeutig durch eine gewichtete Kombination der Grundbegriffe dargestellt werden kann. Diese Tatsache erweist sich im folgenden Abschnitt bei der Analyse unscharfer Mengen als sehr nützlich.

Zusammenfassend haben wir gesehen, wie wir mit den Grundbegriffen einer Partition komplexe unscharfe Aussagen auf grobgranularer Ebene bilden können. Trotz der Vergrößerung des Universums ist unsere Ausdruckskraft damit nicht auf einzelne Grundbegriffe beschränkt, sondern wir können die Grundbegriffe zu neuen Aussagen kombinieren. Diese sind zudem leicht verständlich, weil sie mit den Grundbegriffen ausschließlich auf unseren Wortschatz zurückgreifen. Durch die Synthese werden solche grobgranularen Aussagen in unscharfe Mengen übersetzt, welche die Aussagen auf feingranularer Ebene definieren. Damit liegen die Aussagen gleichzeitig in verständlicher Form auf grobgranularer Ebene und wohldefiniert auf der Ebene der Einzelheiten vor. Den Rückschritt von unscharfen Mengen zu grobgranularen Aussagen leiten wir im folgenden Abschnitt her.

### 4.2.3 Analyse

Im Fall der synthetischen Darstellung eines unscharfen Wertes durch eine Kombination von Grundbegriffen wird der Wert auf grobgranularer Ebene formuliert. In Anwendungen liegen aber oft Beobachtungen vor, die in Form einer Plausibilitätsdichte und damit auf feingranularer Ebene gegeben sind. Der Zusammenhang zwischen der Plausibilitätsdichte und den Grundbegriffen ist dann unklar. Mittels einer *Analyse* soll dieser Zusammenhang hergestellt werden, indem die unscharfe Beobachtung durch eine Kombination der Grundbegriffe approximiert wird, vgl. Abb. 4.2. Als Ergebnis liegt dann eine leicht verständliche Beschreibung der gegebenen unscharfen Beobachtung durch die Grundbegriffe vor; eine feingranulare Information wird auf diese Weise vergrößert.

Kehren wir zu dem Beispiel der Körpergröße von Herrn G. zurück: 22 Personen geben ein (möglichst kleines) Intervall an, in dem sie die Körpergröße von Herrn G. vermuten. Wie in Abschnitt 2.3.1 bilden wir daraus die Plausibilitätsdichte  $pl_{\tilde{A}}$ , welche die unscharfe Körpergröße  $\tilde{A}$  des Mannes beschreibt. Das Ergebnis der Umfrage zeigt Abbildung 4.5 mit der größten Plausibilität von 0,68 in  $[187cm, 188cm]$ . Ziel der Analyse ist die Approximation der unscharfen Messung  $\tilde{A}$  der Körpergröße durch eine gewichtete Kombination der Grundbegriffe *sehr klein*, *klein*, *normal*, *groß* und *sehr groß*.

Die unbekanntenen Größen bei einer Analyse sind die Wahrscheinlichkeiten der Grundbegriffe  $\tilde{B} \in \mathcal{P}$ . Wir kennzeichnen sie bei der Analyse durch  $P_{\mathcal{P}}(\tilde{B}|\tilde{A})$  statt  $P_{\mathcal{P}}(\tilde{B})$  um zu verdeutlichen, dass wir nicht direkt die Wahrscheinlichkeit von  $\tilde{B}$  angeben, sondern die bedingte Wahrscheinlichkeit von  $\tilde{B}$  unter der Beobachtung  $\tilde{A}$ .

Wir erwarten von der Analyse, also der Bestimmung der Begriffswahrscheinlichkeiten, dass sie konsistent mit der Synthese aus Abschnitt 4.2.2 ist. Konsistenz

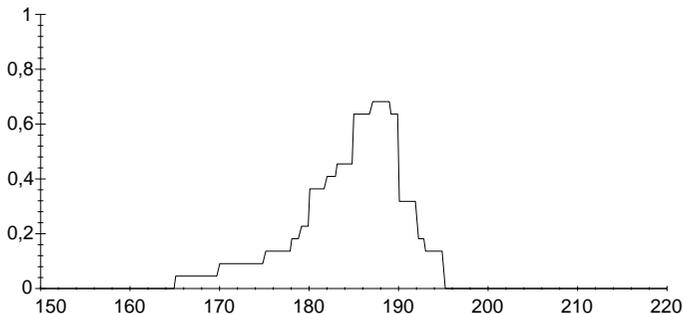


Abbildung 4.5: Plausibilitätsdichte der Größe von Herrn G. aus einer Umfrage

bedeutet dabei Folgendes: Wenn wir einen unscharfen Wert durch eine gewichtete Kombination der Grundbegriffe synthetisieren und die resultierende Plausibilitätsdichte wieder analysieren, so soll die Analyse die Begriffswahrscheinlichkeiten der Synthese rekonstruieren. Diese Forderung ist eine zwingende Voraussetzung, um das Verständnis unscharfer Aussagen sicherzustellen. Denn wenn wir wissen, dass eine Aussage durch Grundbegriffe exakt dargestellt werden kann, dann soll die Analyse diese Darstellung auch finden. Ansonsten sind Synthese und Analyse semantisch nicht verträglich und die Bedeutung unscharfer Aussagen ist unklar. Wir definieren daher formal die Analyse eines unscharfen Wertes mit

**Definition 6** Sei  $\mathcal{P}$  eine unscharfe Partition des Universums  $X$  sowie  $\tilde{A}$  eine unscharfe Menge auf  $X$ . Die **Analyse** von  $\tilde{A}$  berechnet die Begriffswahrscheinlichkeiten  $P_{\mathcal{P}}(\tilde{B}|\tilde{A})$  so, dass die gewichtete Kombination der Grundbegriffe  $\tilde{B} \in \mathcal{P}$  die unscharfe Menge  $\tilde{A}$  approximiert. Unter der Annahme, dass  $\tilde{A}$  bereits eine Kombination der Grundbegriffe  $\tilde{B} \in \mathcal{P}$  mit Gewichten  $P'_{\mathcal{P}}(\tilde{B})$  ist, soll gelten:

$$\forall \tilde{B} \in \mathcal{P} : P_{\mathcal{P}}(\tilde{B}|\tilde{A}) = P'_{\mathcal{P}}(\tilde{B}) \quad (4.13)$$

Bei der Rekonstruktion der Begriffswahrscheinlichkeiten in (4.13) nutzen wir die Annahme von S. 47 aus, dass die Plausibilitätsdichten der Grundbegriffe linear unabhängig sind. Andernfalls wären die Begriffswahrscheinlichkeiten, die  $\tilde{A}$  definieren, nicht eindeutig bestimmt.

Grundsätzlich sind zwei Ansätze für die Durchführung der Analyse denkbar. Der erste Ansatz arbeitet auf der Ebene der Maßbasen. Hierbei werden die vagen Charakteristika  $\Gamma_{\tilde{A}}$  der Beobachtung mit den vagen Charakteristika  $\Gamma_{\tilde{B}}$  der Grundbegriffe verglichen. Auf dieser Basis werden die Kontextwahrscheinlichkeiten der Grundbegriffe neu bestimmt. Im zweiten Ansatz geben wir uns eine Ebene höher und operieren nur mit den Plausibilitätsdichten, ähnlich der Vorgehensweise bei

der Synthese von Aussagen. Für den ersten Ansatz variieren wir bekannte Verfahren aus der Literatur. Er führt zwar in gewisser Hinsicht zu präziseren Ergebnissen als der zweite Ansatz, verletzt aber die Forderung (4.13), wie sich zeigen wird. Im zweiten Ansatz bilden wir den Analysevorgang auf ein Problem der Funktionsapproximation ab, wobei Forderung (4.13) erfüllt wird.

**Ansatz 1** Wir benutzen in diesem Ansatz die vagen Charakteristika der Grundbegriffe, um die Beobachtung darzustellen. Dazu müssen wir die Gewichte der Charakteristika neu bestimmen, d. h. die Wahrscheinlichkeiten der Kontexte anpassen. Die Beobachtung  $\tilde{A}$  basiere auf einem Kontextmaßraum  $(\mathcal{D}, 2^{\mathcal{D}}, P_{\mathcal{D}})$ . Betrachten wir exemplarisch das vage Charakteristikum  $\Gamma_{\tilde{B}}(c)$  eines Grundbegriffs. Die Frage ist, wie stark ein vages Charakteristikum  $\Gamma_{\tilde{B}}(c)$  zur Darstellung der Beobachtung beitragen kann. Offenbar kann es nur für die vagen Charakteristika  $\Gamma_{\tilde{A}}(d)$  der Beobachtung als Vertreter dienen, mit denen es eine gemeinsame Teilmenge besitzt:  $\Gamma_{\tilde{B}}(c) \cap \Gamma_{\tilde{A}}(d) \neq \emptyset$ . Die Relevanz von  $\Gamma_{\tilde{B}}(c)$  für die Beobachtung wird dann gerade durch seine Plausibilität

$$Pl_{\tilde{A}}(\Gamma_{\tilde{B}}(c)) = \sum_{\substack{d \in \mathcal{D}: \\ \Gamma_{\tilde{B}}(c) \cap \Gamma_{\tilde{A}}(d) \neq \emptyset}} P_{\mathcal{D}}(d) \quad (4.14)$$

gemessen, wobei  $Pl_{\tilde{A}}$  das Plausibilitätsmaß der Beobachtung ist. Jedes vage Charakteristikum  $\Gamma_{\tilde{A}}(d)$ , das eine nichtleere Schnittmenge mit  $\Gamma_{\tilde{B}}(c)$  besitzt, trägt so mit seiner Wahrscheinlichkeit  $P_{\mathcal{D}}(d)$  zur Relevanz von  $\Gamma_{\tilde{B}}(c)$  bei.

Für die Durchführung des Verfahrens definieren wir wie bei der Synthese einen neuen Kontextmaßraum. Die neue Kontextmenge  $\mathcal{C}'$  und die vagen Charakteristika  $\Gamma'$  werden durch die Vorschriften (4.8) und (4.11) festgelegt. Für die Kontextwahrscheinlichkeit gilt dann

$$P_{\mathcal{C}'}(c(\tilde{B})) := \omega(c(\tilde{B})) \cdot P_{\mathcal{C}}(c) \quad (4.15)$$

$$\text{mit } \omega(c(\tilde{B})) = \frac{Pl_{\tilde{A}}(\Gamma'(c(\tilde{B})))}{\sum_{c'(\tilde{B}) \in \mathcal{C}'} Pl_{\tilde{A}}(\Gamma'(c'(\tilde{B})))} \cdot P_{\mathcal{C}}(c') \quad (4.16)$$

für alle  $c(\tilde{B}) \in \mathcal{C}'$ , wobei  $c(\tilde{B})$  wieder das Duplikat des Kontextes  $c \in \mathcal{C}$  für den Grundbegriff  $\tilde{B}$  kennzeichnet. Der Nenner von  $\omega(c)$  sorgt als Normierungsfaktor dafür, dass  $P_{\mathcal{C}'}$  ein Wahrscheinlichkeitsmaß ist.

In der Literatur existieren Ansätze für ähnliche Probleme, die noch berücksichtigen, wie groß die Schnittmenge  $\Gamma_{\tilde{B}}(c) \cap \Gamma_{\tilde{A}}(d)$  in (4.14) ist [Sha76, BMP95]. Dabei liegt aber die Annahme zugrunde, dass die Elemente eines vagen Charakteristikums gleichverteilt sind. Diese Annahme trifft schon für unser Beispiel mit Körpergrößen nicht zu. Man würde hier eher eine Wahrscheinlichkeitsdichte annehmen, die ihr Maximum in der Mitte der angegebenen Mengen hat. Man kann

die Verfahren so erweitern, dass solche Kenntnisse über die Verteilung berücksichtigt werden können, was man dann auch ausnutzen sollte. Für den häufig vorkommenden Fall, dass die Verteilung nicht bekannt ist, bleiben wir aber bei der Plausibilität als Gewichtung.

Vorschrift (4.15) haben wir benutzt, um die unscharfe Beobachtung der Körpergröße von Herrn G. aus Abbildung 4.5 zu approximieren. Die dünne Kurve in Abbildung 4.6 zeigt die Plausibilitätsdichte der Approximation. Im Intervall  $[185\text{cm}, 186\text{cm})$  besteht die größte Plausibilität mit 0,66, in  $[186\text{cm}, 190\text{cm})$  eine unwesentlich geringere von 0,65.

Ein Vergleich der Gleichung (4.15) mit der entsprechenden Gleichung (4.9) bei der synthetischen Kombination von Begriffen zeigt, dass diese Form der Aggregation von Informationen i. A. keine Kombination von Grundbegriffen ist. Der Unterschied liegt darin, dass bei der Kombination alle Kontextwahrscheinlichkeiten eines Begriffs  $\tilde{B}$  mit demselben Faktor, nämlich der Begriffswahrscheinlichkeit  $P_p(\tilde{B})$ , gewichtet werden. Bei Ansatz 1 für die Analyse hingegen ist jedem Kontext  $c$  ein eigenes Gewicht  $\omega(c)$  zugeordnet. Dadurch wird die Relation der Kontextwahrscheinlichkeiten eines Grundbegriffs zerstört und das Ergebnis kann nicht mehr durch die Grundbegriffe ausgedrückt werden.

Prinzipiell wäre es denkbar, die Begriffswahrscheinlichkeiten  $P_p(\tilde{B}|\tilde{A})$  aus den  $\omega(c)$  zu berechnen. Die Entscheidung für einen Ansatz wie Minimum-, Maximum- oder Durchschnittsbildung ist aber schwer zu treffen und hängt stark von der Anwendung ab. Hinzu kommt, dass diese Form der Approximation einer unscharfen Beobachtung die Forderung (4.13) verletzt. Wir können das leicht am Beispiel der Körpergröße von Herrn G. verdeutlichen. In der Synthese wurde seine Größe allein durch die Grundbegriffe *normal* und *groß* beschrieben (s. S. 45). Da diese Kombination wegen der linearen Unabhängigkeit der Plausibilitätsdichten von  $\mathcal{P}$  eindeutig ist, muß nach Forderung (4.13) auch die Analyse mit diesen zwei Grundbegriffen auskommen. Da sich aber z. B. die vagen Charakteristika der Beobachtung  $\tilde{A}$  und des Grundbegriffs *klein* überlappen, wird auch der Begriff *klein* mit positiver Wahrscheinlichkeit belegt. Die Unschärfe der Aussage wird also durch den Analysevorgang größer als nötig; insbesondere ist das Ergebnis nicht mit der Synthese von  $\tilde{A}$  konsistent. Wir müssen also feststellen, dass dieser Ansatz 1 keine Analyse im Sinne der Definition 6 ist.

**Ansatz 2** In der Praxis verarbeitet man Begriffe oft auf der Basis der zugehörigen Plausibilitätsdichten, anstatt immer auf die Kontexte und vagen Charakteristika zurückzugreifen. Das ist insbesondere dann unausweichlich, wenn das Speichern aller vagen Charakteristika wegen einer zu großen Anzahl von Kontexten nicht handhabbar ist oder wenn die Plausibilitätsdichten wie in Abschnitt 6.1 durch eine

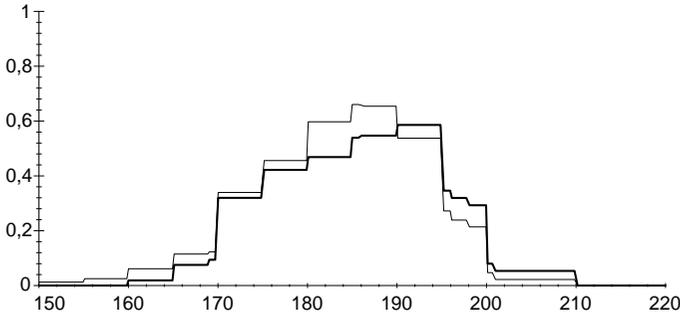


Abbildung 4.6: Approximationen der Größe von Herrn G. nach Ansatz 1 (dünn) und Ansatz 2 (fett)

Gruppierung von Daten gewonnen werden. Im zweiten Fall liegen nur die Plausibilitätsdichten vor. Wir müssen also davon ausgehen, dass die Grundbegriffe im Zweifelsfall nur in Form ihrer Plausibilitätsdichte gegeben sind. Wenn wir nun eine unscharfe Beobachtung mithilfe der Grundbegriffe ausdrücken wollen, dann ist das gleichbedeutend damit, die Plausibilitätsdichte der Beobachtung mit den Dichten der Grundbegriffe darzustellen. Wir können daher die Beschreibung der Beobachtung durch Grundbegriffe als Funktionsapproximation interpretieren: Berechne die Gewichte  $P_{\mathcal{P}}(\tilde{B}|\tilde{A})$  so, dass die gewichtete Kombination

$$\sum_{B \in \mathcal{P}} P_{\mathcal{P}}(\tilde{B}|\tilde{A}) \cdot pl_{\tilde{B}}(u) \quad (4.17)$$

die Beobachtung  $pl_{\tilde{A}}(u)$  möglichst gut annähert.

Funktionsapproximation ist eine klassische Aufgabe aus der Numerik. Dort werden Linearkombinationen von Basisfunktionen, z. B. Polynome, benutzt um Funktionen zu approximieren. Wir sehen, dass die semantische Kombination von Grundbegriffen mathematisch identisch ist mit der algebraischen Kombination von Basisfunktionen. Wir können damit die Kenntnisse aus der Mathematik nutzen, um die Approximationsaufgabe zu lösen.

Zur Bestimmung der Gewichte greifen wir auf Funktionsapproximation in Prähilberträumen zurück. Bei diesem Ansatz wird die Fehlerfunktion

$$F = \|pl_{\tilde{A}} - \sum_{B \in \mathcal{P}} \alpha_B \cdot pl_B\|_2 \quad (4.18)$$

minimiert. Aus der Theorie ist bekannt, dass dieses Optimierungsproblem eine eindeutige Lösung hat, wenn die Funktionen  $pl_B$  linear unabhängig sind, was wir vorausgesetzt haben. In der Praxis wird man unter Umständen die diskrete Version des Optimierungsverfahrens bevorzugen, die unter dem Namen *Methode der*

*kleinsten Quadrate*<sup>2</sup> bekannt ist. Eine Zusammenfassung der nötigen Theorie und Lösungsverfahren befindet sich in Anhang B.2.

Die Optimierung bestimmt die Gewichte  $\alpha_{\tilde{B}}$ , mit denen wir die Approximation  $\tilde{A}'$  der Beobachtung  $\tilde{A}$  berechnen:

$$pl_{\tilde{A}}(u) \approx pl_{\tilde{A}'}(u) = \sum_{\tilde{B} \in \mathcal{P}} \alpha_{\tilde{B}} \cdot pl_{\tilde{B}}(u) \quad (4.19)$$

Ziel der Analyse ist die Berechnung der Begriffswahrscheinlichkeiten  $P_{\mathcal{P}}(\tilde{B}|\tilde{A})$ , die wir formal scheinbar mit den Gewichten  $\alpha_{\tilde{B}}$  identifizieren können, wie der Vergleich von (4.17) und (4.19) verdeutlicht. Ein bekanntes Problem bei den genannten Approximationsverfahren ist aber, dass sie u. U. negative Gewichte  $\alpha_{\tilde{B}}$  berechnen, die wir nicht als Wahrscheinlichkeiten interpretieren können (vgl. auch [NK98]).

Aus Sicht der Funktionsapproximation ist der Ursprung negativer Gewichte leicht verständlich. Dazu betrachten wir Abbildung 4.7, in der wir Grundbegriffe  $\tilde{B}_1$ ,  $\tilde{B}_2$ ,  $\tilde{B}_3$  und eine unscharfe Menge  $\tilde{A}$  sehen. Der Träger von  $\tilde{A}$  überschneidet sich

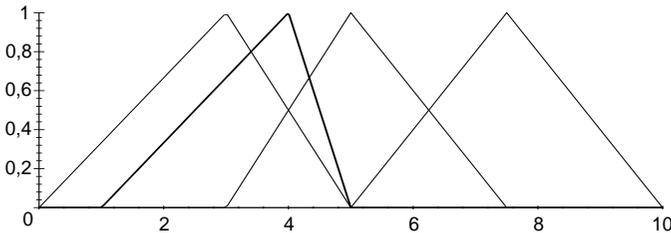


Abbildung 4.7: Grundbegriffe  $\tilde{B}_1$ ,  $\tilde{B}_2$ ,  $\tilde{B}_3$  (dünn, von links nach rechts) und eine unscharfe Menge  $\tilde{A}$  (fett)

nur mit den Trägern von  $\tilde{B}_1$  und  $\tilde{B}_2$ . In der gewichteten Kombination der Plausibilitätsdichten können daher nur  $\tilde{B}_1$  und  $\tilde{B}_2$  einen positiven Beitrag liefern. Die Funktionsapproximation liefert dementsprechend  $\alpha_1 = 0,77$  und  $\alpha_2 = 0,23$ . Der Grundbegriff  $\tilde{B}_3$  hingegen wird mit  $\alpha_3 = -0,06$  negativ gewichtet. Die Gewichte wurden so normiert, dass die Summe der positiven Werte eins ergibt.

Die Ursache für die negative Gewichtung finden wir in dem Bereich  $S = \text{supp}(\tilde{B}_2) \cap \text{supp}(\tilde{B}_3)$ . Das positive Gewicht von  $\tilde{B}_2$  bewirkt, dass die Approximation von  $pl_{\tilde{A}}$  im Bereich  $S$  positiv ist;  $pl_{\tilde{A}}$  selbst ist in  $S$  konstant null. Wenn wir

<sup>2</sup>GAUSS benutzte das Verfahren 1801, um die Bahn des Planetoiden Ceres zu berechnen. Veröffentlicht wurde es zuerst 1806 von LEGENDRE [HH92].

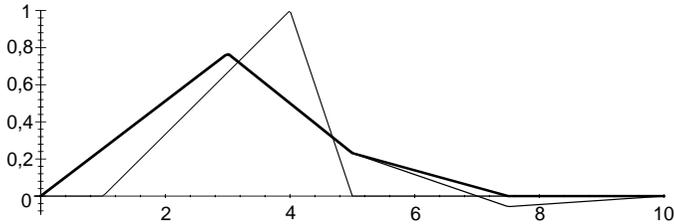


Abbildung 4.8: Beobachtung (dünn), Approximation mit positiven Gewichten (dick) und inklusive negativem Gewicht (mittel, mit negativem Bereich)

$\tilde{B}_3$  negativ gewichten, dann können wir die Approximation in  $S$  kleiner und damit insgesamt präziser machen.

Wir können uns also größere Approximationsgenauigkeit mit zwei Nachteilen erkaufen: Ein negatives Gewicht ist nicht als Wahrscheinlichkeit interpretierbar und die Approximation von  $pl_{\tilde{A}}$  hat negative Bereiche. Da diese beiden Fakten in unserer Formalisierung unscharfer Informationen nicht tolerierbar sind, benutzen wir ein Verfahren, das negative Gewichte ausschließt. Es beruht auf der Feststellung, dass ein Grundbegriff nur dann einen positiven Beitrag zur Approximation der Beobachtung leisten kann, wenn sich die Träger des Grundbegriffs und der Beobachtung überschneiden.

### Analyseverfahren nach Ansatz 2

1. Bestimme die Menge  $M$  der Grundbegriffe, deren Träger einen nichtleeren Schnitt mit dem Träger von  $\tilde{A}$  haben:

$$M := \{\tilde{B} \in \mathcal{P} \mid \text{supp}(\tilde{B}) \cap \text{supp}(\tilde{A}) \neq \emptyset\} \quad (4.20)$$

2. Berechne durch eine Minimierung der Fehlerfunktion (4.18) die Gewichte  $\alpha_{\tilde{B}}$  für alle  $\tilde{B} \in \mathcal{P}$ .
3. Falls negative Gewichte  $\alpha_{\tilde{B}}$  auftreten, entferne die zugehörigen Grundbegriffe aus  $M$ .
4. Setze  $\alpha_{\tilde{B}} = 0$  für alle  $\tilde{B} \notin M$  und normiere die Gewichte  $\alpha_{\tilde{B}}$ , so dass  $\sum_{\tilde{B} \in \mathcal{P}} \alpha_{\tilde{B}} = 1$ .
5. Definiere die Begriffswahrscheinlichkeiten:

$$\forall \tilde{B} \in \mathcal{P} : P_{\mathcal{P}}(\tilde{B} | \tilde{A}) := \alpha_{\tilde{B}} \quad (4.21)$$

Dieses Verfahren zur Berechnung der Begriffswahrscheinlichkeiten erfüllt die gestellten Anforderungen an eine Analyse:

**Satz 7** Die Bestimmung der Begriffswahrscheinlichkeiten  $P_{\mathcal{P}}(\tilde{B}|\tilde{A})$  durch (4.21) mit dem Analyseverfahren nach Ansatz 2 ist eine Analyse nach Definition 6.

*Beweis:*  $P_{\mathcal{P}}$  ist offensichtlich ein Wahrscheinlichkeitsmaß auf  $(\mathcal{P}, 2^{\mathcal{P}})$ . Wir müssen noch zeigen, dass Forderung (4.13) erfüllt wird. Nach den Voraussetzungen von Def. 6 gilt

$$pl_{\tilde{A}} \equiv \sum_{\tilde{B} \in \mathcal{P}} P'_{\mathcal{P}}(\tilde{B}) \cdot pl_{\tilde{B}}$$

Mit  $\alpha_{\tilde{B}} = P'_{\mathcal{P}}(\tilde{B})$  existieren also Gewichte, welche die Fehlerfunktion (4.18) zu null minimieren und somit Lösung des Optimierungsproblems sind. Da die Lösung des Optimierungsproblems wegen der linearen Unabhängigkeit der Plausibilitätsdichten eindeutig ist, gilt  $P_{\mathcal{P}}(\tilde{B}|\tilde{A}) := \alpha_{\tilde{B}} = P'_{\mathcal{P}}(\tilde{B})$ .  $\square$

Wenden wir das Verfahren auf die Körpergröße von Herrn G. aus Abbildung 4.5 an, so ergibt sich für die Approximation  $pl_{\tilde{A}}(u) = 0,414 \cdot pl_{normal}(u) + 0,586 \cdot pl_{gro\beta}(u)$ , s. Abb. 4.6. Dieses Ergebnis bestätigt die Umfrage aus der Synthese in Abschnitt 4.2.2, bei der 59% der Befragten Herrn G. für *groß* und 41% für *normal* groß hielten. Im Vergleich mit dem Ergebnis von Ansatz 1 ist das Ergebnis erwartungsgemäß etwas ungenauer. Bei Ansatz 1 kann jedem vagen Charakteristikum ein eigenes Gewicht zugeordnet werden. Im Gegensatz dazu müssen bei Ansatz 2 alle vagen Charakteristika eines Grundbegriffs gleich gewichtet werden. Die Approximation aus Ansatz 2 können wir dafür aber als eine Kombination von *normal* und *groß* interpretieren, während das Ergebnis von Ansatz 1 nicht mit den Grundbegriffen beschrieben werden kann. Da die Interpretierbarkeit ein wesentlicher Bestandteil unscharfer Systeme im Allgemeinen und dieser Arbeit im Speziellen ist, wählen wir von nun an immer Ansatz 2 für die Analyse einer unscharfen Beobachtung.

Mit Ansatz 2 haben wir ein Verfahren entworfen, das per Analyse unscharfe Mengen auf feingranularer Ebene in verständliche unscharfe Aussagen auf grobgranularer Ebene übersetzt. Das Vorgehen ist mit der Synthese konsistent: Wir können eine symbolische Kombination unscharfer Grundbegriffe durch die Synthese in eine unscharfe Menge übersetzen, und deren Analyse berechnet wiederum dieselbe Kombination als symbolische Interpretation, vgl. Abb. 4.2. Als Ergebnis können wir unscharfe Informationen effizient und verständlich auf grobgranularer Ebene verarbeiten, ohne die Exaktheit der feingranularen Ebene zu verlieren.

### 4.3 Zusammenfassung

In diesem Kapitel haben wir gezeigt, wie man ein Universum vergrößert. Wir abstrahieren von Einzelheiten und partitionieren das Universum durch wenige unscharfe Mengen, die mit unscharfen Begriffen belegt sind. Die unscharfen Begriffe haben in der jeweiligen Anwendung eine konkrete Bedeutung. Daher begeben wir uns mithilfe einer Partitionierung von der feingranularen Ebene der Einzelheiten auf die grobgranulare Ebene unseres Verständnisses.

Damit eine unscharfe Partition einerseits das gesamte Universum erfasst und andererseits unterscheidbare Aussagen trennt, haben wir die Eigenschaften Vollständigkeit und Konsistenz von scharfen Partitionen verallgemeinert. Daraus resultiert die Überdeckungseigenschaft unscharfer Partitionen, die bisher nur als Daumenregel vorgeschlagen wurde [DHR93] oder implizit Teil von Modellannahmen war [DMP93].

Die unscharfen Begriffe einer unscharfen Partition nennen wir Grundbegriffe. Sie stellen den Wortschatz dar, mit dessen Hilfe wir unscharfe Beobachtungen auf leicht verständliche Weise beschreiben. Zu diesem Zweck schlagen wir in dieser Arbeit als grundlegend neues Verfahren die gewichtete Kombination von Grundbegriffen vor. Die Grundbegriffe der Partition bilden ein Erzeugendensystem für unscharfe Aussagen, die einerseits leicht verständlich durch die Grundbegriffe beschrieben werden und andererseits exakt auf der feingranularen Ebene der Einzelheiten repräsentiert sind. Formal erweitern wir dafür das Kontextmodell durch einen wahrscheinlichkeitstheoretischen Aufbau. Damit steht das Verfahren auf der soliden Basis einer konkreten Semantik. Aus Sicht der Praxis ist dies wichtig, weil damit erstens die Bedeutung von Unschärfe wohldefiniert ist und wir zweitens praktische Verfahren angeben können, um unscharfe Aussagen aus Daten zu generieren.

Mit der Synthese unscharfer Mengen haben wir gezeigt, wie man aus der gewichteten Kombination von Grundbegriffen eine unscharfe Menge berechnet, also eine grobgranulare in eine feingranulare Darstellung übersetzt. Umgekehrt können wir mit der Analyse eine unscharfe Menge durch eine gewichtete Kombination der Grundbegriffe verständlich machen. Bestehende Ansätze versuchen zu diesem Zweck, die unscharfe Menge zu klassifizieren. Man könnte etwa eine Form von Konsistenz der unscharfen Menge mit jedem Grundbegriff messen [CFP82]. Wir hingegen rekonstruieren die unscharfe Menge möglichst exakt durch die Grundbegriffe. So können wir erstmalig sicherstellen, dass feingranulare und grobgranulare Aussagen in beiden Richtungen eindeutig ineinander überführbar sind und somit die Semantik unscharfer Aussagen wohldefiniert ist. Verständlichkeit und Exaktheit unscharfer Aussagen sind damit immer gleichzeitig gegeben.

---

In den folgenden Kapiteln erweist sich die Idee, alle unscharfen Aussagen durch die Grundbegriffe darzustellen, als äußerst elegant. Alle Berechnungen mit unscharfen Aussagen können wir auf Berechnungen mit den Grundbegriffen zurückführen. Das sichert erstens die Effizienz der Verfahren, da wir in Anwendungen normalerweise nur wenige Grundbegriffe definieren. Zweitens ist aber auch die Verständlichkeit unscharfer Ergebnisse gewährleistet, denn sie werden durch bekannte Begriffe ausgedrückt.



## Kapitel 5

---

# Schließen mit unscharfen Begriffen

Für das Schließen mit unscharfen Begriffen beschäftigen wir uns in diesem Kapitel mit zwei Problemstellungen: der Modellierung von Regeln für die Wissensbasis und der Herleitung eines Inferenzmechanismus. Wir greifen dabei auf die Definitionen von Wissensbasis, Beobachtung und Inferenz aus Abschnitt 2.2 zurück.

In Abschnitt 5.1 zeigen wir, wie Zusammenhänge zwischen Variablen durch Regeln auf der Basis unscharfer Begriffe dargestellt werden. Dabei müssen wir sicherstellen, dass sich die Semantik der Regeln mit der Semantik unscharfer Begriffe und insbesondere der gewichteten Kombination von Grundbegriffen verträgt. Wir erreichen dies durch die Modellierung von Regeln als bedingte Wahrscheinlichkeiten unscharfer Begriffe.

Der zweite Teil des Kapitels widmet sich mit Abschnitt 5.2 der Inferenz, also der Frage, wie man aus mehrstufigen Regelbasen und Beobachtungen Schlüsse zieht. Dabei haben wir uns als Ziel gesetzt, dass der Inferenzmechanismus die Probleme herkömmlicher Verfahren beseitigt: Um die Effizienz der Berechnungen zu gewährleisten, sollen mehrstufige Inferenzen auf grobgranularer Ebene berechnet werden. Darüber hinaus soll der Schluss des Systems verständlich auf grobgranularer Ebene und wohldefiniert auf feingranularer Ebene vorliegen.

Um diesen Anforderungen zu genügen, benutzen wir als Antezedenzen und Konsequenzen der Regeln Grundbegriffe und berechnen die Inferenz auf grobgranularer Ebene über die Begriffswahrscheinlichkeiten, die wir in Kapitel 4 eingeführt haben. Ein unscharfer Schluss ist dann als gewichtete Kombination von Grundbegriffen leicht verständlich und kann durch eine Synthese in eine unscharfe Menge übersetzt werden. Damit zeigen wir erstmals, wie man auf grobgranularer Ebene mit den Grundbegriffen symbolisch „rechnen“ kann, ohne die feingranulare Exaktheit der Ergebnisse zu verlieren.

In Abschnitt 5.2 haben wir angenommen, dass die Antezedenzen und Konsequenzen der Regeln Grundbegriffe sind. Diese Einschränkung heben wir in Abschnitt 5.3 auf. Wir führen Regeln mit beliebigen unscharfen Begriffen als Antezedenzen und Konsequenzen auf Regeln mit unscharfen Grundbegriffen zurück. Auf diese Weise können wir beliebige Regeln verständlich mit den Grundbegriffen darstellen und effizient mit unserem Inferenzmechanismus verarbeiten.

Bei praktischen Anwendungen unscharfen Schließens ist es wichtig, die Zuverlässigkeit der Ergebnisse beurteilen zu können. Damit ein Maß für die Zuverlässigkeit aussagekräftig ist, muss es über eine konkrete Semantik verfügen. In Abschnitt 5.4 zeigen wir, wie man aus der Semantik unscharfer Begriffe ein Zuverlässigkeitsmaß für unscharfe Schlüsse ableitet, das diese Anforderung erfüllt.

Abschließend vergleichen wir in Abschnitt 5.5 den Rechenaufwand unseres Inferenzverfahrens mit dem Aufwand anderer Ansätze. Wie in Abschnitt 1.3 gefordert, hängt der Aufwand der Inferenz von der Anzahl der Grundbegriffe bzw. der Regeln ab und nicht wie bei anderen Verfahren von der Anzahl der Einzelheiten.

## 5.1 Grobgranulare Regeln

Wie bereits in den Grundlagen in Abschnitt 2.2 besprochen, verstehen wir in dieser Arbeit unter einer Regel einen gerichteten Zusammenhang zwischen unscharfen Aussagen, den wir durch eine Formulierung der Bauart „Wenn  $x = \tilde{A}$ , dann  $y = \tilde{B}$ “ ausdrücken.  $\tilde{A}$  und  $\tilde{B}$  sind dabei unscharfe Begriffe über den Universen der Variablen  $x$  bzw.  $y$ . Der Einfachheit halber schreiben wir statt der Regel „Wenn  $x = \tilde{A}$ , dann  $y = \tilde{B}$ “ kurz  $\tilde{A} \longrightarrow \tilde{B}$ .

Konsistent mit der gewichteten Kombination unscharfer Begriffe bauen wir unsere Theorie zum unscharfen Schließen auf der Interpretation einer Regel als bedingte Begriffswahrscheinlichkeit auf. In Abschnitt 6.3 zeigen wir dann im Rahmen verschiedener Repräsentationen von Wissen, wie man partielle Abbildungen und logische Verknüpfungen mit bedingten Wahrscheinlichkeiten darstellt. In dieser Weise kann regelbasiertes Wissen auf verschiedenen Semantiken beruhen und trotzdem durch den gleichen mathematischen Formalismus repräsentiert werden.

In diesem Abschnitt, wie auch bei der Inferenz in Abschnitt 5.2, gehen wir davon aus, dass die unscharfen Begriffe  $\tilde{A}$  und  $\tilde{B}$  den unscharfen Partitionen  $\mathcal{P}_X = \{\tilde{A}_1, \tilde{A}_2, \dots\}$ ,  $\mathcal{P}_Y = \{\tilde{B}_1, \tilde{B}_2, \dots\}$  der Universen  $X$ ,  $Y$  entnommen sind. Unsere Regeln basieren damit nur auf Grundbegriffen. Aus Sicht der Praxis ist dieses Vorgehen unmittelbar einleuchtend, da die Partitionen gerade so gewählt werden, dass auch die Zusammenhänge zwischen verschiedenen Variablen direkt durch die Grundbegriffe ausgedrückt werden können. Insbesondere wenn die Informationen über die Zusammenhänge aber in gemessenen Daten bestehen, sind die Zusammenhänge u. U. von den Grundbegriffen losgelöst. In diesem Fall – die unscharfen Werte  $\tilde{A}$  und  $\tilde{B}$  sind nicht in  $\mathcal{P}_X$  bzw.  $\mathcal{P}_Y$  enthalten – werden wir sie durch Analyse oder Synthese als Kombination von Grundbegriffen ausdrücken, was in Abschnitt 5.3 ausführlich diskutiert wird. Die Beschränkung auf unscharfe Grundbegriffe in den Antezedenzien und Konsequenzen der Regeln beschränkt also nicht die Anwendbarkeit unserer Theorie.

Wir betrachten zunächst nur Regeln mit eindimensionalem Antezedens und eindimensionalem Konsequens. Mehrdimensionale Antezedenzen werden in Abschnitt 5.2.3 behandelt. Eine Regel mit mehrdimensionalem Konsequens „Wenn  $x = \tilde{A}$ , dann  $y = \tilde{B}$  und  $z = \tilde{C}$ “ ersetzen wir o. B. d. A. durch die beiden Regeln „Wenn  $x = \tilde{A}$ , dann  $y = \tilde{B}$ “ und „Wenn  $x = \tilde{A}$ , dann  $z = \tilde{C}$ “.

Eine Regel  $\tilde{A} \longrightarrow \tilde{B}$  weist der Größe  $y$  den Wert  $\tilde{B}$  unter der Annahme zu, dass  $x$  den Wert  $\tilde{A}$  annimmt. In dieser Hinsicht deuten wir die Regel wie z. B. bei BAYESSCHEN Netzen [Pea88, RN95] als sichere bedingte Wahrscheinlichkeit: Wenn  $\tilde{B}$  die einzig mögliche Schlussfolgerung für  $y$  unter der Annahme von  $\tilde{A}$  ist, so gilt  $P_{\mathcal{P}}(\tilde{B}|\tilde{A}) = 1$ . Um Unsicherheit über das Konsequens ausdrücken zu können, verallgemeinern wir die Modellierung, indem wir beliebige Wahrscheinlichkeiten  $P_{\mathcal{P}}(\tilde{B}|\tilde{A})$  zulassen. Formal definieren wir zu diesem Zweck wiederum einen Begriffsmaßraum  $(\mathcal{P}, 2^{\mathcal{P}}, P_{\mathcal{P}})$ . In diesem Fall bilden wir das kartesische Produkt  $\mathcal{P} = \mathcal{P}_X \times \mathcal{P}_Y$ , auf dem die Regeln als bedingte Wahrscheinlichkeiten definiert werden. Für eine gewichtete Regel notieren wir  $\tilde{A} \xrightarrow{P_{\mathcal{P}}(\tilde{B}|\tilde{A})} \tilde{B}$ . Aus der Wahrscheinlichkeitsrechnung folgt die Eigenschaft

$$\forall \tilde{A} \in \mathcal{P}_X : \sum_{\tilde{B} \in \mathcal{P}_Y} P_{\mathcal{P}}(\tilde{B}|\tilde{A}) = 1, \quad (5.1)$$

die anschaulich bedeutet, dass bei Gültigkeit von  $x = \tilde{A}$  die Variable  $y$  nur Werte aus  $\mathcal{P}_Y$  annehmen kann. Der Hintergrund dieser Eigenschaft ist die Annahme einer abgeschlossenen Welt<sup>1</sup>.

Als Beispiel betrachten wir den Zusammenhang zwischen der Körpergröße  $x$  und dem Gewicht  $y$  eines Menschen. Wir gehen dazu von den Partitionen  $\mathcal{P}_X = \{\text{sehr klein, klein, normal, groß, sehr groß}\}$  und  $\mathcal{P}_Y = \{\text{sehr leicht, leicht, durchschnittlich, schwer, sehr schwer}\}$  aus. Wir könnten nun eine Gruppe von Experten bitten, jeder unscharfen Körpergröße aus  $\mathcal{P}_X$  ein oder mehrere unscharfe Körpergewichte aus  $\mathcal{P}_Y$  zuzuordnen. Aus dieser Umfrage können wir direkt die bedingten Wahrscheinlichkeiten  $P_{\mathcal{P}}(\tilde{B}_j|\tilde{A}_i)$  als relative Häufigkeiten ablesen. Ein mögliches Ergebnis zeigt Tab. 5.1 ausschnittsweise.

Das Beispiel deutet schon an, dass es bei gewichteten Regeln um mehr geht als den reinen Ausdruck von Unsicherheit über das Konsequens. Wenn wir den Inferenzmechanismus hergeleitet haben, werden wir in Abschnitt 5.3 sehen, dass wir auf diese Weise Regeln konstruieren können, die nicht auf den Grundbegriffen basieren. So können wir z. B. die drei Regeln mit dem Antezedens *klein* aus Tab. 5.1 zu einer Regel zusammenfassen. Das Konsequens bilden wir durch eine gewichtete Kombination der Grundbegriffe *sehr leicht, leicht* und *durchschnittlich*. Wir

<sup>1</sup>engl. *closed world assumption*

|            |                      |                  |
|------------|----------------------|------------------|
| sehr klein | $\xrightarrow{0,7}$  | sehr leicht      |
| sehr klein | $\xrightarrow{0,3}$  | leicht           |
| klein      | $\xrightarrow{0,2}$  | sehr leicht      |
| klein      | $\xrightarrow{0,5}$  | leicht           |
| klein      | $\xrightarrow{0,3}$  | durchschnittlich |
| normal     | $\xrightarrow{0,25}$ | leicht           |
| normal     | $\xrightarrow{0,5}$  | durchschnittlich |
| normal     | $\xrightarrow{0,25}$ | schwer           |
|            |                      | etc.             |

Tabelle 5.1: Teil einer Regelbasis für den Zusammenhang zwischen Körpergröße und Körpergewicht

werden in Abschnitt 5.3 mit Satz 10 beweisen, dass wir die Regelgewichte als Wahrscheinlichkeiten der drei Grundbegriffe interpretieren können. Dazu benötigen wir den Inferenzmechanismus, den wir im folgenden Abschnitt herleiten.

**Bemerkung 1** *Die vorgestellte Formalisierung von Regeln unterscheidet sich grundlegend von der Art, die bei der Inferenz nach dem Prinzip von Kombination/Projektion benutzt wird. Wie auch bei unscharfen Produktionsregeln stellen wir eine Regel auf grobgranularer Ebene dar. Beim Prinzip von Kombination/Projektion hingegen wird eine grobgranular formulierte Regel zwischen Variablen  $x$  und  $y$  in einen feingranularen Zusammenhang zwischen  $x$  und  $y$  übersetzt und auch auf dieser Ebene verarbeitet. Daraus resultieren die Effizienzprobleme und die mangelnde Verständlichkeit der Schlüsse, wie wir in Abschnitt 3 gesehen haben. Wie der nächste Abschnitt zeigt, können wir diese Probleme mit unserem Ansatz lösen.*

## 5.2 Inferenz

Aus der Modellierung von Regeln als bedingte Wahrscheinlichkeiten leiten wir in diesem Abschnitt ein Inferenzverfahren ab. Wir gehen von einem Regelsatz als Wissensbasis aus, im Fall eines einstufigen Systems von gewichteten Regeln  $\tilde{A}_i \xrightarrow{P_P(\tilde{B}_j|\tilde{A}_i)} \tilde{B}_j$ . Aus der Kenntnis, dass  $x$  einen unscharfen Wert  $\tilde{A}$  annimmt, möchten wir durch Inferenz einen Schluss für den Wert von  $y$  ziehen. Das Basisverfahren der Inferenz beschreiben wir in 5.2.1 für einstufige Inferenzen mit eindimensionalen Antezedenzen. Diese Einschränkung wird in den folgenden Abschnitten aufgehoben. Weiterhin zeigen wir, dass das Inferenzverfahren im Gegensatz zu den meisten anderen Verfahren den Modus Ponens auch bei paralleler Aus-

wertung mehrerer Regeln erfüllt. Besonders bei mehrstufigem Schließen ist dies wichtig, weil andernfalls mit jeder weiteren Stufe die Unsicherheit der Schlüsse zunimmt.

Das Inferenzverfahren bauen wir auf den Begriffswahrscheinlichkeiten auf, weshalb sich die Inferenz vollständig auf grobgranularer Ebene bewegt. Damit sichern wir die Effizienz des Verfahrens und die Verständlichkeit der Schlüsse. Insbesondere bei mehrstufigen Inferenzen müssen wir aber gewährleisten, dass die feingranularen Ergebnisse exakt bleiben, wenn wir auf grobgranularer Ebene schließen. In Abschnitt 5.2.2 beweisen wir, dass wir keinen Verlust an Exaktheit erleiden. Auf diese Weise gewinnen wir Effizienz und Verständlichkeit gegenüber anderen Inferenzverfahren, ohne ungenauer zu werden.

### 5.2.1 Basisverfahren der Inferenz

Da wir die Regeln als bedingte Wahrscheinlichkeiten modelliert haben, lehnen wir das Inferenzverfahren an die *Formel von der totalen Wahrscheinlichkeit* an:

$$P(B_j) = \sum_i P(B_j|A_i) \cdot P(A_i) \quad (5.2)$$

mit scharfen Mengen  $A_i \subseteq \mathcal{X}$ ,  $B_j \subseteq \mathcal{Y}$ . Für die Gültigkeit der Formel müssen die  $A_i$  paarweise disjunkt sein und das Universum  $\mathcal{X}$  überdecken:  $\bigcup_i A_i = \mathcal{X}$ . Aus der Kenntnis der bedingten Wahrscheinlichkeiten  $P(B_j|A_i)$  und der Beobachtungen  $P(A_i)$  über das Eintreffen der Ereignisse  $A_i$  lässt sich damit auf die Eintrittswahrscheinlichkeit  $P(B_j)$  schließen. In unserem Modell sind die Ereignisse  $A_i$  und  $B_j$  unscharfe Grundbegriffe  $\tilde{A}_i \in \mathcal{P}_{\mathcal{X}}$ ,  $\tilde{B}_j \in \mathcal{P}_{\mathcal{Y}}$ , die durch Plausibilitätsdichten dargestellt werden. Die bedingten Wahrscheinlichkeiten  $P_{\mathcal{P}}(\tilde{B}_j|\tilde{A}_i)$  modellieren Regeln zwischen den Grundbegriffen wie in Abschnitt 5.1 dargestellt. Wir operieren also ausschließlich im Begriffsmaßraum  $(\mathcal{P}, 2^{\mathcal{P}}, P_{\mathcal{P}})$ . Die Grundbegriffe werden in diesem als diskrete Elemente von  $\mathcal{P}$  dargestellt, wie wir in Abschnitt 4.2.1 beschrieben haben, und sind damit einerseits disjunkt und überdecken andererseits die Menge  $\mathcal{P}$ . Wir dürfen also die Formel von der totalen Wahrscheinlichkeit anwenden.

Bei Kenntnis der Regeln müssen wir die Wahrscheinlichkeiten  $P_{\mathcal{P}}(\tilde{A}_i)^2$  aus einer i. A. unscharfen Information über die Variable  $x$  bestimmen, um Formel (5.2) für die Inferenz anwenden zu können. Diesen Schritt haben wir bereits in Abschnitt 4.2 erledigt. Im Fall der Synthese sind die Wahrscheinlichkeiten  $P_{\mathcal{P}}(\tilde{A}_i)$  direkt durch die Beobachtung gegeben. Liegt dagegen eine unscharfe Beobachtung  $\tilde{A}$  vor, so berechnen wir die  $P_{\mathcal{P}}(\tilde{A}_i)$  durch eine Analyse, wobei wir wie in Abschnitt 4.2.3  $P_{\mathcal{P}}(\tilde{A}_i)$  mit  $P_{\mathcal{P}}(\tilde{A}_i|\tilde{A})$  identifizieren.

<sup>2</sup>als Kurzform von  $P_{\mathcal{P}}(\{\tilde{A}_i\} \times \mathcal{P}_{\mathcal{Y}})$

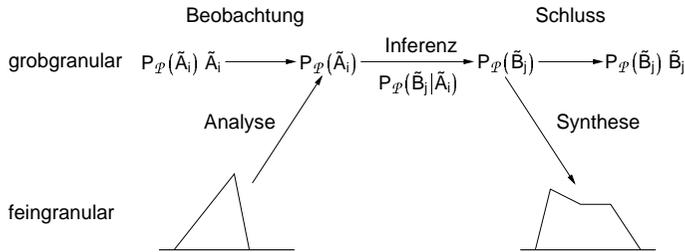


Abbildung 5.1: Schließen mit uncharfen Begriffen

Insgesamt müssen wir zur Berechnung eines Schlusses  $\tilde{B}$  also drei Schritte durchführen, wie Abb. 5.1 illustriert – die Bezeichnung  $P_P(\tilde{A}_i)$   $\tilde{A}_i$  in der Abbildung steht dabei für die grobgranularen Aussagen „Der Grundbegriff  $\tilde{A}_i$  gilt mit Wahrscheinlichkeit  $P_P(\tilde{A}_i)$ .“ Im ersten Schritt der Berechnung werden die Begriffswahrscheinlichkeiten  $P_P(\tilde{A}_i)$  bestimmt, die im zweiten Schritt mit den bedingten Wahrscheinlichkeiten  $P_P(\tilde{B}_j|\tilde{A}_i)$  verknüpft werden:

$$P_P(\tilde{B}_j) = \sum_i P_P(\tilde{B}_j|\tilde{A}_i) \cdot P_P(\tilde{A}_i) \quad (5.3)$$

Resultat sind die Wahrscheinlichkeiten  $P_P(\tilde{B}_j)$ , mit denen im dritten Schritt durch gewichtete Kombination der Begriffe  $\tilde{B}_j$  die Konklusion  $\tilde{B}$  berechnet wird:

$$pl_{\tilde{B}}(v) = \sum_j P_P(\tilde{B}_j) \cdot pl_{\tilde{B}_j}(v) \quad (5.4)$$

Die Berechnungsvorschrift (5.3) für den zweiten Schritt, den Kern der Inferenz, können wir als Multiplikation einer Matrix mit einem Vektor darstellen. Mit den Vektoren  $\vec{a} = (P_P(\tilde{A}_i))_i$ ,  $\vec{b} = (P_P(\tilde{B}_j))_j$  und der Matrix  $G = (P_P(\tilde{B}_j|\tilde{A}_i))_{ij}$  gilt

$$\vec{b} = G^t \cdot \vec{a} \quad (5.5)$$

Wir sehen, dass die Inferenz selbst nur auf der Ebene der Begriffswahrscheinlichkeiten, der grobgranularen Ebene, berechnet wird. Zur Bestimmung der Wahrscheinlichkeiten  $P_P(\tilde{A}_i)$  müssen wir uns nur im Fall einer Analyse auf die Ebene der Einzelheiten begeben, weil der nötige Approximationsvorgang eine Optimierung über dem Universum  $\mathcal{X}$  ist. Nach der Inferenz liegt der berechnete Schluss als Kombination von Grundbegriffen vor. Je nach Anwendung verwerten wir das Ergebnis auf grobgranularer Ebene – z. B. „Der Täter ist mit Wahrscheinlichkeit 0,7 groß und mit 0,3 sehr groß.“ – oder wir berechnen die zugehörige Plausibilitätsdichte der Kombination auf feingranularer Ebene, um z. B. die scharfen Körpergrößen mit der höchsten Plausibilität bestimmen zu können. Das Beispiel in Anhang A illustriert diese beiden Möglichkeiten.

Unsere Interpretation einer Regel haben wir an den Modus Ponens geknüpft: Aus  $\tilde{A}$  und  $\tilde{A} \rightarrow \tilde{B}$  schließen wir  $\tilde{B}$ . Insbesondere verlangen wir bei einem Satz von Regeln  $\tilde{A}_i \rightarrow \tilde{B}_i$  für alle  $i$ , dass aus  $\tilde{A}_i$  nach Anwendung aller Regeln der Schluss  $\tilde{B}_i$  folgt. Wie folgender Satz zeigt, gilt der Modus Ponens in unserem Inferenzmechanismus.

**Satz 8** Seien  $\tilde{A}_i \in \mathcal{P}_X$ ,  $\tilde{B}_j \in \mathcal{P}_Y$  die Grundbegriffe auf den Universen  $X$  und  $Y$ . Die gewichteten Regeln der Wissensbasis seien durch die bedingten Wahrscheinlichkeiten  $P_{\mathcal{P}}(\tilde{B}_j|\tilde{A}_i)$  definiert. Dann wird eine Eingabe  $\tilde{A}_i$  durch das Inferenzverfahren gemäß (5.3) und (5.4) auf eine gewichtete Kombination der Grundbegriffe  $\tilde{B}_j$  mit  $P_{\mathcal{P}}(\tilde{B}_j) = P_{\mathcal{P}}(\tilde{B}_j|\tilde{A}_i)$  abgebildet. Für die Plausibilitätsdichte des Schlusses  $\tilde{B}^l$  gilt somit

$$\forall v \in \mathcal{Y} : pl_{\tilde{B}^l}(v) = \sum_{\tilde{B} \in \mathcal{P}_Y} P_{\mathcal{P}}(\tilde{B}|\tilde{A}_i) \cdot pl_{\tilde{B}}(v) \quad (5.6)$$

Insbesondere gilt im Fall einfacher Regeln  $\tilde{A}_i \xrightarrow{1} \tilde{B}_j$ , d. h.  $\forall i \exists j : P_{\mathcal{P}}(\tilde{B}_j|\tilde{A}_i) = 1$  : Das Inferenzverfahren bildet  $\tilde{A}_i$  auf  $\tilde{B}_j$  ab.

*Beweis:* Die Eingabe eines Grundbegriffs  $\tilde{A}_l$  resultiert durch eine Analyse in den Begriffswahrscheinlichkeiten  $P_{\mathcal{P}}(\tilde{A}_i) = 1_{i=l}$ , da  $\tilde{A}_i$  sich selbst perfekt darstellt. Die Behauptungen folgen direkt durch Einsetzen der  $P_{\mathcal{P}}(\tilde{A}_j)$  in die Gleichungen (5.3) und (5.4).  $\square$

**Bemerkung 2** Im Fall des Prinzips von Kombination/Projektion wird die Forderung des Modus Ponens z. B. benutzt, um aus gegebenen Antezedenzien  $\tilde{A}_i$  und Konsequenzen  $\tilde{B}_i$  eine unscharfe Relation  $\tilde{R}$  zu bestimmen, die das Wissen aller Regeln zusammenfasst.  $\tilde{R}$  wird so berechnet, dass der Modus Ponens unter Anwendung der sup-min-Komposition erfüllt wird:  $\tilde{A}_i \circ \tilde{R} = \tilde{B}_i$  für alle  $i$  [Ped91]. Diese Gleichung unscharfer Relationen muss allerdings nicht lösbar sein; der Modus Ponens wird dann nicht erfüllt.

Speziell von MAMDANIS Ansatz ist bekannt, dass der Modus Ponens bei einer Wissensbasis aus mehreren Regeln nicht gilt. Bildet man beispielsweise eine identische Abbildung über einen Satz von Regeln  $\tilde{A}_i \rightarrow \tilde{A}_i$  nach, so resultiert die Inferenz eines  $\tilde{A}_j$  i. A. nicht in  $\tilde{A}_j$  [WS00]. Das Ergebnis ist unschärfer, da neben  $\tilde{A}_j$  auch andere  $\tilde{A}_i$  einen Beitrag zum Schluss liefern. Diese Eigenschaft bereitet insbesondere bei mehrstufigen Inferenzen Schwierigkeiten. Wiederholte Anwendungen der Regeln auf den Schluss vergrößern nämlich die Unschärfe weiter, bis irgendwann alle  $\tilde{A}_i$  zum Ergebnis beitragen. Der Schluss ist dann so unspezifisch, dass er keinerlei Informationen mehr trägt. Aus praktischer Sicht ist MAMDANIS Verfahren damit nicht für mehrstufiges Schließen geeignet. Diese Feststellung gilt für

alle Spezialfälle des Prinzips von Kombination/Projektion, bei denen die Schlüsse mehrerer Regeln durch Vereinigung zusammengefasst werden.

Da der Modus Ponens gilt, sind die Schlüsse unseres Inferenzverfahrens nicht unsicherer als die Regeln vorgeben. Mehrstufiges Schließen kann als Sequenz einstufiger Inferenzen daran nichts ändern. So können wir die identische Abbildung aus Bemerkung 2 beliebig oft auf eine unscharfe Eingabe anwenden, ohne dass sich die Unschärfe der Ergebnisse vergrößert. Unser Verfahren zum Schließen mit unscharfen Begriffen erfüllt damit eine notwendige Voraussetzung für mehrstufiges Schließen.

## 5.2.2 Mehrstufige Inferenz

In vielen Anwendungen ist es sinnvoll, Zusammenhänge zwischen Variablen mehrstufig über Zwischengrößen zu formulieren anstatt sie direkt anzugeben, wie wir bereits in Abschnitt 1.3 diskutiert haben und wie auch das Beispiel in Anhang A zeigt. Nehmen wir an, wir hätten drei Variablen  $x \in \mathcal{X}$ ,  $y \in \mathcal{Y}$  sowie  $z \in \mathcal{Z}$  und suchten einen Zusammenhang zwischen  $x$  und  $z$ . Wir formulieren unser Wissen in Form von Regeln zwischen Begriffen  $\tilde{A}_i \in \mathcal{P}_{\mathcal{X}}$  und  $\tilde{B}_j^1 \in \mathcal{P}_{\mathcal{Y}}^1$  und Regeln zwischen Begriffen  $\tilde{B}_k^2 \in \mathcal{P}_{\mathcal{Y}}^2$  und  $\tilde{C}_l \in \mathcal{P}_{\mathcal{Z}}$ . Der Inferenzmechanismus muss aus einer Beobachtung für  $x$  auf Werte für  $y$  und daraus auf Werte für  $z$  schließen. Abbildung 5.2 illustriert das naive Vorgehen, die zweistufige Inferenz als Abfolge von zwei einstufigen Inferenzen zu realisieren.

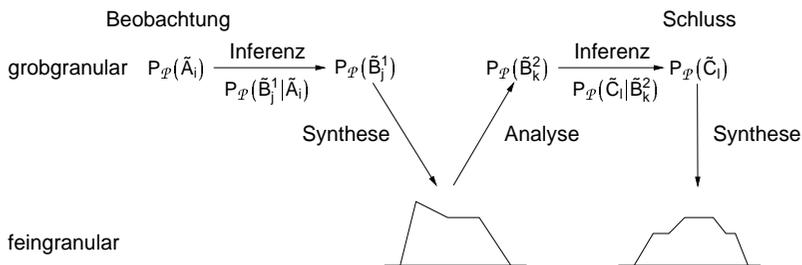


Abbildung 5.2: Naives Vorgehen bei mehrstufiger Inferenz

Das Problem dieser Methode ist, dass der Analysevorgang am Eingang der zweiten Stufe sehr aufwändig ist, weil die notwendigen Berechnungen auf dem Universum  $\mathcal{Y}$  durchgeführt werden müssen. Eines der Ziele dieser Arbeit besteht aber gerade darin, dies zu vermeiden. Wir möchten bei der mehrstufigen Inferenz diesen Analysevorgang zwischen den Inferenzstufen vermeiden, indem wir direkt aus den

Wahrscheinlichkeiten  $P_P(\tilde{B}_j^1)$  die Wahrscheinlichkeiten  $P_P(\tilde{B}_k^2)$  bestimmen. Gelingt uns das, dann sollte der Berechnungsaufwand der Inferenz nur von der Anzahl der Begriffe bzw. Regeln und nicht von der Größe des Universums abhängen.

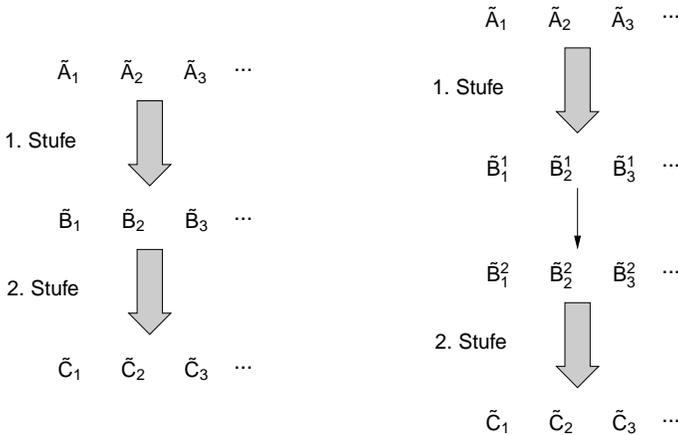


Abbildung 5.3: Mehrstufige Inferenz; links der Standardfall, rechts der allgemeine Fall

**Standardfall** Der Standardfall liegt dann vor, wenn die Konsequenzen der Regeln der ersten Stufe mit den Antezedenzen der Regeln der zweiten Stufe übereinstimmen:  $\mathcal{P}_Y^1 = \mathcal{P}_Y^2 := \mathcal{P}_Y$ . Dieser Fall ist in der Praxis die Regel, denn normalerweise wird ein Universum nur einmal partitioniert. Mehrfache Partitionierungen eines Universums sind nur dann sinnvoll, wenn sie in verschiedenen Zusammenhängen benutzt werden. Aussagen über das Gewicht einer Person zum Beispiel könnten auf fünf Grundbegriffen beruhen, wie in Abschnitt 5.1 beschrieben. Wenn wir auf Basis des Gewichtes entscheiden müssen, ob die Federkraft einer Skibindung ausreicht, dann würden hingegen zwei Begriffe für das Gewicht der Person ausreichen: *passend* und *nicht passend*. Diese beiden Begriffe sind tatsächlich unscharf, da die Entscheidung außerdem von Fahrkönnen und Knochenbau der Person abhängt. Diesen allgemeineren Fall mit zwei Partitionierungen behandeln wir weiter unten. Abbildung 5.3 zeigt links den Standardfall und rechts den allgemeinen Fall.

Im Standardfall resultiert die Inferenz einer Information durch die erste Stufe in den Begriffswahrscheinlichkeiten  $P_P(\tilde{B}_j)$ . Diese bilden auch die Eingabe für die zweite Stufe. Wir müssen also nicht wie in Abb. 5.2 einen feingranularen Schluss  $\tilde{B}$  synthetisieren und wieder aufwändig analysieren. Stattdessen können die Begriffswahrscheinlichkeiten  $P_P(\tilde{B}_j)$  direkt von der ersten zur zweiten Stufe weiter-

gereicht werden, ohne dass wir uns auf die feingranulare Ebene des unterliegenden Universums begeben müssen. Abbildung 5.4 zeigt diese Situation.

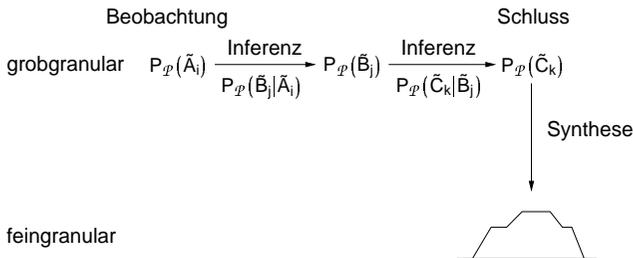


Abbildung 5.4: Schnelle, mehrstufige Inferenz im Standardfall

Wenn  $G_1 = (P_{\mathcal{P}}(\tilde{A}_i|\tilde{B}_j))_{ij}$  und  $G_2 = (P_{\mathcal{P}}(\tilde{B}_j|\tilde{C}_k))_{jk}$  die Gewichte der Regeln beider Stufen bezeichnen, ergibt sich für die zweistufige Inferenz einer Eingabe  $\vec{a}$

$$\vec{c} = G_2 G_1^t \cdot \vec{a} = G^t \cdot \vec{a} \quad \text{mit} \quad G = G_1 G_2 \quad (5.7)$$

Insbesondere können wir die mehrstufige Inferenz also durch eine einstufige Inferenz ersetzen, indem wir die Gewichtsmatrizen durch Matrixmultiplikation zusammenfassen. Auf diese Weise ersparen wir uns nicht nur die Berechnungen auf feingranularer Ebene zwischen den Stufen, sondern können sogar die Anzahl der Stufen auf eins reduzieren. Es ist offensichtlich, dass diese Reduktion allgemein für  $s$ -stufige Regelsysteme gilt. Zusätzlich können wir die Elemente der Matrix  $G$  als bedingte Wahrscheinlichkeiten  $P_{\mathcal{P}}(\tilde{A}_i|\tilde{C}_k)$  deuten. Damit haben wir grobgranulare Regeln, die den direkten Zusammenhang zwischen  $x$  und  $z$  in verständlicher Form beschreiben. Dies ist insbesondere in Anwendungen interessant, bei denen wir den Zusammenhang zwischen  $x$  und  $z$  nicht direkt formulieren können und die Zwischenvariable  $y$  nur mangels besseren Wissens einführen müssen.

**Allgemeiner Fall** Wir gehen nun davon aus, dass die Partitionierungen  $\mathcal{P}_y^1 = \mathcal{P}_y^2$  nicht übereinstimmen. Wie bereits erwähnt, ist dieser Fall in praktischen Anwendungen wesentlich seltener anzufinden als der Standardfall; wir sollten ihn trotzdem handhaben können.

Eine unscharfe Eingabe  $\tilde{A}$  resultiert durch Inferenz in der ersten Stufe in einem Schluss  $\tilde{B}$  nach Gleichung (5.4):

$$pl_{\tilde{B}}(v) = \sum_j P_{\mathcal{P}}(\tilde{B}_j^1) \cdot pl_{\tilde{B}_j^1}(v) \quad (5.8)$$

$\tilde{B}$  muss anschließend in einem Analysevorgang durch die Antezedenzien  $\tilde{B}_k^2$  der zweiten Stufe dargestellt werden. Mit dem Analyseverfahren nach Ansatz 2 bestimmen wir die Begriffswahrscheinlichkeiten  $P_{\mathcal{P}}(\tilde{B}_k^2)$ , mit denen die Kombination der  $\tilde{B}_k^2$  die unscharfe Menge  $\tilde{B}$  approximiert:

$$pl_{\tilde{B}}(v) \approx \sum_k P_{\mathcal{P}}(\tilde{B}_k^2) \cdot pl_{\tilde{B}_k^2}(v) \quad (5.9)$$

Abbildung 5.2 illustriert dieses Vorgehen. Wir werden nachfolgend zeigen, dass wir den rechentechnisch aufwändigen Analysevorgang umgehen können, indem wir die Gewichte  $P_{\mathcal{P}}(\tilde{B}_k^2)$  direkt aus den Wahrscheinlichkeiten  $P_{\mathcal{P}}(\tilde{B}_j^1)$  berechnen. Die Grundidee unseres Verfahrens lautet, mithilfe einer Analyse vorab jeden Begriff  $\tilde{B}_j^1$  durch eine Kombination der  $\tilde{B}_k^2$  zu approximieren:

$$pl_{\tilde{B}_j^1}(v) \approx \sum_k w_{jk} \cdot pl_{\tilde{B}_k^2}(v) \quad (5.10)$$

Für die Bestimmung der Gewichte  $w_{jk}$  führen wir ausschließlich Schritt 2 des Analyseverfahrens gemäß Ansatz 2 durch, also eine reine Funktionsapproximation. Die Gewichte  $w_{jk}$  können damit negativ sein und sind nicht normiert. Der folgende Satz zeigt, wie wir die Wahrscheinlichkeiten  $P_{\mathcal{P}}(\tilde{B}_k^2)$  mithilfe der Gewichte  $w_{jk}$  berechnen können.

**Satz 9** *Seien  $w_{jk}$  die Gewichte aus Gleichung (5.10) und  $P_{\mathcal{P}}(\tilde{B}_k^2)$  die Gewichte, die durch die Analyse von  $\tilde{B}$  nach Ansatz 2 mit den Begriffen  $\tilde{B}_k^2 \in \mathcal{P}_{\mathcal{Y}}^2$  entstehen (Gleichung (5.9)). Sei  $I := \{k \mid \text{supp}(\tilde{B}_k^2) \cap \text{supp}(\tilde{B}) \neq \emptyset\}$ . Wenn die Plausibilitätsdichten der Grundbegriffe  $\tilde{B}_j^1 \in \mathcal{P}_{\mathcal{Y}}^1$  und  $\tilde{B}_k^2 \in \mathcal{P}_{\mathcal{Y}}^2$  stetig sind, gilt für alle  $k$*

$$P_{\mathcal{P}}(\tilde{B}_k^2) = \begin{cases} \frac{\max\{0, \sum_j w_{jk} \cdot P_{\mathcal{P}}(\tilde{B}_j^1)\}}{\sum_{i \in I} \max\{0, \sum_j w_{ji} \cdot P_{\mathcal{P}}(\tilde{B}_j^1)\}}, & \text{falls } k \in I \\ 0 & \text{sonst} \end{cases} \quad (5.11)$$

*Beweis:* Für den Beweis begeben wir uns in den Prä-Hilbertraum  $(C(\mathcal{Y}), \|\cdot\|_2)$  mit innerem Produkt  $\langle f, f \rangle^{1/2} = \|f\|_2$ , vgl. Anhang B.2. Da wir in (5.10) die Bestapproximationen der Begriffe  $\tilde{B}_j^1$  berechnet haben, gilt für alle  $j, l$  die Normalengleichung

$$\sum_k w_{jk} \cdot \langle pl_{\tilde{B}_k^2}, pl_{\tilde{B}_l^1} \rangle = \langle pl_{\tilde{B}_j^1}, pl_{\tilde{B}_l^1} \rangle \quad (5.12)$$

Wie folgende Rechnung zeigt, löst  $\alpha_k := \sum_j w_{jk} \cdot P_{\mathcal{P}}(\tilde{B}_j^1)$  für alle  $l$  die Normalengleichung:

$$\sum_k \sum_j w_{jk} \cdot P_{\mathcal{P}}(\tilde{B}_j^1) \cdot \langle pl_{\tilde{B}_k^2}, pl_{\tilde{B}_l^2} \rangle = \sum_j P_{\mathcal{P}}(\tilde{B}_j^1) \sum_k w_{jk} \cdot \langle pl_{\tilde{B}_k^2}, pl_{\tilde{B}_l^2} \rangle \quad (5.13)$$

$$\stackrel{(5.12)}{=} \sum_j P_{\mathcal{P}}(\tilde{B}_j^1) \cdot \langle pl_{\tilde{B}_j^1}, pl_{\tilde{B}_l^2} \rangle \quad (5.14)$$

$$\stackrel{(5.8)}{=} \langle pl_{\tilde{B}}, pl_{\tilde{B}_l^2} \rangle \quad (5.15)$$

Da wir immer die lineare Unabhängigkeit der Plausibilitätsdichten voraussetzen, ist die Lösung eindeutig. Die  $\alpha_k$  minimieren damit die Fehlerfunktion  $F = \|pl_{\tilde{B}} - \sum_k \alpha_k \cdot pl_{\tilde{B}_k^2}\|_2$ ; wir haben auf diese Weise Schritt 2 des Analyseverfahrens nach Ansatz 2 zur Bestimmung der Wahrscheinlichkeiten  $P_{\mathcal{P}}(\tilde{B}_k^2)$  durchgeführt. Die Schritte 3–6 des Analyseverfahrens werden in Gleichung (5.11) abgearbeitet. Damit gilt die Behauptung.  $\square$

Satz 9 können wir in gleicher Weise für die diskrete Version der Funktionsapproximation formulieren, s. Anhang B.2. Wir können dann auf die Stetigkeit der Plausibilitätsdichten verzichten und mit diskreten Universen arbeiten.

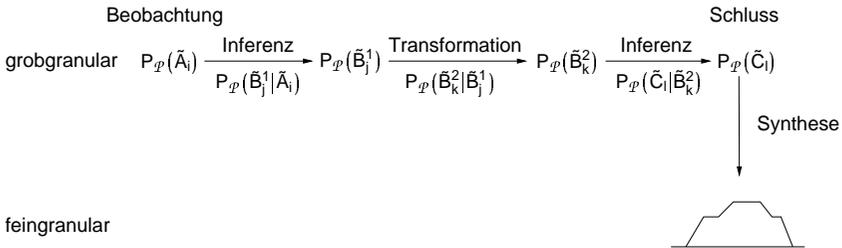


Abbildung 5.5: Schnelle, mehrstufige Inferenz im allgemeinen Fall

Dem Satz folgend können wir die Wahrscheinlichkeiten  $P_{\mathcal{P}}(\tilde{B}_k^2)$  ohne einen Analysevorgang aus den  $P_{\mathcal{P}}(\tilde{B}_j^1)$  bestimmen. Wir bezeichnen das Verfahren als **Transformation**, weil es die Darstellung von  $\tilde{B}$  durch die Grundbegriffe  $\tilde{B}_j^1$  in eine Darstellung mit den Grundbegriffen  $\tilde{B}_k^2$  überführt. Voraussetzung dafür ist eine Vorberechnung der  $w_{jk}$ . In Abbildung 5.5 sehen wir im Vergleich mit der naiven Inferenz von Abb. 5.2, dass wir den teuren Schritt auf feingranularer Ebene sparen.

Wenn wir auf die Gewichte  $w_{jk}$  die Schritte 1 und 3–5 des Analyseverfahrens anwenden, dann können wir sie als bedingte Wahrscheinlichkeiten  $P_{\mathcal{P}}(\tilde{B}_k^2 | \tilde{B}_j^1)$  deuten.

Formel (5.11) vereinfacht sich in diesem Fall zur Formel von der totalen Wahrscheinlichkeit:

$$P_{\mathcal{P}}(\tilde{B}_k^2) = \sum_j P_{\mathcal{P}}(\tilde{B}_k^2 | \tilde{B}_j^1) \cdot P_{\mathcal{P}}(\tilde{B}_j^1) \quad (5.16)$$

Wir können die Transformation damit als weitere Inferenzstufe interpretieren, die zwischen der ersten und zweiten Stufe liegt. Die Gewichte  $w_{jk}$  entsprechen als bedingte Wahrscheinlichkeiten  $P_{\mathcal{P}}(\tilde{B}_k^2 | \tilde{B}_j^1)$  gewichteten Regeln  $\tilde{B}_j^1 \xrightarrow{P_{\mathcal{P}}(\tilde{B}_k^2 | \tilde{B}_j^1)} \tilde{B}_k^2$ . Wenn alle Gewichte  $w_{jk}$  nichtnegativ sind, dann müssen wir sie nur normieren, um sie als  $P_{\mathcal{P}}(\tilde{B}_k^2 | \tilde{B}_j^1)$  zu verstehen. Schritt 3 und der erste Punkt von Schritt 4 des Analyseverfahrens bleiben wirkungslos. In diesem Fall ist die Berechnung der Wahrscheinlichkeiten  $P_{\mathcal{P}}(\tilde{B}_k^2)$  mit (5.16) exakt. Falls aber einige Gewichte  $w_{jk}$  negativ sind, dann führen die Schritte 3 und 4 des Analyseverfahrens dazu, dass wir die gesuchten Wahrscheinlichkeiten  $P_{\mathcal{P}}(\tilde{B}_k^2)$  nicht mehr exakt mit (5.16) berechnen können. Man muss hier also abwägen, ob man negative Gewichte zulässt und die Transformation nur als ein Hilfsmittel betrachtet, um exakt und schnell rechnen zu können, oder ob man ein Inferenzmodell haben möchte, das vollständig auf der Wahrscheinlichkeitsrechnung beruht, dafür aber i. A. nicht exakt ist.

Gleichung (5.16) können wir durch eine Matrixmultiplikation darstellen:

$$\vec{b}_2 = W^t \cdot \vec{b}_1 \quad (5.17)$$

mit  $\vec{b}_1 = (P_{\mathcal{P}}(\tilde{B}_j^1))_j$ ,  $\vec{b}_2 = (P_{\mathcal{P}}(\tilde{B}_k^2))_k$  und  $W = (P_{\mathcal{P}}(\tilde{B}_k^2 | \tilde{B}_j^1))_{jk}$ . Wir bezeichnen die Matrix  $W$  als **Transformationsmatrix**. Insgesamt gilt dann für eine zweistufige Inferenz mit Transformation

$$\vec{c} = G_2^t \cdot W^t \cdot G_1^t \cdot \vec{a} = G^t \cdot \vec{a} \quad \text{mit} \quad G = G_1 \cdot W \cdot G_2, \quad (5.18)$$

wobei  $\vec{c} = (P_{\mathcal{P}}(\tilde{C}_l))_l$ . Wie im Standardfall, bei dem die Konsequenzen der ersten Regelstufe gleichzeitig die Antezedenzen der zweiten Stufe sind, kann die Inferenz also auch im allgemeinen Fall immer exakt auf grobgranularer Ebene berechnet werden.

Darüber hinaus können wir mehrere Inferenzstufen zu einer Stufe zusammenfassen, wenn wir alle Gewichte als Wahrscheinlichkeiten interpretieren. Auf diese Weise wird nicht nur die Berechnung beschleunigt, sondern wir gewinnen zusätzlich den direkten Zusammenhang zwischen den Eingangsgrößen der ersten Regelstufe und den Ausgangsgrößen der letzten Stufe. Dargestellt wird dieser Zusammenhang durch gewichtete Regeln auf Basis der Grundbegriffe.

**Bemerkung 3** *Beim Regler von MAMDANI ist es ebenfalls möglich, mehrere Regelstufen zu einer zusammenzufassen [LTT98]. Das Verfahren beschränkt sich allerdings auf den Standardfall von Mehrstufigkeit in Abb. 5.3, und die resultierenden Regeln sind schwer verständlich, weil die Grundbegriffe in der Darstellung verloren gehen.*

### 5.2.3 Regelnetzwerke

Bisher haben wir nur eindimensionale Antezedenzen betrachtet. In diesem Abschnitt werden wir zeigen, wie man mit mehrdimensionalen Antezedenzen umgeht, deren Elemente konjunktiv verknüpft sind. In Verbindung mit mehrstufigem Schließen können wir so Netzwerke von Regeln aufbauen, die Baumstruktur haben. Am Ende dieses Abschnitts beschreiben wir, wie Inferenzen in solchen Strukturen berechnet werden.

Eine Regel mit mehrdimensionalem Antezedens hat die Form „Wenn  $x = \tilde{A}$  und  $y = \tilde{B}$  und ... , dann  $z = \tilde{C}$ “. Wir modellieren solche Regeln über mehrdimensionale Begriffswahrscheinlichkeiten. Auf diese Weise wird das eindimensionale Verfahren direkt auf mehrdimensionale Probleme übertragen. Der einfacheren Darstellung wegen beschreiben wir nur Regeln mit zweidimensionalen Antezedenzen; Regeln mit vieldimensionalen Antezedenzen werden analog behandelt. Wir betrachten damit Regeln der Form  $(\tilde{A}_i, \tilde{B}_j) \xrightarrow{\gamma_{ijk}} \tilde{C}_k$ , die wir als bedingte Wahrscheinlichkeiten  $\gamma_{ijk} = P_{\mathcal{P}}(\tilde{C}_k | \tilde{A}_i, \tilde{B}_j)$  interpretieren. Die Inferenz einer Beobachtung von  $(x, y)$  wird wieder mit der Formel von der totalen Wahrscheinlichkeit berechnet:

$$P_{\mathcal{P}}(\tilde{C}_k) = \sum_{i,j} P_{\mathcal{P}}(\tilde{C}_k | \tilde{A}_i, \tilde{B}_j) \cdot P_{\mathcal{P}}(\tilde{A}_i, \tilde{B}_j) \quad (5.19)$$

Um Formel (5.19) anwenden zu können, müssen wir die Begriffswahrscheinlichkeiten  $P_{\mathcal{P}}(\tilde{A}_i, \tilde{B}_j)$  unter der Kenntnis der Beobachtung bestimmen. Wenn die Variablen  $x$  und  $y$  unabhängig sind, dann können wir die gemeinsamen Wahrscheinlichkeiten durch  $P_{\mathcal{P}}(\tilde{A}_i, \tilde{B}_j) = P_{\mathcal{P}}(\tilde{A}_i) \cdot P_{\mathcal{P}}(\tilde{B}_j)$  berechnen. In diesem Fall beobachten wir für  $x$  und  $y$  unabhängig voneinander unscharfe Werte  $\tilde{A}$  und  $\tilde{B}$ .  $P_{\mathcal{P}}(\tilde{A}_i)$  und  $P_{\mathcal{P}}(\tilde{B}_j)$  finden wir dann in gleicher Weise wie bei eindimensionalen Antezedenzen durch Synthese oder Analyse der Beobachtungen  $\tilde{A}$ ,  $\tilde{B}$ .

Wenn  $x$  und  $y$  abhängig sind, müssen wir die gemeinsamen Wahrscheinlichkeiten  $P_{\mathcal{P}}(\tilde{A}_i, \tilde{B}_j)$  direkt bestimmen. Falls die Beobachtung als Synthese vorliegt, kennen wir die gemeinsamen Wahrscheinlichkeiten bereits. Anderenfalls ist die Beobachtung eine zweidimensionale Plausibilitätsdichte  $p_{\mathcal{L}_{x,y}}(u, v)$  auf  $\mathcal{X} \times \mathcal{Y}$ , die wir analysieren müssen. Wegen der Abhängigkeit von  $x$  und  $y$  sind auch die Grundbegriffe

zweidimensional. Sie sind Elemente einer Partition des Raums  $\mathcal{X} \times \mathcal{Y}$  und werden durch zweidimensionale Plausibilitätsdichten  $pl_{\bar{A}_i, \bar{B}_j}(u, v)$  dargestellt. Für die Analyse der beobachteten Plausibilitätsdichte  $pl_{x,y}$  benutzen wir das Analyseverfahren aus Abschnitt 4.2.3. Es ist unabhängig von der Dimension der Plausibilitätsdichte. Insgesamt werden Regeln mit mehrdimensionalen Antezedenzen damit analog zu Regeln mit eindimensionalen Antezedenzen dargestellt und entsprechend auch im Inferenzverfahren verarbeitet.

Mehrstufige Regelbasen mit mehrdimensionalen Antezedenzen können wir als Baum darstellen, wie Abbildung 5.6 illustriert. Abhängige Variablen in den Antezedenzen fassen wir dabei zu einem Vektor zusammen und behandeln sie formal wie eine einfache Variable. Die Blätter des Baums repräsentieren jene Variablen,

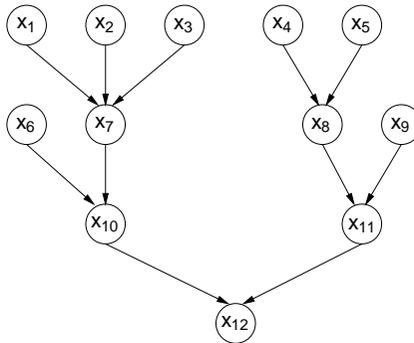


Abbildung 5.6: Regelnetzwerk in Baumstruktur mit Variablen  $x_1, x_2, \dots, x_{12}$

die beobachtet werden und als Eingangswerte des unscharfen Systems dienen. Der unscharfe Schluss des Systems wird in den Variablen der Wurzel gebildet. Um den Schluss zu berechnen, arbeiten wir den Baum schichtweise von den Blättern zur Wurzel<sup>3</sup> durch lokale Inferenzen ab. Das Beispiel in Anhang A verdeutlicht dieses Vorgehen noch einmal.

Wie bereits in Abschnitt 5.1 erwähnt, ersetzen wir Regeln mit mehrdimensionalem Konsequens durch mehrere Regeln mit eindimensionalem Konsequens. Global betrachtet substituieren wir einen Mehrfachbaum durch mehrere einfache Bäume, in denen wir getrennt die Inferenzen berechnen können. Liegt dagegen ein gerichteter azyklischer Graph vor, so müssen wir den globalen Inferenzmechanismus anpassen, um der Wahrscheinlichkeitstheorie zu genügen. Da solche Strukturen in regelbasierten Systemen selten vorkommen und die entsprechenden Inferenzverfahren

<sup>3</sup>engl. *bottom up*

aus der Theorie BAYESSCHER Netze bekannt sind [Pea88, RN95], diskutieren wir diesen Fall hier nicht weiter.

### 5.3 Allgemeine Regeln

In Abschnitt 5.1 haben wir gesehen, wie man Regeln  $\tilde{A} \longrightarrow \tilde{B}$  als bedingte Wahrscheinlichkeiten modelliert. Wir sind dabei davon ausgegangen, dass  $\tilde{A}$  und  $\tilde{B}$  Grundbegriffe sind. Diese Einschränkung heben wir in diesem Abschnitt auf.

In Anwendungen basieren Regeln normalerweise auf den Grundbegriffen, da sie die vergrößerte Sicht auf die Universen darstellen, die unserem Verständnis entspricht. Wenn allerdings ein Zusammenhang zwischen Universen nicht durch Experten festgelegt wird, sondern in Form von Daten gegeben ist, gibt es zwei Möglichkeiten für die Verarbeitung. Wir können die Daten benutzen, um die unscharfen Partitionen selbst *und* die Regeln zwischen den Partitionen festzulegen. In diesem Fall basieren die Regeln wie in Abschnitt 5.1 auf den Grundbegriffen.

Die zweite Möglichkeit besteht darin, den Zusammenhang mit bereits definierten Grundbegriffen zu beschreiben. Aus den Daten resultiert eine Regel mit unscharfen Begriffen  $\tilde{A}$  und  $\tilde{B}$ , die keine Elemente der Partitionen sind. Diese Regel ersetzen wir durch Regeln mit Grundbegriffen. Dieser Fall wird in der Praxis nicht so häufig vorkommen, denn er bedeutet, dass die vorab definierten Grundbegriffe unabhängig von den Daten über den Zusammenhang festgelegt werden. Es ist fraglich, ob das Problem dann angemessen modelliert ist. Aus zwei Gründen werden wir uns in diesem Abschnitt trotzdem dieser Aufgabe annehmen. Erstens werden beliebige Regeln verständlich, wenn wir sie mit den Grundbegriffen darstellen, und zweitens können wir sie dann effizient mit unserem Inferenzverfahren verarbeiten.

Wir beschreiben in diesem Abschnitt, wie man die bedingten Wahrscheinlichkeiten  $P_{\mathcal{P}}(\tilde{B}_j|\tilde{A}_i)$  so bestimmt, dass die resultierenden Regeln  $\tilde{A}_i \xrightarrow{P_{\mathcal{P}}(\tilde{B}_j|\tilde{A}_i)} \tilde{B}_j$  einer oder mehreren Regeln  $\tilde{A} \longrightarrow \tilde{B}$  für beliebige unscharfe Mengen  $\tilde{A}$  und  $\tilde{B}$  entsprechen. Die Grundidee des Verfahrens lautet, ein Antezedens  $\tilde{A}$  mit den Grundbegriffen  $\tilde{A}_i$  und ein Konsequens  $\tilde{B}$  mit den Grundbegriffen  $\tilde{B}_j$  durch gewichtete Kombinationen darzustellen. Aus den Begriffswahrscheinlichkeiten werden wir dann die bedingten Wahrscheinlichkeiten für die Regeln berechnen.

Wir wenden uns zunächst dem einfacheren Fall zu, bei dem das Antezedens ein Grundbegriff  $\tilde{A}_i$  und das Konsequens  $\tilde{B}$  eine beliebige unscharfe Menge auf  $\mathcal{Y}$  ist:  $\tilde{A}_i \longrightarrow \tilde{B}$ . Gesucht sind zunächst die Begriffswahrscheinlichkeiten  $P_{\mathcal{P}}(\tilde{B}_j)$ . Das Konsequens  $\tilde{B}$  ist entweder direkt als unscharfe Menge gegeben oder wir haben es mit den Grundbegriffen synthetisiert. Im ersten Fall können wir mit einer Analyse die Begriffswahrscheinlichkeiten bestimmen, im zweiten Fall sind sie direkt

gegeben. Wir haben somit

$$pl_{\tilde{B}}(v) \approx \sum_j P_{\mathcal{P}}(\tilde{B}_j) \cdot pl_{\tilde{B}_j}(v) \quad (5.20)$$

Im zweiten Fall liegt sogar Gleichheit vor. Um die bedingten Wahrscheinlichkeiten  $P_{\mathcal{P}}(\tilde{B}_j|\tilde{A}_i)$  zu bestimmen, betrachten wir den Modus Ponens. Wir suchen gewichtete Regeln  $\tilde{A}_i \xrightarrow{P_{\mathcal{P}}(\tilde{B}_j|\tilde{A}_i)} \tilde{B}_j$ , die mit unserem Inferenzverfahren den Grundbegriff  $\tilde{A}_i$  auf das Konsequens  $\tilde{B}$  abbilden. Folgender Satz zeigt, dass wir aus dieser Anforderung eindeutig auf die bedingten Wahrscheinlichkeiten schließen können, falls  $\tilde{B}$  eine Kombination der Grundbegriffe  $\tilde{B}_j$  ist.

**Satz 10** *Gegeben seien Partitionen  $\mathcal{P}_X, \mathcal{P}_Y$  der Universen  $X, Y$  und eine Regel  $\tilde{A} \rightarrow \tilde{B}$  mit  $\tilde{A} \in \mathcal{P}_X$  und  $\tilde{B} \in \mathcal{UM}(Y)$ , wobei  $\tilde{B}$  eine durch die Begriffswahrscheinlichkeiten  $P_{\mathcal{P}}(\tilde{B}_j)$  gewichtete Kombination der Grundbegriffe  $\tilde{B}_j \in \mathcal{P}_Y$  ist. Das durch die Gleichungen (5.3) und (5.4) beschriebene Inferenzverfahren erfüllt genau dann den Modus Ponens, d. h. es bildet  $\tilde{A}$  auf  $\tilde{B}$  ab, wenn für die bedingten Wahrscheinlichkeiten gilt:*

$$\forall j: P_{\mathcal{P}}(\tilde{B}_j|\tilde{A}) = P_{\mathcal{P}}(\tilde{B}_j) \quad (5.21)$$

*Beweis:* Dass der Modus Ponens aus Gleichung (5.21) folgt, haben wir schon mit Satz 8 gezeigt. Nehmen wir nun an, dass der Modus Ponens gilt. Dann gibt es bedingte Wahrscheinlichkeiten  $P_{\mathcal{P}}(\tilde{B}_j|\tilde{A})$  mit

$$\sum_j P_{\mathcal{P}}(\tilde{B}_j|\tilde{A}) \cdot pl_{\tilde{B}_j} \equiv pl_{\tilde{B}} \equiv \sum_j P_{\mathcal{P}}(\tilde{B}_j) \cdot pl_{\tilde{B}_j}$$

Die linke Seite der Gleichung erhält man aus den Gleichungen (5.3) und (5.4), wenn man  $P_{\mathcal{P}}(\tilde{A}) = 1$  und  $P_{\mathcal{P}}(\tilde{A}') = 0$  für alle  $\tilde{A} \neq \tilde{A}' \in \mathcal{P}_X$  einsetzt, d. h.  $x = \tilde{A}$  annimmt. Die rechte Seite der Gleichung ist Voraussetzung des Satzes. Aus der linearen Unabhängigkeit der Plausibilitätsdichten  $pl_{\tilde{B}_j}$  folgt die Identität (5.21).  $\square$

Falls wir das Konsequens  $\tilde{B}$  nur näherungsweise mit den Grundbegriffen  $\tilde{B}_j$  darstellen können, dann können wir den Modus Ponens mit Regeln auf der Basis von Grundbegriffen ebenfalls nur näherungsweise erfüllen. Mehr als das Ergebnis von Satz 10 können wir nicht erwarten, denn die Vergrößerung der Universen durch Partitionierungen kann nicht beliebige unscharfe Aussagen präzise wiedergeben. Der Satz gilt nicht nur für eine einzelne Regel  $\tilde{A} \rightarrow \tilde{B}$ , sondern für einen ganzen Satz von Regeln  $\tilde{A}_i \rightarrow \tilde{B}'_i$  für alle  $\tilde{A}_i \in \mathcal{P}_X$  mit unscharfen Mengen  $\tilde{B}'_i \in \mathcal{UM}(Y)$ , die als Kombination der Grundbegriffe gebildet werden können.

Wir lösen uns nun von der Annahme, dass die Antezedenzien der Regeln Grundbegriffe sind. Wir betrachten  $r$  Regeln der Art  $\tilde{A}'_k \longrightarrow \tilde{B}'_k$  mit beliebigen unscharfen Mengen  $\tilde{A}'_k \in \mathcal{UM}(\mathcal{X})$  und  $\tilde{B}'_k \in \mathcal{UM}(\mathcal{Y})$ . Wir werden sehen, dass in diesem allgemeinen Fall die Wahl der bedingten Wahrscheinlichkeiten  $P_{\mathcal{P}}(\tilde{B}_j|\tilde{A}_i)$  nicht eindeutig ist. Dies liegt daran, dass es Konflikte zwischen verschiedenen Regeln geben kann, die im einfachen Fall von oben nicht auftreten.

Ausgangspunkt des Verfahrens ist ein Analysevorgang nach Abschnitt 4.2.3, mit dem wir die Wahrscheinlichkeiten  $P_{\mathcal{P}}(\tilde{A}_i|\tilde{A}'_k)$  und  $P_{\mathcal{P}}(\tilde{B}_j|\tilde{B}'_k)$  bestimmen. Die Regelgewichte  $P_{\mathcal{P}}(\tilde{B}_j|\tilde{A}_i)$  sollen wieder so gewählt werden, dass das Inferenzverfahren aus Abschnitt 5.2.1 eine Eingabe  $\tilde{A}'_k$  auf das gewünschte Konsequens  $\tilde{B}'_k$  abbildet, d. h. für jede Regel der Modus Ponens erfüllt wird:

$$\forall j, k: \sum_i P_{\mathcal{P}}(\tilde{B}_j|\tilde{A}_i) \cdot P_{\mathcal{P}}(\tilde{A}_i|\tilde{A}'_k) = P_{\mathcal{P}}(\tilde{B}_j|\tilde{B}'_k) \quad (5.22)$$

Für jedes  $j$  erhalten wir ein lineares Gleichungssystem

$$\begin{pmatrix} (P_{\mathcal{P}}(\tilde{A}_i|\tilde{A}'_1))_i \\ (P_{\mathcal{P}}(\tilde{A}_i|\tilde{A}'_2))_i \\ \vdots \\ (P_{\mathcal{P}}(\tilde{A}_i|\tilde{A}'_r))_i \end{pmatrix} \begin{pmatrix} \gamma_{1j} \\ \gamma_{2j} \\ \vdots \\ \gamma_{nj} \end{pmatrix} = \begin{pmatrix} P_{\mathcal{P}}(\tilde{B}_j|\tilde{B}'_1) \\ P_{\mathcal{P}}(\tilde{B}_j|\tilde{B}'_2) \\ \vdots \\ P_{\mathcal{P}}(\tilde{B}_j|\tilde{B}'_r) \end{pmatrix}, \quad (5.23)$$

wobei wir statt der Wahrscheinlichkeiten  $P_{\mathcal{P}}(\tilde{B}_j|\tilde{A}_i)$  zunächst  $\gamma_{ij}$  schreiben, da die Lösungen des Gleichungssystems i. A. kein Wahrscheinlichkeitsmaß definieren. Wir nehmen an, dass die Anzahl  $n$  der Grundbegriffe auf  $\mathcal{X}$  größer als die Anzahl  $r$  der allgemeinen Regeln ist. Ansonsten ist die Granularisierungsstufe der Universen den Regeln nicht angemessen, denn die Regeln unterscheiden offenbar mehr Fälle, als wir mit den Grundbegriffen ausdrücken können. Wenn das Gleichungssystem lösbar ist, dann ist es wegen  $r < n$  unterbestimmt. Es gibt somit keine eindeutige Lösung.

Das folgende Verfahren wählt eine Lösung aus und sorgt gleichzeitig dafür, dass die Regelgewichte  $\gamma_{ij}$  nicht negativ sind. Wir betrachten dazu Gleichung (5.22) für ein festes  $k$ , also für die  $k$ -te Regel. Sie besitzt i. A. keine eindeutige Lösung für die  $\gamma_{ij}$ . Da wir uns im Kontext von Wahrscheinlichkeiten befinden, müssen wir ohne zusätzliches Wissen eine Gleichverteilung annehmen, d. h.  $\gamma_{ij}$  ist konstant bzgl.  $i$ . Mit  $\gamma_j = \gamma_{ij}$  gilt dann nach Gleichung 5.22:

$$P_{\mathcal{P}}(\tilde{B}_j|\tilde{B}'_k) = \sum_i \gamma_j \cdot P_{\mathcal{P}}(\tilde{A}_i|\tilde{A}'_k) \quad (5.24)$$

$$= \gamma_j \sum_i P_{\mathcal{P}}(\tilde{A}_i|\tilde{A}'_k) \quad (5.25)$$

$$\stackrel{(5.1)}{=} \gamma_j \quad (5.26)$$

Wir definieren also für alle  $i, j$ :

$$\gamma_{ij}^k := \begin{cases} P_{\mathcal{P}}(\tilde{B}_j | \tilde{B}'_k), & \text{wenn } P_{\mathcal{P}}(\tilde{A}_i | \tilde{A}'_k) > 0 \\ 0, & \text{wenn } P_{\mathcal{P}}(\tilde{A}_i | \tilde{A}'_k) = 0 \end{cases} \quad (5.27)$$

Im Fall  $P_{\mathcal{P}}(\tilde{A}_i | \tilde{A}'_k) = 0$  trägt der Term  $\gamma_{ij}^k$  nicht zur Summe in (5.24) bei. Da keine Evidenz für  $\gamma_{ij}^k > 0$  vorliegt, setzen wir den Term auf null. Der Index  $k$  des Gewichtes  $\gamma_{ij}^k$  markiert, dass diese Definition nur für die  $k$ -te Regel gilt. Dies ist wichtig, weil unterschiedliche Regeln die Gewichte  $\gamma_{ij}$  u. U. verschieden definieren. Es handelt sich bei diesem Phänomen um eine Form von *Inkonsistenz* zwischen den Regeln, die durch die Vergrößerung des Wissens bei der Granularisierung entsteht. Wie bei Wahrscheinlichkeiten üblich, bilden wir die endgültigen bedingten Wahrscheinlichkeiten durch eine Durchschnittsbildung über alle Regeln. Damit tragen wir dem Unwissen Rechnung, welches Gewicht richtig ist. Hinzu kommt die Normierung auf  $\sum_j \gamma_{ij} = 1$ . Wir definieren also

$$P_{\mathcal{P}}(\tilde{B}_j | \tilde{A}_i) := \frac{\sum_k \gamma_{ij}^k}{\sum_l \sum_k \gamma_{il}^k} \quad (5.28)$$

Jedes  $\gamma_{ij}^k$  wird auf diese Weise als gleichberechtigte Information über  $P_{\mathcal{P}}(\tilde{B}_j | \tilde{A}_i)$  aufgefasst. Diese Berechnungsvorschrift ist konsistent mit dem Vorgehen am Anfang dieses Abschnitts, in dem wir angenommen haben, dass das Antezedens der Regel ein Grundbegriff ist. Wenn wir dies hier ebenfalls annehmen, d. h.  $\tilde{A}'_k \in \mathcal{P}_{\mathcal{X}}$ , dann sind die Vorschriften (5.28) und (5.21) zur Berechnung der bedingten Wahrscheinlichkeiten identisch. Die Gültigkeit des Modus Ponens können wir i. A. nicht sicherstellen. Das liegt daran, dass das Antezedens und das Konsequens einer beliebigen Regel nicht exakt mit den Grundbegriffen dargestellt werden können. Dies zeigt auf technischer Ebene, dass die Universen abhängig von den Regeln partitioniert werden sollten.

Zusammenfassend stellen wir fest, dass eine Regel  $\tilde{A} \rightarrow \tilde{B}$  in unserem Ansatz dargestellt werden kann, auch wenn  $\tilde{A}$  und  $\tilde{B}$  nicht zu den Grundbegriffen gehören. Wir haben bereits mehrfach darauf hingewiesen, dass die Formulierung solcher Regeln in der Praxis nicht der Normalfall ist; trotzdem sollten wir damit konsistent zum üblichen Verfahren umgehen können, was wir mit diesem Abschnitt gezeigt haben.

## 5.4 Zuverlässigkeit von Schlussfolgerungen

Wenn wir einen unscharfen Schluss ziehen, dann sind wir nicht nur an dem Schluss selbst, sondern auch an seiner *Zuverlässigkeit* interessiert. Wir suchen einen An-

haltspunkt dafür, wieviel Vertrauen wir in den Schluss setzen können. Dazu unterscheiden wir zwei Typen von Schlüssen. Bei einigen Anwendungen reicht die Entscheidung für einen Grundbegriff als Schluss. Wir sind in diesem Fall mit einem grobgranularen Ergebnis zufrieden. Bei anderen Problemstellungen hingegen werden scharfe Schlüsse erwartet. Ein technisches System etwa benötigt einen scharfen Wert wie eine Spannung, mit der es beispielsweise gesteuert wird. In diesem Fall liegt die Entscheidung auf feingranularer Ebene vor. In Anhang A werden beide Möglichkeiten, grob- und feingranulare Schlüsse, an einem Beispiel illustriert.

**Grobgranulare Schlüsse** Eine Schlussfolgerung  $\tilde{B}^l$  liegt als gewichtete Kombination von Grundbegriffen vor. Wenn wir einen Grundbegriff  $\tilde{B} \in \mathcal{P}_{\mathcal{Y}}$  aus der Kombination als Schluss wählen, so nennen wir  $\tilde{B}$  einen **grobgranularen Schluss**. Im Allgemeinen wird man sich für den Grundbegriff mit der größten Begriffswahrscheinlichkeit entscheiden. Dabei misst die Begriffswahrscheinlichkeit  $P_{\mathcal{P}}(\tilde{B})$  die Zuverlässigkeit des grobgranularen Schlusses  $\tilde{B}$ :

$$\forall \tilde{B} \in \mathcal{P}_{\mathcal{Y}} : Z_g(\tilde{B}) := P_{\mathcal{P}}(\tilde{B}) \quad (5.29)$$

Im Beispiel in Anhang A entspricht  $Z_g(\tilde{B})$  der relativen Anzahl von Experten, die  $\tilde{B}$  als Schluss unterstützen.

Je größer die Unsicherheit des Wissens ist, desto größer ist die Anzahl der Entscheidungsalternativen, die durch die Anzahl der positiven Wahrscheinlichkeiten  $P_{\mathcal{P}}(\tilde{B})$  bestimmt wird. Mit der Anzahl der Entscheidungsmöglichkeiten sinkt aber automatisch die Zuverlässigkeit jeder einzelnen Alternative. Das liegt daran, dass wir uns in einem wahrscheinlichkeitstheoretischen Modell bewegen, in dem die Summe der Zuverlässigkeiten über die Alternativen eins ergibt. Folglich misst  $Z_g(\tilde{B}^c) = 1 - Z_g(\tilde{B})$  die Zuverlässigkeit der Aussage, dass  $\tilde{B}$  nicht der richtige Schluss ist.

**Feingranulare Schlüsse** Die Plausibilitätsdichte  $pl_{\tilde{B}^l}$  eines Schlusses  $\tilde{B}^l$  bezeichnen wir als **feingranularen Schluss**.  $pl_{\tilde{B}^l}(v)$  misst die Zuverlässigkeit der Kontexte, die  $v$  als scharfen Schluss für richtig halten. Im Beispiel in Anhang A sind die Kontexte Experten, die i. A. gleiches Vertrauen genießen.  $pl_{\tilde{B}^l}(v)$  ist dann die relative Anzahl von Experten, die  $v$  als scharfen Schluss vorschlagen.

Die Zuverlässigkeit des feingranularen Schlusses, der durch  $pl_{\tilde{B}^l}(v)$  repräsentiert wird, können wir durch die maximale Plausibilität messen:

$$\forall \tilde{B}^l \in \mathcal{UM}(\mathcal{Y}) : Z_f(\tilde{B}^l) := \sup_{v \in \mathcal{Y}} pl_{\tilde{B}^l}(v) \quad (5.30)$$

$Z_f$  gibt die Plausibilität der scharfen Schlüsse an, in die wir unser größtes Vertrauen legen. In dieser Weise ist  $Z_f$  ein Maß für die Zuverlässigkeit von  $\tilde{B}^l$ , denn in

vielen Anwendungen wählt man eines der  $v \in \mathcal{V}$  mit maximaler Plausibilität als scharfen Schluss aus. Wir kennen dieses Vorgehen bereits aus Abschnitt 2.2 als Maximummethode. Ist  $Z_f$  sehr klein, so ist das  $v$  mit maximaler Plausibilität zwar der beste Schluss, den das System bei gegebenem Wissen vorschlagen kann; unser Vertrauen in diesen Schluss ist aber nicht sehr groß. Dieser Fall tritt ein, wenn das Wissen, auf dem der Schluss basiert, sehr unsicher oder vage ist.

Insgesamt stellen wir fest, dass wir die Zuverlässigkeit oder das Vertrauen in die Schlüsse unseres unscharfen Systems messen können, unabhängig davon, ob sie grob- oder feingranular sind. Wenn wir z. B. annehmen, dass alle für den Schluss verwendeten Informationen von Experten stammen, dann sagt der Zuverlässigkeitswert aus, wieviel Prozent der Experten dem jeweiligen Schluss zustimmen würden. Damit wird dem Anwender ein aussagekräftiger Wert an die Hand gegeben, mit dem er seine Entscheidungen, die auf dem Schluss des unscharfen Systems basieren, leicht rechtfertigen kann.

## 5.5 Rechenaufwand

Insbesondere bei Echtzeitanwendungen spielt die Effizienz des Inferenzverfahrens eine wichtige Rolle. Zu große Berechnungszeiten resultieren darin, dass die Anzahl der Schlüsse pro Zeiteinheit reduziert wird. Das unscharfe System kann also nicht so oft in den umgebenden Prozess eingreifen, was die Leistungsfähigkeit drastisch reduzieren kann oder sogar die Anwendung des unscharfen Systems ausschließt.

Um den Rechenaufwand unseres Inferenzverfahrens zu beurteilen, vergleichen wir es mit dem Prinzip von Kombination/Projektion (PKP) sowie dem Verfahren nach MAMDANI als effizientem, sehr oft verwendeten Spezialfall des PKPs und unscharfer Produktionsregeln. Dazu nehmen wir eine  $m$ -stufige Regelbasis mit  $k$  Grundbegriffen in jeder Schicht an. Antezedenzien und Konsequenzen seien eindimensional. Weiterhin nehmen wir eine Diskretisierung der Universen von jeweils  $n$  Punkten an.

|         | variable Regeln          | feste Regeln             |          |
|---------|--------------------------|--------------------------|----------|
|         |                          | allgemein                | Standard |
| PKP     | $O(k \cdot n^3 \cdot m)$ | $O(n^2 \cdot m)$         | –        |
| MAMDANI | $O(k \cdot n^2 \cdot m)$ | $O(k^2 \cdot m)$         | $O(k^2)$ |
| SPOTT   | $O(k^2 \cdot m)$         | $O(k^2), O(k^2 \cdot m)$ | $O(k)$   |

Tabelle 5.2: Vergleich des Rechenaufwands verschiedener Verfahren

Tabelle 5.2 zeigt den Rechenaufwand der drei Verfahren unter verschiedenen Vor-

aussetzungen. „Variable Regeln“ bezeichnet den Fall, dass die Inferenz komplett berechnet werden muss. Dieser Fall tritt z. B. während der Entwicklung eines unscharfen Systems ein, bei der die Regeln jederzeit geändert werden können. In unserem Fall heißt das, dass die bedingten Wahrscheinlichkeiten variabel sind. Sind die Regeln jedoch fixiert („feste Regeln“), so dass nur zu gegebener Beobachtung der Schluss berechnet werden muss, dann können einige Schritte vorberechnet werden. Im Wesentlichen sind dies die Zusammenfassung mehrerer Regelstufen und Berechnungen, die den Übergang von einer zur nächsten Regelstufe beschleunigen. Der „Standard“-Fall unterscheidet sich noch einmal vom „allgemeinen“ dadurch, dass hier erstens die Konsequenzen der Regeln mit denen der Antezedenzen in der nächsten Stufe übereinstimmen (s. Abschnitt 5.2.2); zweitens haben alle Regeln das Gewicht eins. Letzteres bedeutet, dass aus jedem Grundbegriff genau ein anderer Grundbegriff folgt. Dies ist der Standardfall in mehrstufigen, unscharfen Regelbasen. Der Berechnungsaufwand des PKP und von MAMDANI ist [MC89, Ari96, CP96] entnommen.

Bei allen Verfahren haben wir in Tab. 5.2 den Aufwand weggelassen, den wir benötigen, um eine unscharfe Beobachtung für die Inferenz vorzubereiten. Auf diese Weise können wir den Aufwand der reinen Inferenz besser vergleichen. Bei MAMDANI und dem PKP müssen wir die Übereinstimmung der Beobachtung mit den Antezedenzen der Regeln in der ersten Stufe berechnen. Der Aufwand beträgt  $O(k \cdot n)$ . Bei unserem Verfahren müssen wir die Beobachtung mit den Grundbegriffen der Antezedenzen beschreiben, d. h. die Begriffswahrscheinlichkeiten berechnen. Liegt die Beobachtung als Synthese vor, so kennen wir die Begriffswahrscheinlichkeiten schon, wir haben also keinerlei Aufwand. Im Fall einer unscharfen Menge müssen wir eine Analyse durchführen, was das Lösen eines linearen Gleichungssystems erfordert (s. Anhang B.2). Im schlechtesten Fall beträgt der Aufwand der Analyse  $O(n^3)$ . Für gewöhnlich überlappen sich aber an jeder Stelle  $u \in X$  nur zwei oder drei Plausibilitätsdichten einer Partition. Damit ist die Matrix des LGS dünn besetzt (drei oder vier Einträge pro Zeile, die ungleich null sind). Solche Gleichungssysteme können rekursiv mit einem Aufwand von  $O(n)$  gelöst werden.

In der Praxis wird meist mit scharfen Beobachtungen gearbeitet, die bei allen Verfahren mit einem Aufwand von  $O(k)$  eingebracht werden, was im Vergleich zur Inferenz vernachlässigbar ist, vgl. dazu auch Abschnitt 6.2.

In Tab. 5.2 wird deutlich, dass unser Inferenzverfahren nur von der Anzahl der Grundbegriffe  $k$  abhängt und im schlechtesten Fall noch von der Anzahl der Inferenzstufen  $m$ . Das PKP hängt immer von der Anzahl der Diskretisierungspunkte  $n$  ab, die je nach Anwendung erheblich größer als  $k$  ist. Wenn  $n$  nicht zu groß ist, könnte man versuchen, den Faktor zwischen  $k$  und  $n$  durch Parallelverarbeitung

auszugleichen. Wenn wir aber an Anwendungen denken, bei denen sich die Größe des Universums im Bereich von  $10^6$  Elementen bewegt, wie etwa den Kunden einer Telefongesellschaft, dann scheidet diese Alternative aus.

MAMDANI und das PKP erlauben die Vorberechnung einer globalen Regel aus den  $k$  Einzelregeln. Der Aufwand fester Regeln ist damit geringer als der variabler Regeln. Bei MAMDANI kann man dann wie bei unserem Ansatz noch auf eine Diskretisierung der Universen verzichten [CP96] und erreicht im Standardfall einen Aufwand von  $O(k^2)$ , da dann auch mehrere Regelstufen zu einer zusammengefasst werden können [LTT98].

Unser Verfahren ist bis auf eine Ausnahme in jedem Fall besser als die anderen. Im Fall variabler Regeln müssen wir  $2m - 1$  Multiplikationen einer Matrix mit einem Vektor durchführen (Abschnitt 5.2.2):  $O(k^2 \cdot m)$ . Am Eingang jeder Regelstufe werden mit Aufwand  $O(k)$  negative Gewichte entfernt. Zusammen erhalten wir damit  $O(k^2 \cdot m)$  als Aufwand. Bei festen Regeln ändert sich daran nichts, falls die Transformationsmatrizen negative Gewichte enthalten. Wenn alle Gewichte positiv sind, müssen wir am Eingang einer Regelstufe keine negativen Gewichte streichen und können alle Matrizen zu einer Matrix zusammenfassen (Gleichung (5.18)). Die Inferenz beschränkt sich dann auf die Multiplikation eines Vektors mit einer Matrix, der Aufwand beträgt  $O(k^2)$ . Im Standardfall enthält die Regelmatrix in jeder Zeile maximal eine Eins; die gesamte Inferenz beschränkt sich auf  $k$  Zuordnungen.

Insgesamt heben wir uns nicht nur vom Prinzip von Kombination/Projektion ab, weil der Rechenaufwand statt von den Einzelheiten nur von der Anzahl der Grundbegriffe abhängt. Wir gewinnen bei variablen Regeln und Standardregeln sogar gegenüber dem schnellen Verfahren von MAMDANI eine Größenordnung im Berechnungsaufwand.

## 5.6 Zusammenfassung

Bei unscharfem Schließen werden aus unscharfen Beobachtungen und einer Wissensbasis durch Inferenz Informationen abgeleitet. In unserem Fall besteht die Wissensbasis aus einer Menge unscharfer Regeln der Form „Wenn  $x = \tilde{A}$ , dann  $y = \tilde{B}$ “ oder kurz  $\tilde{A} \rightarrow \tilde{B}$ . Wir interpretieren eine Regel als bedingte Wahrscheinlichkeit  $P_p(\tilde{B}|\tilde{A}) = 1$ . Diese Darstellung erweitern wir insofern, als wir Regelgewichte einführen, welche die Unsicherheit über das Konsequens  $\tilde{B}$  der Regel ausdrücken:  $\tilde{A} \xrightarrow{P_p(\tilde{B}|\tilde{A})} \tilde{B}$ . Durch die Modellierung der Gewichte als bedingte Wahrscheinlichkeiten  $P_p(\tilde{B}|\tilde{A})$  besitzen sie eine konkrete Semantik, die zudem konsistent mit der gewichteten Kombination von Grundbegriffen ist. Die meisten ande-

ren Ansätze geben keine konkrete Semantik an. So wird beispielsweise in [YT97] die Bedeutung der Regelgewichte nicht erläutert, womit auch die Art ihrer Weiterverarbeitung fragwürdig ist.

Für das Inferenzverfahren greifen wir die Idee aus Kapitel 4 auf, alle unscharfen Informationen mit den Grundbegriffen darzustellen. Indem wir als Antezedenzen und Konsequenzen der Regeln Grundbegriffe verwenden, sind die Regeln auf der grobgranularen Ebene der Begriffswahrscheinlichkeiten definiert. Eine unscharfe Beobachtung stellen wir durch Synthese oder Analyse ebenfalls über die Wahrscheinlichkeiten der Grundbegriffe dar. Mithilfe der Wahrscheinlichkeitsrechnung schließen wir daraus auf die Wahrscheinlichkeiten der Konsequenzen. Der Schluss liegt dann als gewichtete Kombination von Grundbegriffen vor und kann mittels einer Synthese in eine unscharfe Menge übersetzt werden. Alle Berechnungen der Inferenz werden auf der symbolischen Ebene der Begriffswahrscheinlichkeiten durchgeführt. Somit ist das Verfahren effizient und die Schlüsse werden verständlich mit den Grundbegriffen dargestellt. Darüber hinaus haben wir gezeigt, dass unser Inferenzverfahren im Gegensatz zu anderen Ansätzen den Modus Ponens auch dann erfüllt, wenn mehrere Regeln parallel ausgewertet werden.

Bei mehrstufigen Inferenzen werden die Stufen einzeln nacheinander abgearbeitet, indem der unscharfe Schluss einer Stufe als Eingabe für die nächste Stufe benutzt wird. Dieser Schritt findet bei den meisten Verfahren mit hohem Berechnungsaufwand auf feingranularer Ebene statt. Wir zeigen, wie man durch einmalige Vorberechnungen auch diesen Schritt auf grobgranularer Ebene berechnen kann. Dabei können wir die Exaktheit unscharfer Schlüsse auf feingranularer Ebene garantieren. Der Rechenaufwand unserer mehrstufigen Inferenz ist erheblich geringer als bei allen anderen unscharfen Inferenzverfahren. Im besten Fall, der auch gleichzeitig der wichtigste für die Praxis ist, hängt der Rechenaufwand linear von der Anzahl der Regeln ab. Bei dem schnellen Verfahren von MAMDANI ist der Aufwand schon quadratisch und beim Prinzip von Kombination/Projektion geht er quadratisch mit der Anzahl der Einzelheiten und zusätzlich linear mit der Anzahl der Regelstufen.

Insbesondere bei mehrstufigem Schließen offenbart sich die Wichtigkeit des Modus Ponens, der z. B. von MAMDANIS Ansatz nicht erfüllt wird. Verfahrensbedingt erhöht sich bei MAMDANI die Unsicherheit mit jedem weiteren Schluss, bis irgendwann ein völlig unspezifischer Schluss vorliegt: alles ist möglich. Indem unser Verfahren auch über mehrere Stufen den Modus Ponens erfüllt, sind Schlüsse nur so unsicher wie die Regeln; der Inferenzmechanismus fügt keine Unsicherheit hinzu. So bleiben auch Schlüsse über mehrere Stufen aussagekräftig.

Im Normalfall sind die Antezedenzen und Konsequenzen der Regeln Grundbegriffe. Die Universen werden so partitioniert, dass die resultierenden Grundbe-

griffe die Basis für die Regeln bilden. Andererseits könnte es sinnvoll sein, von diesem Schema abzuweichen, wenn z. B. die Grundbegriffe das grobe Verständnis eines Anwenders von einem Universum wiedergeben, die Relation zwischen den Universen aber unabhängig davon aus Daten resultiert. In diesem Fall entsteht eine allgemeine Regel  $\tilde{A} \longrightarrow \tilde{B}$ , deren Antezedens und Konsequens beliebige unscharfe Mengen sind.

Wir haben gezeigt, wie man eine allgemeine Regel durch Regeln auf Basis der Grundbegriffe ersetzt, um sie erstens verständlich zu machen und zweitens die effiziente Verarbeitung mit unserem Inferenzverfahren zu gewährleisten. Auf diese Weise erweitern wir die Mächtigkeit der bisherigen Verfahren, deren Regeln ausschließlich die vordefinierten Begriffe benutzen.

Um die Akzeptanz eines unscharfen Systems sicherzustellen, müssen wir ein Maß für die Zuverlässigkeit der Schlüsse, die das System berechnet, angeben können. Sowohl für grobgranulare als auch für geschärfte, feingranulare Schlüsse haben wir Zuverlässigkeitsmaße definiert, die durch ihre Semantik auf der Basis von Wahrscheinlichkeiten aussagekräftig und leicht verständlich zugleich sind. Wenn beispielsweise die Wissensbasis durchgängig von Experten definiert wird, misst die Zuverlässigkeit den relativen Anteil der Experten, die den Schluss für richtig halten.

Insgesamt erhalten wir eine neues Verfahren für unscharfe Inferenz, das wie schon die Kombination von Grundbegriffen darauf beruht, dass Informationen durch eine Basis von Grundbegriffen dargestellt werden. Das Verfahren ist wesentlich effizienter als andere, weil es durchgängig auf der grobgranularen Ebene operiert. Dabei stellt es erstmalig auch die Verständlichkeit unscharfer Schlüsse sicher, ohne an Präzision zu verlieren. Kurzum: Wir können symbolisch mit unscharfen Begriffen schließen. Im folgenden Kapitel 6 erhärten wir diese Aussage noch insofern, als wir zeigen, dass das Verfahren in gewissem Rahmen unabhängig von der unterliegenden Semantik der Grundbegriffe und der Regeln ist.



## Kapitel 6

---

# Semantik und Repräsentation von Wissen

Eine wichtige Frage bei Anwendungen informationsverarbeitender Systeme ist, wie Wissen repräsentiert wird. Besonders interessant sind hierbei folgende zwei Aspekte. Einmal handelt es sich um die Frage, wie Wissen mathematisch so formalisiert werden kann, dass die Semantik adäquat wiedergegeben wird. Insbesondere ist es bei Anwendungen wichtig, dass eine Theorie der Informationsverarbeitung Wissen mit verschiedenen Semantiken darstellen kann. Anderenfalls leidet die Qualität der Ergebnisse, weil nicht alle Informationen verarbeitet werden können. Der zweite Aspekt der Repräsentation von Wissen hat einen implementierungstechnischen Hintergrund. Eine mathematische Formalisierung von Wissen kann aus der Sicht des Speicher- oder Rechenaufwands ungeeignet für eine rechnergestützte Verarbeitung sein. In diesem Fall müssen wir die Theorie den praktischen Anforderungen anpassen. Wir müssen einen Kompromiss zwischen der Exaktheit der mathematischen Darstellung und dem Aufwand der rechnergestützten Verarbeitung finden. Beides, den semantischen und den implementierungstechnischen Aspekt der Wissensrepräsentation, werden wir in diesem Kapitel näher beleuchten.

Bei Anwendungen liegen oftmals gleichzeitig Informationen vor, die sich durch

1. die Art der Unschärfe (*Vagheit* oder *Unsicherheit*),
2. die Ordnung der Unschärfe (scharf, unsharp  $n$ -ter Ordnung),
3. die Semantik der Beschreibung (bei Regeln z. B. logische Formeln, Abbildungsvorschriften, positive/negative Regeln)

unterscheiden. Ein Beispiel sind Systeme, die den Dollarkurs vorhersagen. Wissen über das Verhalten des Kurses besteht aus statistischen Daten in Form von Zeitreihen aus der Vergangenheit, vagen oder unsicheren Einschätzungen von Experten, scharfen Informationen wie dem Dollarkurs des vorangehenden Tages usw. Die gemeinsame Verarbeitung aller Informationen in einem unscharfen System kann nur dann zu vertrauenswürdigen Ergebnissen führen, wenn wir für alle Formen von Wissen eine gemeinsame mathematische Formalisierung oder Repräsentation finden. Auf die Notwendigkeit dieser Anforderung haben wir bereits auf S. 7 in

der Einleitung mit dem Zitat aus [DP97] hingewiesen. Anderenfalls sind die Informationen nicht vergleichbar und können daher nicht zusammengefasst werden. In diesem Kapitel stellen wir verschiedene Repräsentationen von Wissen aus obiger Aufzählung vor, die bereits in Anwendungen unscharfer Systeme benutzt werden. Dabei zeigen wir, wie sie in unsere Theorie eingebettet und gemeinsam verarbeitet werden können.

Bei der Semantik unscharfer Mengen variieren wir Art und Ordnung der Unschärfe. In Abschnitt 6.1 greifen wir die Interpretation von Unschärfe als *Ähnlichkeit* auf und beschreiben einen bekannten *Gruppierungsalgorithmus*, der in der Praxis zum Generieren unscharfer Mengen aus Daten eingesetzt wird [Höp99]. Wir zeigen, dass auf diese Weise aus Daten unscharfe Partitionen im Sinne dieser Arbeit erzeugt werden können. Nachdem wir uns in den vorangehenden Kapiteln ausführlich mit der Verarbeitung unscharfer Informationen auseinandergesetzt haben, beantworten wir in Abschnitt 6.2 die Frage, wie man scharfe Werte im Kontext grobgranularen Wissens darstellt und verarbeitet. Scharfe Werte kommen in Anwendungen sehr häufig vor und können mit wesentlich geringerem Rechenaufwand verarbeitet werden als unscharfe Informationen, was besonders für technische Systeme mit Echtzeitanforderungen sehr wichtig ist.

In Abschnitt 6.3 diskutieren wir verschiedene Semantiken von Regeln. Wir beleuchten die Bedeutung der logischen Konjunktion und der Implikation als Grundbausteine *positiver* und *negativer Regeln* und stellen sie mit unserem wahrscheinlichkeitstheoretischen Modell für Regeln dar. In der Praxis hat sich die Kombination positiver und negativer Regeln als äußerst nützlich erwiesen, um Regelwerke kompakt und verständlich zu formulieren [CDP96, Kie97, SSR00b]. Mit der Interpretation unseres Inferenzverfahrens als unscharfes Interpolationsverfahren in Abschnitt 6.3.2 leisten wir einen wichtigen Beitrag für die Modellierung technischer Systeme. Die zugrunde liegende Art der Beschreibung einer Regel als funktionaler Zusammenhang entspricht viel eher der Denkweise eines Systemingenieurs als die regelbasierte Sprechweise der Logik. Dabei subsumieren wir einige Verfahren aus dem Bereich unscharfer Regelung wie den Regler von TAKAGI/SUGENO und Neuro-Fuzzy-Ansätze, die wir so mit einer konkreten Semantik versehen und als Verfahren für mehrstufiges unscharfes Schließen deuten können.

Abschließend widmen wir uns in Abschnitt 6.4.1 dem implementierungstechnischen Aspekt der Repräsentation von Wissen. Probleme bei der Implementierung beschränken sich dabei auf die Darstellung unscharfer Grundbegriffe. Wenn Grundbegriffe durch Daten definiert werden, kann der Speicheraufwand für die Definition so groß werden, dass er nicht mehr handhabbar ist. Als Lösung dieses Problems stellen wir ein Parametermodell vor, das den Speicheraufwand drastisch reduziert und sich semantisch nahtlos in unsere Theorie einfügt.

## 6.1 Unscharfe Gruppen

In praktischen Anwendungen von Techniken der Informationsverarbeitung liegt Wissen in vielen Fällen in Form scharfer Daten vor. Man ist aus zwei Gründen daran interessiert, die Daten zu vergrößern:

1. In der Anwendung können verschiedene Daten aufgrund ihrer Ähnlichkeit nicht unterschieden werden, oder eine Unterscheidung ist für die weitere Verarbeitung nicht notwendig.
2. Der Umfang der Daten ist für die Verarbeitung zu groß.

Ein Beispiel sind Anbieter von Musik-CDs im Internet, auf deren WWW-Seiten Kunden CDs bewerten können.<sup>1</sup> Auf Basis dieser Bewertungen werden Gruppen von Kunden gebildet, die ähnliche Musik hören. Ziel dieser Systeme ist, den Kunden CDs zu empfehlen. Einerseits können dies CDs sein, die anderen Kunden der gleichen Gruppe gefallen, andererseits aber auch CDs, welche die Musikexperten des Anbieters dem Geschmack der Gruppe zuordnen. Dabei ist der einzelne Kunde für die Empfehlung einer CD nicht wichtig, es geht nur um die Vorlieben der Gruppe. Berücksichtigt man, dass sich die Anzahl der Kunden im Bereich von mehreren Zehntausend<sup>2</sup> bewegt, so wird auch der Effizienzgewinn durch Gruppenbildungen deutlich.

Ziel dieses Abschnitts ist zu zeigen, wie man durch Gruppierung ähnlicher Daten ein Universum unscharf partitioniert. Insbesondere werden wir darlegen, dass die Elemente der Partition als Plausibilitätsdichten interpretiert werden können. Damit erreichen wir zwei Dinge. Erstens können wir unscharfe Mengen aus Daten erzeugen und mit unserem Schließmechanismus verarbeiten, weil sie die gleiche Semantik aufweisen wie unscharfe Begriffe. Zweitens zeigen wir die Verwandtschaft der Konzepte Unsicherheit und Ähnlichkeit, was die Allgemeinheit unserer Theorie untermauert.

Bei der Gruppierung von Daten werden Einzelheiten zu Mengen, den **Gruppen**, zusammengefasst. Die Elemente einer Gruppe zeichnen sich dadurch aus, dass sie in irgendeiner Hinsicht als ähnlich angesehen werden. Der Ähnlichkeitsbegriff kann sich dabei auf die Einzelheiten selbst beziehen oder auf das, was aus den Einzelheiten in der weiteren Verarbeitung folgt. So können Pilze einerseits ähnlich sein, weil sie derselben Gattung angehören; andererseits können wir Pilze verschiedener Gattungen der gleichen Gruppe zuordnen, weil sie giftig sind und deshalb nicht gegessen werden dürfen.

<sup>1</sup>beispielsweise LAUNCH; URL: <http://www.launch.de> (Oktober 2000)

<sup>2</sup>Bei FIREFLY, dem Vorläufer von LAUNCH, wurden 1998 Zahlen um 25000 genannt.

Unabhängig von der Art des Ähnlichkeitsbegriffs dient Gruppierung zur Vergrößerung von Universen. Nach einer Gruppierung unterscheiden wir nicht mehr Einzelheiten, die in Form von Daten vorliegen, sondern Gruppen, denen die Einzelheiten angehören. Dabei geht Vergrößerung durch Gruppierung den Weg vom Feinen zum Groben. Bei der Partitionierung von Universen in Abschnitt 4.1 sind wir umgekehrt vorgegangen. Wir haben das Universum zunächst symbolisch vergrößert, indem wir unscharfe Begriffe auf grobgranularer Ebene festgelegt haben. Im zweiten Schritt haben wir die Begriffe als unscharfe Mengen auf feingranularer Ebene definiert. Während in diesem Fall Anzahl und Bedeutung der groben Informationseinheiten den Ausgangspunkt bilden, sind es bei der Gruppierung die scharfen Daten. Insbesondere kann man Gruppen u. U. keine anschauliche Bedeutung zuordnen, wie es bei unscharfen Begriffen der Fall ist.

Klassische Gruppierungsverfahren wie *c-Means* und ISODATA fassen Daten zu scharfen Gruppen zusammen [DH73]. Im einfachsten Fall gehört jedes Datum genau einer Gruppe an. Da in vielen Fällen die Daten aber verrauscht oder gestört sind, sind sie inhärent mit Unsicherheit versehen. Unscharfe Gruppierungsverfahren verzichten daher auf die binäre Zuordnung von Daten zu Gruppen und gehen von unscharfen Gruppen aus. Eine unscharfe Gruppe ist eine unscharfe Menge; die Zugehörigkeit eines Datums zu einer unscharfen Gruppe wird mit einem Zugehörigkeitsgrad aus dem Einheitsintervall gemessen. Im Beispiel der Musikliebhaber ist die Notwendigkeit unscharfer Gruppen offensichtlich, denn ihre Vorlieben beschränken sich selten auf eine Musikrichtung. Damit gehören sie mit unterschiedlichen Zugehörigkeitsgraden mehreren Kundengruppen an.

Die klassischen Verfahren für unscharfe Gruppierung sind *Fuzzy-c-Means* [Bez73] und die Algorithmen von GUSTAFSON und KESSEL [GK79] sowie von GATH und GEVA [GG89], die in dieser Reihenfolge aufeinander aufbauen. Einen Überblick geben die Bücher [Bez92, Höp99]. Wir beschränken uns in dieser Arbeit auf das probabilistische Verfahren von GATH und GEVA, dessen Grundzüge bereits einige Jahre früher im Zusammenhang mit BAYESSCHEN Klassifikatoren bekannt waren [Bez81]. Das Ziel dieses Abschnitts ist dabei nicht die detaillierte Beschreibung des Algorithmus, sondern zu zeigen, dass die generierten unscharfen Gruppen als unscharfe Begriffe im Sinne dieser Arbeit aufgefasst werden können.

Wir gehen von einer kompakten Teilmenge  $X$  des  $\mathbb{R}^p$  als Universum aus. Seien  $n$  Daten  $u_i \in X$  gegeben, die zu  $c$  Gruppen zusammengefasst werden sollen. GATH und GEVA nehmen an, dass die Daten Realisierungen  $p$ -dimensionaler, normalverteilter Zufallsvariablen sind. Die Normalverteilungen bezeichnen wir mit  $N_j$ , die zugehörigen Erwartungswerte mit  $m_j$  und die Kovarianzmatrizen mit  $C_j$ . Wenn wir jedes Datum  $u_i$  genau einer Normalverteilung  $N_j$  zurechnen, entspricht dies binären Zugehörigkeiten  $\mu_{N_j}(u_i) \in \{0, 1\}$ . Die Erwartungswerte und Kovarianzma-

trizen der Normalverteilungen können wir mit statistischen Schätzern berechnen:

$$m_j = \frac{\sum_{i=1}^n \mu_{N_j}(u_i) \cdot u_i}{\sum_{i=1}^n \mu_{N_j}(u_i)} \quad (6.1)$$

$$C_j = \frac{\sum_{i=1}^n \mu_{N_j}(u_i) \cdot (u_i - m_j)(u_i - m_j)^t}{\sum_{i=1}^n \mu_{N_j}(u_i)} \quad (6.2)$$

Weiterhin wird davon ausgegangen, dass die Normalverteilung  $N_j$  mit der a-priori-Wahrscheinlichkeit  $P(N_j)$  gewählt wird, um ein Datum zu erzeugen:

$$P(N_j) = \frac{\sum_{i=1}^n \mu_{N_j}(u_i)}{n} \quad (6.3)$$

Damit ergibt sich für die a-posteriori-Wahrscheinlichkeit  $P(N_j|u_i)$  nach der Formel von BAYES

$$P(N_j|u_i) = \frac{P(N_j) \cdot P(u_i|N_j)}{\sum_{k=1}^c P(N_k) \cdot P(u_i|N_k)}, \quad (6.4)$$

wobei  $P(u_i|N_j)$  mit der Dichte der Normalverteilung  $N_j$  identifiziert wird:

$$P(u_i|N_j) = \frac{1}{(2\pi)^{p/2} \sqrt{\det(C_i)}} \exp\left(-\frac{1}{2}(u_i - m_j)^t C_j^{-1} (u_i - m_j)\right) \quad (6.5)$$

Die unscharfen Gruppen werden durch die Likelihood-Funktionen  $P(N_j|u)$  repräsentiert. Der Gruppierungsalgorithmus bestimmt die Parameter  $m_j$ ,  $C_j$  und  $P(N_j)$  – und damit die Likelihood-Funktion  $P(N_j|u)$  – so, dass folgende Zielfunktion minimiert wird:

$$Z := \sum_{i=1}^n \sum_{j=1}^c P(N_j|u_i)^2 \cdot d^2(u_i, m_j) \quad (6.6)$$

mit einer *Distanzfunktion*  $d^2(u_i, m_j)$ , die den Abstand zwischen dem Datum  $u_i$  und dem Erwartungswert  $m_j$  der Gruppe  $j$  misst. Die Zielfunktion ist eine Verallgemeinerung der Zielfunktion für die Methode der kleinsten Quadrate, die wir auch in Abschnitt 4.2.3 verwendet haben.  $Z$  zu minimieren bedeutet, die Gruppen so zu wählen, dass die Elemente der Gruppen einen möglichst kleinen Abstand zu ihrem Erwartungswert haben. Um die Zielfunktion zu minimieren, verallgemeinern GATH und GEVA die statistischen Schätzer (6.1) und (6.2). Sie lösen sich von der Annahme, dass jedes Datum von genau einer Normalverteilung erzeugt wird, indem sie die scharfen Zugehörigkeiten  $\mu_{N_j}(u_i)$  durch die quadrierten Likelihood-Funktionen  $P(N_j|u_i)^2$  ersetzen. Dies hat sich als praktikable Heuristik für die Optimierung erwiesen. Damit variieren sie die Maximum-Likelihood-Schätzer der

BAYES-Klassifikatoren [Bez81]. Dabei werden die Parameter  $m_j$ ,  $C_j$  und  $P(N_j)$  näherungsweise so bestimmt, dass die Eintrittswahrscheinlichkeit des Datensatzes  $u_i \in \mathcal{X}$  maximiert wird. Die a-priori-Wahrscheinlichkeiten  $P(N_j)$  bestimmen sie als Verallgemeinerung von (6.3):

$$P(N_j) = \frac{\sum_{i=1}^n P(N_j|u_i)^2}{\sum_{i=1}^n \sum_{j=1}^c P(N_j|u_i)^2} \quad (6.7)$$

Der Gesamtalgorithmus für die Gruppierung ist iterativ. Ausgehend von einer initialen Gruppierung werden Erwartungswerte, Kovarianzmatrix und a-priori-Wahrscheinlichkeit berechnet und damit neue Gruppen als Likelihood-Funktionen definiert. Dieser Vorgang wird als alternierende Optimierung so oft wiederholt, bis die maximale Änderung der Likelihood-Funktionen einen Schwellenwert unterschreitet.

Der Zusammenhang zwischen unscharfen Grundbegriffen im Sinne dieser Arbeit und unscharfen Gruppen nach GATH und GEVA bzw. BAYES-Klassifikatoren ist sehr einfach, weil sowohl die Grundbegriffe als auch die Gruppen durch Likelihood-Funktionen  $P(N_j|u)$  repräsentiert werden. Mathematisch gesehen sind unscharfe Gruppen und Grundbegriffe also identisch. Die Likelihood-Funktionen bilden wie unscharfe Grundbegriffe eine unscharfe Partition von  $\mathcal{X}$ , wie wir bereits in Abschnitt 4.1 gesehen haben.

Folgende Überlegungen zeigen, dass die Unschärfe der Gruppen durch Ähnlichkeit induziert wird. Der Erwartungswert  $m_j$  der Normalverteilung  $N_j$  dient als Prototyp der Gruppe  $j$ . GATH und GEVA definieren die Distanzfunktion  $d^2(u, m_j)$  umgekehrt proportional zu  $P(u|N_j)$ . Da  $P(u|N_j)$  nach Gleichung (6.5) die Dichte einer Normalverteilung ist, ist die Distanzfunktion  $d^2(u, m_j)$  eine monotone Transformation des EUKLIDISCHEN Abstands  $|u - m_j|$ . Ein kleiner Wert  $P(u|N_j)$  bedeutet, dass der Abstand zwischen dem Datum  $u$  und dem Prototypen  $m_j$  der Gruppe  $j$  sehr groß ist, und ein großer Wert entspricht einem kleinen Abstand von  $u$  und  $m_j$ . Auf diese Weise misst  $P(u|N_j)$  die Ähnlichkeit zwischen dem Datum  $u$  und dem Prototyp  $m_j$  der Gruppe. Die Wahrscheinlichkeitsdichten  $P(u|N_j)$  selbst bilden keine unscharfe Partition des Universums  $\mathcal{X}$ , weil sie die Überdeckungseigenschaft (4.3) nicht erfüllen. Daher werden sie durch Gewichtung und Normierung in die Likelihood-Funktionen  $P(N_j|u)$  überführt, die das Universum  $\mathcal{X}$  partitionieren und die Zugehörigkeitsfunktionen der unscharfen Gruppen bilden.

Insgesamt sehen wir, dass wir durch unscharfe Gruppierung unscharfe Partitionen aus Daten gewinnen können. Die Gruppen beruhen auf einer Semantik, die Unschärfe durch die Ähnlichkeit von Daten ausdrückt. Die mathematische Formalisierung der Gruppen ist dabei mit der Formalisierung unscharfer Grundbegriffe identisch. Damit haben wir die Darstellung unscharfer Mengen auf der Basis von

Unsicherheit und Ähnlichkeit in einem Ansatz integriert und können sie gemeinsam mit unserem Inferenzmechanismus verarbeiten.

## 6.2 Darstellung scharfer Werte

Speziell bei technischen Systemen liegen Beobachtungen häufig als scharfe Messwerte vor. Selbst wenn wir eine Beobachtung wegen potentieller Messfehler nicht als perfekt annehmen können, wird i. A. mit dem scharfen Messwert gerechnet, weil der Berechnungsaufwand für die Verarbeitung sehr gering ist. Die Alternative ist, falls vorhanden, Wissen über die Unsicherheit der scharfen Beobachtung einzubringen. Das Ergebnis ist normalerweise eine Wahrscheinlichkeitsdichte, die als spezielle Plausibilitätsdichte die unscharfe Beobachtung darstellt. Diese Transformation eines scharfen Wertes in einen unscharfen Wert nennt man **Abstumpfung**<sup>3</sup>. In der Praxis muss man abwägen, ob man mit dem potenziell richtigen, abgestumpften Wert rechnet oder den unsicheren scharfen Wert benutzt, was den Berechnungsaufwand stark reduziert. In diesem Abschnitt zeigen wir, wie man scharfe Werte in unserem Ansatz zum Schließen mit unscharfen Begriffen repräsentiert und effizient verarbeitet.

Für die Weiterverarbeitung im Inferenzmechanismus haben wir unscharfe Beobachtungen in Abschnitt 4.2.3 durch eine gewichtete Kombination der Grundbegriffe dargestellt. Dieses Vorgehen ist für scharfe Beobachtungen semantisch nicht vernünftig, denn wir können eine scharfe Beobachtung nicht annähernd exakt mit unscharfen Begriffen erklären. Auf technischer Seite offenbart sich dies insofern, als Ansatz 2 zur Analyse keine sinnvollen Ergebnisse liefert. Bei der Analyse einer unscharfen Beobachtung berechnen wir Begriffswahrscheinlichkeiten so, dass die gewichtete Kombination der Grundbegriffe die unscharfe Beobachtung möglichst gut rekonstruiert. Mathematisch gesehen approximieren wir die beobachtete unscharfe Menge durch eine Kombination der unscharfen Mengen der Grundbegriffe. Wenn nun der Grenzfall eintritt, dass die unscharfe Beobachtung ein scharfer Wert ist, dann versuchen wir bei der Analyse, einen scharfen Wert mit unscharfen Mengen zu approximieren. Es ist klar, dass dies nicht sinnvoll ist. Das Optimierungsverfahren von Ansatz 2 berechnet in der kontinuierlichen Fassung alle Begriffswahrscheinlichkeiten zu null. Der Grund liegt in der Fehlerfunktion (4.18). Ein scharfer Wert  $u_0$  wird durch die Plausibilitätsdichte dargestellt, die an der Stelle  $u_0$  gleich eins ist und an allen anderen Stellen null. Diese einzelne Spitze hat in der Fehlerfunktion keine Auswirkung, da die Norm  $\|\cdot\|_2$  in der Fehlerfunktion ein Integral ist. Als LEBESGUESISCHE Nullmenge können wir nämlich die Spitze an der Stelle  $u_0$  wegnehmen, ohne den Wert des Integrals zu ändern. Insbesondere ändert

---

<sup>3</sup>engl. *fuzzification*

sich die Fehlerfunktion also nicht, wenn wir statt der Plausibilitätsdichte des scharfen Wertes die Nullfunktion einsetzen, die sich nur an der Stelle  $u_0$  unterscheidet. Damit approximieren wir aber die Nullfunktion, was naturgemäß dazu führt, dass alle Begriffswahrscheinlichkeiten null werden.

Als Lösung des Problems schlagen wir vor, einen scharfen Wert vor der weiteren Verarbeitung abzustumpfen. Diese Transformation sollte im Sinne der unscharfen Partition  $\mathcal{P}$  des Universums  $\mathcal{X}$  durchgeführt werden. Dazu klassifizieren wir die scharfe Beobachtung, indem wir die Zugehörigkeit zu den Grundbegriffen berechnen. Unter der Annahme einer scharfen Beobachtung  $u_0$  suchen wir dazu die Begriffswahrscheinlichkeiten  $P_{\mathcal{P}}(\tilde{B})$  für alle  $\tilde{B} \in \mathcal{P}$ , d. h. die bedingten Wahrscheinlichkeiten  $P_{\mathcal{P}}(\tilde{B}|u_0)$ . Wir kennen sie bereits aus Abschnitt 2.3.1:

$$P_{\mathcal{P}}(\tilde{B}|u_0) = L(\tilde{B}; u_0) = pl_{\tilde{B}}(u_0) \quad (6.8)$$

mit der Likelihood-Funktion  $L(\tilde{B}; u_0)$ . Wir können also die Begriffswahrscheinlichkeiten als Zugehörigkeit zu den Grundbegriffen direkt von den Plausibilitätsdichten der Grundbegriffe ablesen. Wir benutzen diese Begriffswahrscheinlichkeiten nun, um einen unscharfen Wert zu synthetisieren; wir bilden die gewichtete Kombination der Grundbegriffe. Der unscharfe Wert, den wir auf diese Weise gewinnen, stellt die natürliche Vergrößerung der scharfen Beobachtung dar. Wir stumpfen so den scharfen Wert im Sinne der vorgegebenen Partitionierung des Universums ab und bewegen uns dann auf der Granularitätsstufe der Grundbegriffe.

**Definition 7** *Gegeben sei eine unscharfe Partition  $\mathcal{P}$  des Universums  $\mathcal{X}$ . Die natürliche Abstumpfung eines scharfen Wertes  $u_0 \in \mathcal{X}$  bzgl.  $\mathcal{P}$  ist die Plausibilitätsdichte*

$$pl_{u_0}(u) = \sum_{\tilde{B} \in \mathcal{P}} pl_{\tilde{B}}(u_0) \cdot pl_{\tilde{B}}(u) \quad (6.9)$$

Das Vorgehen können wir somit folgendermaßen zusammenfassen: Wenn wir einen scharfen Wert mit den Grundbegriffen beschreiben möchten, stumpfen wir ihn nach Def. 7 ab. Daraus resultieren die Begriffswahrscheinlichkeiten, mit denen wir den abgestumpften Wert als gewichtete Kombination der Grundbegriffe darstellen. Für die Weiterverarbeitung mit unserem Inferenzverfahren greifen wir direkt auf die Begriffswahrscheinlichkeiten zurück. Insbesondere müssen wir mit diesem Ansatz keine aufwändige Analyse einer scharfen Beobachtung durchführen, sondern werten lediglich die Plausibilitätsdichten der Grundbegriffe an einer Stelle aus.

Abbildung 6.1 zeigt einen scharfen Wert  $u_0$  und seine natürliche Abstumpfung bzgl. der Grundbegriffe. Die resultierende Plausibilitätsdichte  $pl_{u_0}$  ist sehr unscharf in dem Sinne, dass der Träger sehr breit ist. Wir dürfen dieses Vorgehen

nicht mit der Abstumpfung vergleichen, die z. B. bei unscharfer Regelung nach MAMDANI benutzt wird. Dort wird meist der scharfe Wert übernommen. Unsere Form der Abstumpfung wird in MAMDANIS Ansatz implizit im nächsten Schritt durchgeführt, wenn der Grad bestimmt wird, mit dem eine Regel feuert. Wir können leicht unscharfe Beobachtungen konstruieren, welche die gleichen Grade zum Feuern erzeugen wie die scharfe Beobachtung. Es sind diese unscharfen Beobachtungen, die man mit unserer natürlichen Abstumpfung vergleichen muss.

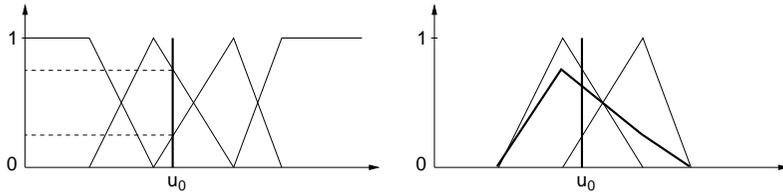


Abbildung 6.1: Natürliche Abstumpfung einer scharfen Beobachtung  $u_0$ ; links vier Grundbegriffe und die scharfe Beobachtung, rechts die relevanten Grundbegriffe (dünn) und die natürliche Abstumpfung (fett) als ihre Kombination

Alternativ zur natürlichen Abstumpfung könnte man einen unscharfen Wert  $u_0$  abstumpfen, indem man ihn z. B. als Erwartungswert einer Normalverteilung definiert oder ein Intervall  $[u_0 - \varepsilon, u_0 + \varepsilon]$  um  $u_0$  legt. Um den abgestumpften Wert mit den Grundbegriffen darzustellen, müssen wir eine Analyse nach Ansatz 2 durchführen. Da das Ergebnis der Analyse aber von der Art der Abstumpfung abhängt, muss möglichst präzises Wissen über die Unsicherheit des scharfen Wertes vorliegen, um ein sinnvolles Ergebnis zu berechnen. Anderenfalls bestimmt die Analyse Begriffswahrscheinlichkeiten, die ebenso willkürlich sind wie die gewählte Abstumpfung. Neben der Berechnungseffizienz ist die natürliche Abstumpfung daher wegen der sauberen Semantik auf Basis der Likelihood-Funktionen vorzuziehen.

### 6.3 Modellierung unscharfer Regeln

Unscharfe Regeln haben wir in Abschnitt 5.1 über bedingte Begriffswahrscheinlichkeiten modelliert. Es gibt viele Möglichkeiten, um diese Wahrscheinlichkeiten zu gewinnen. Dazu zählen Zufallsexperimente, statistische Erhebungen und subjektive Einschätzungen. Diese Methoden sind aus der Wahrscheinlichkeitsrechnung wohlbekannt. Wir werden in diesem Abschnitt zeigen, dass wir auch ganz andere Interpretationen von Regeln mit unserem wahrscheinlichkeitstheoretischen Modell darstellen können.

Zunächst betrachten wir in Abschnitt 6.3.1 noch einmal den Modus Ponens und zeigen, dass man eine Regel sowohl als logische Konjunktion als auch als Implikation verstehen kann, wenn man den Modus Ponens erfüllen möchte. Die Konjunktion verstehen wir dabei als positive Regel, weil sie das Konsequens der Regel unter Annahme des Antezedens unterstützt. Negative Regeln werden durch Implikationen realisiert; wenn das Antezedens erfüllt ist, schließen sie das Komplement des Konsequens aus. In der Possibilitätstheorie wie auch beim praktischen Entwurf von unscharfen Regelwerken werden diese beiden Arten von Regeln in Kombination benutzt, um Wissensbasen aufzubauen. Mit diesem Vorgehen adaptieren wir menschliche Denkweisen. Manche Schlüsse können wir besser durch Ausschlussverfahren ziehen (negative Regeln), bei anderen wiederum schlagen wir direkt den richtigen Schluss vor (positive Regeln). Wir zeigen im folgenden Abschnitt, wie man diese Art der Regelformulierung mit bedingten Wahrscheinlichkeiten realisiert.

Danach gehen wir in Abschnitt 6.3.2 auf die Deutung unscharfen Schließens als unscharfes Interpolationsverfahren ein. Nachdem seit den 70er-Jahren insbesondere der regelbasierte Ansatz von MAMDANI benutzt wurde, um Regler oder die Übertragungsfunktion technischer Systeme mit unscharfen Regeln zu beschreiben, behaupteten sich in den 90er-Jahren zu diesem Zweck Ansätze, die als Funktionsapproximatoren erstens das Expertenwissen besser repräsentieren können und zweitens leichter an Daten angepasst werden können. Dabei handelt es sich um das Verfahren von TAKAGI/SUGENO und Neuro-Fuzzy-Ansätze. Wir zeigen vor diesem Hintergrund, dass unser Inferenzverfahren als ein unscharfes Interpolationsverfahren interpretiert werden kann. Es erlaubt die Vorgabe von unscharfen Stützpunkten auf einer gesuchten unscharfen Abbildung, zwischen denen das Inferenzverfahren interpoliert. Außerdem betten wir den Ansatz von TAKAGI/SUGENO und Neuro-Fuzzy-Modelle in unsere Theorie ein. Aus der Sicht unscharfen Schließens waren diese Verfahren bisher nur intuitiv motiviert. Wir zeigen, dass sie Spezialfälle unserer allgemeinen Theorie unscharfen Schließens sind und stellen sie so auf ein semantisch solides Fundament.

### **6.3.1 Positive und negative Regeln**

Bei der Formulierung von Wissen ist es in vielen Fällen sinnvoll, Evidenzen nicht nur über die Gültigkeit von Aussagen zu sammeln, sondern auch über eine mögliche Ungültigkeit. Wenn wir uns z. B. beim Kauf eines Autos für ein Modell entscheiden müssen, dann teilen wir mögliche Modelle meist in drei Klassen ein: erstens solche Modelle, die uns auf Anhieb gefallen, zweitens diejenigen, die uns nicht zusagen und drittens jene Wagen, zu denen wir noch keine Meinung haben. Der Wahl eines Autos aus der ersten Klasse würden wir sofort zustimmen. Wir

geben also einen direkten Hinweis für die Lösung des Problems. Solche Informationen bezeichnen wir als **positive Informationen**. Auf der anderen Seite können wir die Wahl eines Wagens aus der zweiten Klasse ausschließen. Diese **negative Information** sagt uns, wie die Lösung des Problems *nicht* aussieht. Es bleibt die dritte Klasse von Modellen, die wir zunächst keiner der ersten beiden Klassen zuordnen können. Die Lösung des Problems liegt in der Vereinigung der ersten und der dritten Klasse.

Positives Wissen ist eine Information für das unscharfe System, die nach der Verarbeitung direkt in einem *möglichen* Schluss resultiert; prinzipiell sind aber noch andere Schlüsse denkbar. Positives Wissen ist also im Zweifelsfall nicht vollständig. Auf der anderen Seite schränkt negatives Wissen den Raum möglicher Schlüsse ein. Das unscharfe System schlägt also u. U. Lösungen vor, die falsch oder nicht optimal sind, aber beim vorhandenen Wissensstand nicht ausgeschlossen werden können. Es ist offensichtlich, dass die Kombination von positivem und negativem Wissen mehr Informationen enthält als ein Typ allein.<sup>4</sup> Wenn wir als Beispiel ein Optimierungsproblem betrachten, dann schränkt negatives Wissen den Suchraum der optimalen Lösung ein. Positives Wissen kann die Suche insofern weiter steuern, als es konkrete Vorschläge für die Lösung angibt.

In der Literatur unscharfer Systeme finden sich einige Ansätze, die positives und negatives Wissen vereinigen. In der Possibilitätstheorie werden Possibilitätsdichten sogenannte *Evidenzdichten*<sup>5</sup> gegenübergestellt [Wei95b, Wei95a]. Possibilitätsdichten tragen negative Informationen [Spo94, Spo99], Evidenzdichten repräsentieren positives Wissen. Verfahren zur Kombination von Possibilitäts- und Evidenzdichten werden in [CDP96, SW96, Wei96, WS00] vorgestellt. Diese Ansätze beruhen auf einem axiomatischen Verständnis der Possibilitätstheorie [Spo94, Spo99, Wei95b] und nicht auf einer konkreten Semantik wie der Ansatz dieser Arbeit. Dass die Kombination von positivem und negativem Wissen Vorteile beim praktischen Entwurf unscharfer Systeme besitzt, zeigen z. B. einige Arbeiten aus dem Bereich unscharfer Regelung. In [Kie97] wird darauf hingewiesen, dass der Entwurf eines unscharfen Systems schneller bewerkstelligt werden kann und die resultierenden Regeln leichter verständlich sind. Wenn wir z. B. in einer Landschaft ein Ziel erreichen sollen, dann werden wir es bei Sichtkontakt direkt anvisieren (positive Richtungsinformation). Treffen wir dabei jedoch auf ein Hindernis, dann weichen wir aus – wir laufen *nicht* in der Richtung des Hindernisses und des Zieles weiter (negative Richtungsinformation). Im lernenden Regelungssystem FYNESSE wird positives Wissen in Form aktueller Regelungsstrategien dazu benutzt, um den Suchraum möglicherweise besserer Alternativstrategien ein-

---

<sup>4</sup>Gleiche Mächtigkeit liegt vor, wenn das positive Wissen das Komplement des negativen Wissens ist, wenn also z. B. Informationen entweder richtig oder falsch sind.

<sup>5</sup>engl. *support distributions*

zuschränken [SSR00a, SSLR99, SSR00b]. Der Suchraum stellt dabei negatives Wissen dar, denn er enthält nur solche Strategien, die bei gegebenem Wissenstand nicht ausgeschlossen werden können. Wir sehen an dieser Anwendung, dass positives Wissen genutzt werden kann, um negatives Wissen zu adaptieren, und dass wir tatsächlich beide Formen von Wissen in Kombination benötigen.

Wir beschreiben die Modellierung positiver und negativer Regeln aus Sicht der formalen Logik und übertragen die Erkenntnisse auf die Ansätze dieser Arbeit. Nehmen wir als Ausgangspunkt eine Regel „Wenn Aussage  $a$  gilt, dann ist Aussage  $b$  richtig.“ Aus der Richtigkeit von  $a$  soll die Richtigkeit von  $b$  folgen. Die Frage ist aber, was für  $b$  folgt, wenn  $a$  nicht gilt. In der Wahrheitstabelle 6.1 (links) können wir mit der Regel nur die zweite Zeile definieren. In der klassischen, zweiwertigen Logik wird die Regel normalerweise als *Implikation*  $a \rightarrow b$  gedeutet. In der Wahrheitstabelle 6.1 sehen wir in der Mitte die zugehörigen Wahrheitswerte. Nur wenn das Antezedens  $a$  erfüllt ist und das Konsequens  $b$  falsch ist, ist auch die Implikation falsch. Falls  $a$  falsch ist, können wir daraus keine Aussage über die Richtigkeit von  $b$  treffen. Bei der Implikation kommt dann *ex falso quodlibet* zum Tragen: Alles ist möglich. Lediglich wenn  $a$  wahr ist, können wir ausschließen, dass  $b$  falsch ist. Eine so interpretierte Regel nennen wir **negative Regel**, weil nur das ausgeschlossen wird, was sicher *nicht* gilt. „Aussage  $a$  ist wahr.“ interpretieren wir als „Die Wahrheit von Aussage  $a$  ist möglich.“ bzw. „Die Wahrheit von Aussage  $a$  kann nicht ausgeschlossen werden.“

Dem gegenüber steht die Interpretation der Regel als Konjunktion  $a \wedge b$ . „Aussage  $a$  ist wahr.“ verstehen wir als „Wir unterstützen, dass die Aussage  $a$  wahr ist.“ Wenn wir nicht sicher sind, ob  $a$  gilt, dann nehmen wir an, dass  $a$  falsch ist. Entsprechend füllen wir die erste Zeile der Wahrheitstabelle 6.1 (rechts): Falls  $a$  falsch ist, können wir weder Richtigkeit noch Ungültigkeit von  $b$  unterstützen. Die Konjunktion ist in beiden Fällen falsch. Nur wenn  $a$  wahr ist, können wir aus der Richtigkeit der Regel folgern, dass  $b$  ebenfalls wahr ist. Wir bezeichnen eine so interpretierte Regel als **positive Regel**, weil nur das unterstützt wird, was sicher richtig ist.

Implikation und Konjunktion stellen insofern extreme Interpretationen der Wenn-dann-Regel dar, als sie die unbekanntenen Wahrheitswerte der Regel mit maximalen bzw. minimalen Werten belegen. Den Zusammenhang zwischen Konjunktion und Implikation sehen wir durch einfache logische Umformungen. Für alle Belegungen von  $a, b$  gilt

$$a \rightarrow b = \neg(a \wedge \neg b) \quad \text{bzw.} \quad a \wedge b = \neg(a \rightarrow \neg b) \quad (6.10)$$

Auf diese Weise können wir die Implikation durch eine negierte Konjunktion ausdrücken.

|       |   |   |
|-------|---|---|
| a \ b | 0 | 1 |
| 0     | ? | ? |
| 1     | 0 | 1 |

Wenn a wahr ist,  
dann ist b wahr.

|       |   |   |
|-------|---|---|
| a \ b | 0 | 1 |
| 0     | 1 | 1 |
| 1     | 0 | 1 |

$a \rightarrow b$

|       |   |   |
|-------|---|---|
| a \ b | 0 | 1 |
| 0     | 0 | 0 |
| 1     | 0 | 1 |

$a \wedge b$

Tabelle 6.1: Wahrheitstabellen der Regel (links), von Implikation (Mitte) und Konjunktion (rechts)

In einer mehrwertigen Logik und dementsprechend auch in einer unscharfen Logik gelten die Beziehungen (6.10) i. A. nicht. Aus diesem Grund existieren viele Vorschläge, um z. B. eine unscharfe Implikation zu modellieren. Je nach den Anforderungen der Anwendung wird man sich für eine Realisierung der Implikation entscheiden müssen, was in vielen Fällen nicht einfach ist. In der Possibilitätstheorie haben sich zwei Relationen ausgezeichnet, die eine unscharfe Regel  $\tilde{A} \rightarrow \tilde{B}$  modellieren. Die erste Relation ist die GÖDEL-Relation, die eine Implikation modelliert und für eine negative Interpretation der Regel steht. Innerhalb eines Axiomensystems für die Verarbeitung negativen Wissens kann sie eindeutig als Modell für eine negative unscharfe Regel bewiesen werden [Spo94, Spo99]. Dem gegenüber steht die *Minimum*-Relation. Sie stellt die größte unscharfe Konjunktion dar und modelliert damit eine positive Regel. Ähnlich wie die GÖDEL-Relation kann sie aus einem Axiomensystem für positives Wissen eindeutig als Modell für eine positive unscharfe Regel abgeleitet werden [Wei95b, Wei95a].

Die Berechnung einer Inferenz auf der Basis der GÖDEL-Relation ist wesentlich aufwändiger als mit der *Minimum*-Relation. Daher liegt die Idee nahe, negative Regeln über den Zusammenhang (6.10) als positive Regeln darzustellen und diese auszuwerten [Kie97]. Da wie bereits erwähnt die Darstellungen aber nicht äquivalent sind, erhalten wir nur eine Approximation der ursprünglichen Aussagen.

Wir werden die Modellierung von positiven und negativen Regeln auf den Wahrheitstabellen klassischer Logik aufbauen. Daher gelten in unserem Fall die Beziehungen (6.10). Die Differenzierung positiver und negativer Regeln spielt damit nur für den Entwurf des regelbasierten Systems eine Rolle, nicht jedoch für die Berechnung von Inferenzen.

Das Äquivalent zum Wahrheitswert einer unscharfen Aussage ist in unserer Modellierung ihre Wahrscheinlichkeit. Wenn es keine Evidenz für eine Aussage gibt, ist ihre Wahrscheinlichkeit null. Diese Erkenntnis legt eine Interpretation von  $\tilde{A} \rightarrow \tilde{B}$  als positive unscharfe Regel nahe. Wir modellieren die Regel als bedingte Wahrscheinlichkeiten  $P_p(\tilde{B}|\tilde{A}) = 1$ ,  $P_p(\tilde{B}^c|\tilde{A}) = 0$ , was der zweiten Zeile der Tabellen 6.1 entspricht. Da für den Fall  $P_p(\tilde{A}) = 0$ , also  $P_p(\tilde{A}^c) = 1$ , keine Evidenzen über  $P_p(\tilde{B})$  vorhanden sind, müssten wir analog zu Tab. 6.1 (rechts)

eigentlich  $P_{\mathcal{P}}(\tilde{B}|\tilde{A}^{\mathfrak{C}}) = P_{\mathcal{P}}(\tilde{B}^{\mathfrak{C}}|\tilde{A}^{\mathfrak{C}}) = 0$  definieren. Das steht allerdings im Widerspruch zur Wahrscheinlichkeitsrechnung, denn da wir eine abgeschlossene Welt annehmen, gilt  $P_{\mathcal{P}}(\tilde{B}|\tilde{A}^{\mathfrak{C}}) + P_{\mathcal{P}}(\tilde{B}^{\mathfrak{C}}|\tilde{A}^{\mathfrak{C}}) = 1$ . Anders ausgedrückt: Wenn es nur zwei Alternativen gibt, muss mindestens eine richtig sein. Das Problem können wir formal so lösen, dass wir für alle unscharfen Werte von  $x$  eine Regel annehmen. In diesem Fall bedeutet das, dass eine positive Regel mit dem Antezedens  $\tilde{A}$  nicht allein stehen kann. Sie kann nur zusammen mit einer Menge von Regeln auftreten, deren Antezedenzen vereinigt  $\tilde{A}^{\mathfrak{C}}$  überdecken. Wir bezeichnen diese Anforderung als **Vollständigkeit der Regelbasis**. Unter der Annahme einer vollständigen Regelbasis können wir tatsächlich die bedingten Wahrscheinlichkeiten gemäß der Tab. 6.1 (rechts) definieren. Jede Regel resultiert in einer Tabelle; die Tabellen aller Regeln fassen wir zusammen, indem wir elementweise den Mittelwert bilden und anschließend die Zeilensummen auf eins normieren. Das Ergebnis ist eine Definition aller bedingten Wahrscheinlichkeiten, die alle Regeln integriert und die Anforderungen der Wahrscheinlichkeitsrechnung erfüllt. Wir beschreiben das Verfahren der Aggregation weiter unten genauer. Darüber hinaus illustriert Beispiel 1 dieses Vorgehen für ein System aus drei Regeln.

Wir klären nun, wie wir eine negative Regel mit Wahrscheinlichkeiten darstellen. Wie bei der positiven Regel gelten analog zur Wahrheitstabelle 6.1  $P_{\mathcal{P}}(\tilde{B}|\tilde{A}) = 1$  und  $P_{\mathcal{P}}(\tilde{B}^{\mathfrak{C}}|\tilde{A}) = 0$ . Im Fall von  $P_{\mathcal{P}}(\tilde{A}) = 0$ , d.h.  $P_{\mathcal{P}}(\tilde{A}^{\mathfrak{C}}) = 1$ , sind  $\tilde{B}$  und  $\tilde{B}^{\mathfrak{C}}$  ohne Zusatzwissen gleich wahrscheinlich. Es folgt daher  $P_{\mathcal{P}}(\tilde{B}|\tilde{A}^{\mathfrak{C}}) = P_{\mathcal{P}}(\tilde{B}^{\mathfrak{C}}|\tilde{A}^{\mathfrak{C}}) = 1/2$ . Wir können diese Wahrscheinlichkeiten auch direkt aus der Wahrheitstabelle der Implikation in Tab. 6.1 ableiten. Dazu müssen wir wiederum beachten, dass die Einträge so normiert werden müssen, dass die Zeilensummen eins ergeben. Im Gegensatz zu positiven Regeln sind hier bereits durch eine einzelne Regel die Wahrscheinlichkeiten vollständig spezifiziert. Weitere Regeln präzisieren den Zusammenhang zwischen den Universen  $\mathcal{X}$  und  $\mathcal{Y}$  nur weiter.

Im Normalfall besteht ein regelbasiertes System aus vielen Regeln, die durch die bedingten Wahrscheinlichkeiten  $P_{\mathcal{P}}(\tilde{B}_j|\tilde{A}_i)$  dargestellt werden sollen. Um positive und negative Regeln zu mischen, gehen wir nach dem unten beschriebenen Verfahren vor. Dabei stellen wir für jede Regel eine Wahrheitstabelle wie in Tab. 6.1 auf, die der Anzahl der Grundbegriffe  $\tilde{A}_i \in \mathcal{P}_{\mathcal{X}}$  und  $\tilde{B}_j \in \mathcal{P}_{\mathcal{Y}}$  entsprechend groß ist. Die Einträge der Tabelle bezeichnen wir mit  $\gamma_{ij}^k$ , wobei  $k$  der Index der Regel ist.

### Verfahren für die Definition einer Regelbasis

1. Formales Aufstellen von positiven und negativen Regeln
2. Für jede Regel Bilden der Wahrheitstabelle mit Einträgen  $\gamma_{ij}^k$
3. Falls  $\sum_j \gamma_{ij}^k = 0$  für ein  $i$  und alle  $k$  gilt, dann Regeln ergänzen und zurück zu Punkt 1

4. Für jede Regel zeilenweise Normierung der  $\gamma_{ij}^k$ :

$$\forall k : \gamma_{ij}^k := \gamma_{ij}^k / \sum_l \gamma_{il}^k \tag{6.11}$$

5. Aggregation der Regeln über elementweise Durchschnittsbildung, sei  $n$  die Anzahl der Regeln:

$$P_{\mathcal{P}}(\tilde{B}_j | \tilde{A}_i) := 1/n \sum_k \gamma_{ij}^k \tag{6.12}$$

Die Aggregation der Regeln über die Durchschnittsbildung in Punkt 5 entspricht dem üblichen Vorgehen in der Wahrscheinlichkeitsrechnung, gleichberechtigte Informationen zusammenzufassen. Wir haben dieses Vorgehen bereits in Abschnitt 5.3 benutzt. Punkt 3 des Verfahrens sichert die Vollständigkeit der Regelbasis.

**Beispiel 1** Nehmen wir an, wir hätten drei unscharfe Regeln  $\tilde{A}_i \longrightarrow \tilde{B}_i$  auf den Mengen der Grundbegriffe  $\tilde{A}_i$  und  $\tilde{B}_i$ ,  $i \in \{1, 2, 3\}$ . Die Tabellen 6.2 zeigen die Wahrheitswerte der Interpretation als negative Regeln, die durch Normierung in die Wahrscheinlichkeiten in den Tabellen 6.3 übergehen. In Tabelle 6.4 schließlich werden die Regeln zusammengefasst, indem für jeden Eintrag der Durchschnitt über alle Regeln gebildet wird.

|               |               |               |               |
|---------------|---------------|---------------|---------------|
|               | $\tilde{B}_1$ | $\tilde{B}_2$ | $\tilde{B}_3$ |
| $\tilde{A}_1$ | 1             | 0             | 0             |
| $\tilde{A}_2$ | 1             | 1             | 1             |
| $\tilde{A}_3$ | 1             | 1             | 1             |

$\tilde{A}_1 \longrightarrow \tilde{B}_1$

|               |               |               |               |
|---------------|---------------|---------------|---------------|
|               | $\tilde{B}_1$ | $\tilde{B}_2$ | $\tilde{B}_3$ |
| $\tilde{A}_1$ | 1             | 1             | 1             |
| $\tilde{A}_2$ | 0             | 1             | 0             |
| $\tilde{A}_3$ | 1             | 1             | 1             |

$\tilde{A}_2 \longrightarrow \tilde{B}_2$

|               |               |               |               |
|---------------|---------------|---------------|---------------|
|               | $\tilde{B}_1$ | $\tilde{B}_2$ | $\tilde{B}_3$ |
| $\tilde{A}_1$ | 1             | 1             | 1             |
| $\tilde{A}_2$ | 1             | 1             | 1             |
| $\tilde{A}_3$ | 0             | 0             | 1             |

$\tilde{A}_3 \longrightarrow \tilde{B}_3$

Tabelle 6.2: Wahrheitswerttabellen der negativen Regeln

|               |               |               |               |
|---------------|---------------|---------------|---------------|
|               | $\tilde{B}_1$ | $\tilde{B}_2$ | $\tilde{B}_3$ |
| $\tilde{A}_1$ | 1             | 0             | 0             |
| $\tilde{A}_2$ | 1/3           | 1/3           | 1/3           |
| $\tilde{A}_3$ | 1/3           | 1/3           | 1/3           |

$\tilde{A}_1 \longrightarrow \tilde{B}_1$

|               |               |               |               |
|---------------|---------------|---------------|---------------|
|               | $\tilde{B}_1$ | $\tilde{B}_2$ | $\tilde{B}_3$ |
| $\tilde{A}_1$ | 1/3           | 1/3           | 1/3           |
| $\tilde{A}_2$ | 0             | 1             | 0             |
| $\tilde{A}_3$ | 1/3           | 1/3           | 1/3           |

$\tilde{A}_2 \longrightarrow \tilde{B}_2$

|               |               |               |               |
|---------------|---------------|---------------|---------------|
|               | $\tilde{B}_1$ | $\tilde{B}_2$ | $\tilde{B}_3$ |
| $\tilde{A}_1$ | 1/3           | 1/3           | 1/3           |
| $\tilde{A}_2$ | 1/3           | 1/3           | 1/3           |
| $\tilde{A}_3$ | 0             | 0             | 1             |

$\tilde{A}_3 \longrightarrow \tilde{B}_3$

Tabelle 6.3: Bedingte Wahrscheinlichkeiten  $P_{\mathcal{P}}^k(\tilde{B}_j | \tilde{A}_i)$  der negativen Regeln

|               | $\tilde{B}_1$ | $\tilde{B}_2$ | $\tilde{B}_3$ |
|---------------|---------------|---------------|---------------|
| $\tilde{A}_1$ | 5/9           | 2/9           | 2/9           |
| $\tilde{A}_2$ | 2/9           | 5/9           | 2/9           |
| $\tilde{A}_3$ | 2/9           | 2/9           | 5/9           |

Tabelle 6.4: Aggregation der negativen Regeln mit Durchschnittsbildung

Die Wahrheitswerte der positiven Interpretation der drei Regeln sehen wir in Tab. 6.5, die gleichzeitig die bedingten Wahrscheinlichkeiten unvollständig spezifiziert. Erst die Aggregation der Regeln in Tab. 6.6 ist eine vollständige Festlegung der bedingten Wahrscheinlichkeiten: die Zeilensummen sind eins.

|               | $\tilde{B}_1$ | $\tilde{B}_2$ | $\tilde{B}_3$ |
|---------------|---------------|---------------|---------------|
| $\tilde{A}_1$ | 1             | 0             | 0             |
| $\tilde{A}_2$ | 0             | 0             | 0             |
| $\tilde{A}_3$ | 0             | 0             | 0             |

$\tilde{A}_1 \longrightarrow \tilde{B}_1$

|               | $\tilde{B}_1$ | $\tilde{B}_2$ | $\tilde{B}_3$ |
|---------------|---------------|---------------|---------------|
| $\tilde{A}_1$ | 0             | 0             | 0             |
| $\tilde{A}_2$ | 0             | 1             | 0             |
| $\tilde{A}_3$ | 0             | 0             | 0             |

$\tilde{A}_2 \longrightarrow \tilde{B}_2$

|               | $\tilde{B}_1$ | $\tilde{B}_2$ | $\tilde{B}_3$ |
|---------------|---------------|---------------|---------------|
| $\tilde{A}_1$ | 0             | 0             | 0             |
| $\tilde{A}_2$ | 0             | 0             | 0             |
| $\tilde{A}_3$ | 0             | 0             | 1             |

$\tilde{A}_3 \longrightarrow \tilde{B}_3$

Tabelle 6.5: Wahrheitstabelle der positiven Regeln, gleichzeitig unvollständige Spezifikation der bedingten Wahrscheinlichkeiten  $P_{\tilde{B}_i}^k(\tilde{B}_j|\tilde{A}_i)$ 

|               | $\tilde{B}_1$ | $\tilde{B}_2$ | $\tilde{B}_3$ |
|---------------|---------------|---------------|---------------|
| $\tilde{A}_1$ | 1             | 0             | 0             |
| $\tilde{A}_2$ | 0             | 1             | 0             |
| $\tilde{A}_3$ | 0             | 0             | 1             |

Tabelle 6.6: Aggregation der positiven Regeln

Wenn wir die Tabellen 6.4 und 6.6 vergleichen, wird deutlich, dass die drei negativen Regeln im Beispiel die bedingten Wahrscheinlichkeiten anders definieren als die entsprechenden positiven Regeln. Die Unsicherheit ist bei den negativen Regeln größer, die Dominanz der Diagonale nicht so stark. Die Ursache dafür liegt darin, dass man in der Wahrscheinlichkeitsrechnung Unwissen nicht explizit darstellen kann. Nehmen wir z. B. an, wir hätten eine Urne mit schwarzen und weißen Kugeln. Wir suchen die Wahrscheinlichkeit, mit der wir eine weiße bzw. eine schwarze Kugel ziehen. Wenn wir wissen, dass die Urne ebenso viele weiße wie schwarze Kugeln enthält, so gilt  $P(\text{weiß}) = P(\text{schwarz}) = 0,5$ . Wenn wir keinerlei Informationen über die Farbe der Kugeln haben, wählen wir aus Unwissen die Wahrscheinlichkeiten genauso.<sup>6</sup> Wir können einer Gleichverteilung nicht

<sup>6</sup>insufficient reason principle

ansehen, ob sie aus vollständigem Wissen oder aus Unwissen abgeleitet ist. Bei der Aggregation von Regeln ist dies aber eine wichtige Information. Wenn wir nämlich Regeln wie in Gleichung (6.12) über Durchschnittsbildung aggregieren, dann nehmen wir an, dass die Vertrauenswürdigkeit aller Regeln gleich ist. Wurde aber eine Gleichverteilung aus Unwissen abgeleitet, so haben die zugehörigen Wahrscheinlichkeiten keinerlei Bedeutung im Vergleich mit Wahrscheinlichkeiten, die aus vollständigem Wissen resultieren. In diesem Fall würde man den Wert, der aus Unwissen abgeleitet wurde, nicht für eine Aggregation verwenden.

Wenn wir eine negative Regel wie in Tab. 6.3 formulieren, dann müssen wir festlegen, ob die gleichverteilten Zeileneinträge Unwissen repräsentieren sollen – wir wissen nicht, welches Konsequens richtig ist – oder ob sie tatsächlich bedeuten, dass alle Konsequenzen mit gleicher Wahrscheinlichkeit auftreten können. Wählen wir die zweite Interpretation, dann benutzen wir das oben beschriebene Verfahren für die Aggregation. Falls aber die erste Interpretation gewünscht ist, dann modifizieren wir das Verfahren folgendermaßen: Jeder gleichverteilten Zeile  $(\gamma_{ij}^k)_j$  einer Wahrheitstabelle ordnen wir einen Indikator zu, der festhält, ob die Gleichverteilung vollständiges Wissen oder Unwissen darstellt. Eine Zeile, die vollständiges Wissen repräsentiert, wird unverändert für die Aggregation benutzt. Wenn eine gleichverteilte Zeile aus Unwissen abgeleitet ist, unterscheiden wir zwei Fälle. Im ersten Fall gibt es andere Regeln, die diese Zeile mit Wahrscheinlichkeiten belegen. Wir berücksichtigen dann die gleichverteilte Zeile nicht bei der Aggregation. Der zweite Fall liegt dann vor, wenn die anderen Regeln die Zeile entweder ebenfalls aus Unwissen gleichverteilt definiert haben oder alle Einträge der Zeile null sind. In diesem Fall ist die Gleichverteilung das Ergebnis der Aggregation. Wenn wir dieses Verfahren auf die negativen Regeln in Beispiel 1 anwenden, erhalten wir die bedingten Wahrscheinlichkeiten der positiven Regeln in Tab. 6.6.

Eine Regelbasis resultiert in unserem Ansatz somit immer in der Definition der bedingten Wahrscheinlichkeiten  $P_{\mathcal{P}}(\tilde{B}_j|\tilde{A}_i)$ , unabhängig davon, ob die Regeln positiv oder negativ sind, oder bedingte Wahrscheinlichkeiten teilweise durch andere Quellen wie Daten definiert werden. Das Inferenzverfahren ist ebenso unabhängig von der Quelle der bedingten Wahrscheinlichkeiten. Das Vorgehen ist vergleichbar mit dem Prinzip von Kombination/Projektion in der Possibilitätstheorie: Alle Zusammenhänge oder Regeln werden in einer unscharfen Relation aggregiert, daraus wird ein Schluss abgeleitet. Es gibt aber zwei wesentliche Unterschiede zu unserem Ansatz. Erstens ist die unscharfe Relation in der Possibilitätstheorie auf der Ebene der Einzelheiten definiert, während sich die bedingten Wahrscheinlichkeiten unserer Modellierung auf der grobgranularen Ebene der Begriffe befinden. Damit wird im Ansatz dieser Arbeit einerseits die Verständlichkeit der Zusammenhänge zwischen Variablen sichergestellt, andererseits aber auch der Rechen- und Speicheraufwand drastisch reduziert. Zweitens können wir positive

und negative Regeln in einer Relation vereinen und so gemeinsam verarbeiten. Der Inferenzvorgang wird dadurch transparenter und der Rechenaufwand ist bei beiden Regelarten gleich gering. In der Possibilitätstheorie dagegen haben wir in Abschnitt 5.5 gesehen, dass Inferenzen mit negativen Regeln (Prinzip von Kombination/Projektion) wesentlich aufwändiger sind als mit positiven Regeln (Ansatz von MAMDANI).

Insgesamt haben wir gezeigt, wie man positive und negative Regeln auf der Basis unserer Modellierung mit bedingten Wahrscheinlichkeiten darstellt. Damit werden wir einer bewährten Methode aus der Praxis gerecht, Regeln zu formulieren, ohne dass wir auf die Verständlichkeit der Schlüsse und die Effizienz unseres Verfahrens verzichten müssten.

### 6.3.2 Unschärfe Interpolation

Besonders auf dem Gebiet unscharfer Regelung hat sich gezeigt, dass unscharfe, regelbasierte Systeme dazu benutzt werden können, scharfe Abbildungen zu modellieren. Die Aufgabe besteht darin, eine a priori unbekannte Funktion wie eine Regelstrategie oder die Übertragungsfunktion eines technischen Prozesses zu approximieren. Eine besondere Rolle spielen in diesem Zusammenhang Interpolationsverfahren, weil in vielen Anwendungen Daten vorhanden sind, die die unbekannte Funktion partiell beschreiben. Ein Beispiel ist die Simulation eines technischen Prozesses. Wenn man das physikalische Modell des Prozesses nicht kennt oder die Herleitung zu aufwändig ist, kann man den Prozess im Betrieb beobachten und sein Ein-/Ausgabeverhalten stichprobenhaft in Daten festhalten. Das mathematische Simulationsmodell soll die Daten wiedergeben und für Eingaben, die nicht im Datensatz vorhanden sind, sinnvolle Ausgaben berechnen – es soll zwischen den Daten sinnvoll interpolieren. Sinnvoll heißt, dass die interpolierende Funktion gewünschte analytische Eigenschaften wie Stetigkeit, Linearität u. Ä. erfüllt.

Im Bereich unscharfer Informationsverarbeitung werden für solche Aufgaben der Regler von TAKAGI/SUGENO [TS84] und Neuro-Fuzzy-Ansätze [Jan92, WM92, Jan93, Wan94, BH94, BH95, WM92] benutzt. Dabei handelt es sich um scharfe Funktionsapproximatoren oder Interpolatoren, die nur mit viel Wohlwollen als unscharfe, regelbasierte Ansätze interpretiert werden können. In den letzten Jahren werden zunehmend Verfahren vorgestellt, die unscharfe Interpolation propagieren [KH93, SH98, SK00]. Sie interpolieren zwischen unscharfen Daten wie unsicheren Messungen oder vagen Angaben von Experten. Indem sie bewusst die Unschärfe modellieren, kann man über ein Maß der Unschärfe die Güte der Funktionsapproximation ablesen [SK00].

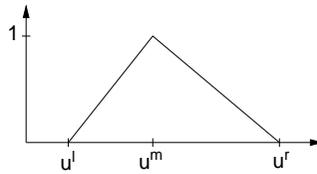


Abbildung 6.2: Dreieckfunktion

In diesem Abschnitt zeigen wir, dass wir unser Inferenzverfahren ebenfalls als unscharfes Interpolationsverfahren interpretieren können, das im Gegensatz zu den anderen Verfahren aber semantisch fundiert ist. Wir beschränken uns in der Darstellung auf abgeschlossene Intervalle auf  $\mathbb{R}$  als Universen  $\mathcal{X}, \mathcal{Y}$ . Formal sieht die scharfe Interpolationsaufgabe wie folgt aus: Gegeben sei eine Menge von Punkten  $(u_i, v_i) \in \mathcal{X} \times \mathcal{Y}$ ,  $i \in \{1, 2, \dots, k\}$ . Gesucht ist eine Funktion  $f: \mathcal{X} \rightarrow \mathcal{Y}$  mit der **Interpolationseigenschaft**  $f(u_i) = v_i$  für alle  $i$ . Welches mathematische Modell für die gesuchte Funktion  $f$  gewählt wird, hängt von den gewünschten analytischen Eigenschaften von  $f$  ab. Wir betrachten in dieser Arbeit Linearkombinationen von Basisfunktionen  $A_i: \mathcal{X} \rightarrow \mathcal{Y}$ :

$$f(u) := \sum_i \alpha_i \cdot A_i(u) \quad (6.13)$$

Die Wahl der Basisfunktionen  $A_i$  bestimmt die analytischen Eigenschaften von  $f$ . Wir widmen uns in diesem Abschnitt linearer Interpolation:  $f$  ist ein Polygonzug mit den Eckpunkten  $(u_i, v_i)$ . Um eine lineare Interpolation zu erreichen, wählt man Dreiecksfunktionen wie in Abb. 6.2 als Basisfunktionen. Eine **Dreieckfunktion** wird durch das Tripel  $(u^l, u^m, u^r)$  mit  $u^l \leq u^m \leq u^r$  festgelegt:

$$A(u) := \begin{cases} 0, & \text{wenn } u < u^l \\ \frac{u - u^l}{u^m - u^l}, & \text{wenn } u^l \leq u \leq u^m \\ 1 - \frac{u - u^m}{u^r - u^m}, & \text{wenn } u^m < u \leq u^r \\ 0, & \text{wenn } u > u^r \end{cases} \quad (6.14)$$

Insbesondere gelten damit  $A(u^l) = A(u^r) = 0$  und  $A(u^m) = 1$ . Die Menge der Basisfunktionen  $\{A_1, A_2, \dots, A_k\}$  wird durch die Stützstellen  $u_0, u_1, u_2, \dots, u_k, u_{k+1}$  mit  $u_0 := u_1$  und  $u_{k+1} := u_k$  definiert. Dabei nehmen wir an, dass die Stützstellen bzgl. der Relation „ $\leq$ “ geordnet sind. Für die definierenden Tripel der Funktionen  $A_i$  gilt dann  $(u_i^l, u_i^m, u_i^r) = (u_{i-1}, u_i, u_{i+1})$ . Die Stützstellen  $u_1$  und  $u_k$  werden dupliziert, weil sie den Rand des Interpolationsbereiches darstellen. Auf diese Weise

entstehen entartete Dreieckfunktionen, wie Abb. 6.3 verdeutlicht. Unter der Annahme dieser Definitionen interpoliert die Linearkombination  $\sum_{i=1}^k v_i \cdot A_i(u)$  von Dreieckfunktionen linear zwischen den  $k$  Punkten  $(u_i, v_i)$  [SB94].

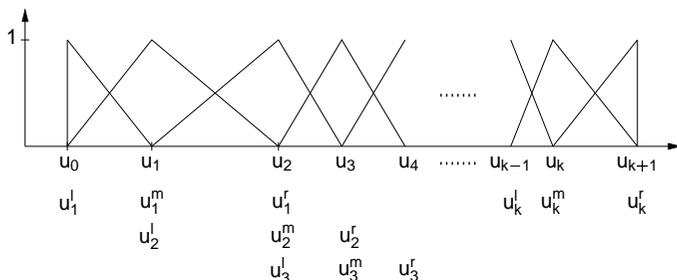


Abbildung 6.3: Familie von Dreieckfunktionen als Basisfunktionen

Hinter dieser Art von Interpolation steckt der Wunsch nach einer strukturerhaltenden Interpolierenden  $f$ , einem Homomorphismus. Ein interpolierter Wert  $v = f(u)$  wird so berechnet, dass  $v$  in der gleichen Beziehung zu den  $v_i$  steht wie  $u$  zu den  $u_i$ :

$$f(u) = \sum_{i=1}^k A_i(u) \cdot v_i \quad \text{mit} \quad u = \sum_{i=1}^k A_i(u) \cdot u_i \quad (6.15)$$

$A_i(u)$  misst die Proportion, mit der  $u \in [u_i, u_{i+1}]$  zwischen  $u_i$  und  $u_{i+1}$  liegt und überträgt sie auf  $v \in [v_i, v_{i+1}]$ . Für polynomielle Interpolationen höheren Grades kann man entsprechende Beziehungen angeben [dB78].

Wir formulieren die Interpolationsaufgabe nun als Regelwerk, um den Zusammenhang von Interpolation und regelbasierten Systemen zu zeigen. Das Wissen über die Relation der Universen  $\mathcal{X}$  und  $\mathcal{Y}$  ist durch die Interpolationseigenschaft als Satz von Regeln „Wenn  $x = u_i$ , dann  $y = v_i$ “ gegeben. Die Interpolation erreichen wir durch die Vorschrift „Wenn  $x \approx u_i$ , dann  $y \approx v_i$ “, die für alle  $i$  erfüllt sein soll. Dabei ist der Operator  $\approx$  ein Maß für die Ähnlichkeit zweier Zahlen. Im Fall linearer Interpolation misst  $A_i(x)$  die Ähnlichkeit von  $x$  und  $u_i$ . Dementsprechend werden die analytischen Eigenschaften der Interpolierenden durch die Ähnlichkeitsmaße bestimmt.

Diese Interpretation von Interpolation als regelbasiertes System erweitern wir in dieser Arbeit insofern, als wir zwischen unscharfen Mengen  $(\tilde{A}_i, \tilde{B}_i)$  interpolieren. Dabei sind die  $\tilde{A}_i, \tilde{B}_i$  unscharfe Grundbegriffe. Wir suchen also eine Abbildung  $f: \mathcal{UM}(\mathcal{X}) \rightarrow \mathcal{UM}(\mathcal{Y})$ , welche die Interpolationseigenschaft  $f(\tilde{A}_i) = \tilde{B}_i$  für

alle  $i$  erfüllt. Daraus resultieren die Regeln  $\tilde{A}_i \longrightarrow \tilde{B}_i$ . Der Einfachheit halber verzichten wir hier auf die allgemeinere Darstellung mit gewichteten Regeln, die wir in Kapitel 5 benutzt haben.

Eine unscharfe Eingabe  $\tilde{A}$  wird gemäß (4.19) durch die unscharfen Mengen  $\tilde{A}_i$  dargestellt. Im Gegensatz zu (6.15) ist die Darstellung i. A. nur eine Approximation:

$$pl_{\tilde{A}}(u) \approx \sum_{i=1}^k P_{\mathcal{P}}(\tilde{A}_i | \tilde{A}) \cdot pl_{\tilde{A}_i}(u) \quad (6.16)$$

Den Schluss  $\tilde{B} = f(\tilde{A})$  berechnen wir mit unserem Inferenzmechanismus aus Abschnitt 5.2 gemäß der Vorschriften (5.3) und (5.4):

$$pl_{\tilde{B}}(v) = \sum_{i=1}^k P_{\mathcal{P}}(\tilde{A}_i | \tilde{A}) \cdot pl_{\tilde{B}_i}(v) \quad (6.17)$$

Dies entspricht dem Vorgehen in (6.15). Nach Satz 8 wird die Interpolationseigenschaft  $f(\tilde{A}_i) = \tilde{B}_i$  erfüllt, die nichts anderes ist als der Modus Ponens. Unser Inferenzmechanismus kann somit als unscharfes Interpolationsverfahren gedeutet werden.

In den folgenden Abschnitten zeigen wir, dass sich in diese Interpretation einige häufig genutzte Ansätze einbetten lassen, die im Bereich unscharfer Regelung oder allgemeiner zur Funktionsapproximation eingesetzt werden.

**Der Regler von TAKAGI/SUGENO und Neuro-Fuzzy-Systeme** Wie bereits in Abschnitt 2.2 beschrieben, formuliert man beim Ansatz von TAKAGI/SUGENO eine Funktion als Satz von Regeln der Art „Wenn  $x = \tilde{A}_i$ , dann  $y = f_i(u)$ “. Der für die Praxis relevante Spezialfall reduziert die reellwertigen Funktionen  $f_i : \mathcal{X} \longrightarrow \mathcal{Y}$  auf Konstante  $v_i \in \mathcal{Y}$  und damit die Regeln auf  $\tilde{A}_i \longrightarrow v_i$ . Der Regler wird schließlich durch die Funktion  $f$  dargestellt:

$$y = f(x) = \frac{\sum_{i=1}^k \mu_{\tilde{A}_i}(x) \cdot v_i}{\sum_{i=1}^k \mu_{\tilde{A}_i}(x)} \quad (6.18)$$

Der Nenner entfällt, wenn die unscharfen Mengen  $\tilde{A}_i$  die Überdeckungseigenschaft (4.3) erfüllen. Gleichung (6.18) entspricht dann dem klassischen Interpolationsverfahren aus Gleichung (6.13).

Wir zeigen nun, dass der Regler von TAKAGI/SUGENO mit konstanten Konsequenzen ein Spezialfall unseres Inferenzverfahrens ist. Die Antezedenzen  $\tilde{A}_i$  werden in beiden Fällen durch unscharfe Mengen dargestellt. Die Konsequenzen  $v_i$  sind als scharfe Werte spezielle unscharfe Mengen bzw. Plausibilitätsdichten:

$pl_{\tilde{B}_i}(v) = \mathbb{1}|_{v=v_i}$ . Wenn die Eingabe  $\tilde{A}$  in unser System ein scharfer Wert  $x$  ist, so bestimmen wir die Wahrscheinlichkeiten  $P_{\mathcal{P}}(\tilde{A}_i|\tilde{A})$  aus Gleichung (6.17) nach Abschnitt 6.2 mit

$$P_{\mathcal{P}}(\tilde{A}_i|\tilde{A}) = P_{\mathcal{P}}(\tilde{A}_i|x) = pl_{\tilde{A}_i}(x) = \mu_{\tilde{A}_i}(x) \quad (6.19)$$

und Gleichung (6.17) geht über in

$$pl_{\tilde{B}}(x) = \sum_{i=1}^k \mu_{\tilde{A}_i}(x) \cdot pl_{\tilde{B}_i}(x) = \sum_{i=1}^k \mu_{\tilde{A}_i}(x) \cdot \mathbb{1}|_{x=v_i} \quad (6.20)$$

Gleichung (6.20) ist eine unscharfe Version von (6.18): In (6.20) werden die scharfen Konklusionen  $v_i$  aus (6.18) durch  $\mathbb{1}|_{x=v_i}$  unscharf repräsentiert; dementsprechend ist auch der Schluss  $\tilde{B}$  unscharf. Eine Schärfung nach der Schwerpunktmethode überführt aber (6.20) in (6.18).

Damit sind die Inferenzverfahren im Fall scharfer Konsequenzen und scharfer Beobachtungen  $x$  identisch. Dies gilt ebenso für klassische Neuro-Fuzzy-Systeme wie ANFIS [Jan92, Jan93], *Fuzzy Basis Functions* [WM92, ZS94, KM95], Spline-basierte Systeme [BH94, BH95, ZK96, ZK99] und *radiale Basisfunktionen (RBF)* [Pow87, PG90, HHG96], die dem Ansatz von TAKAGI/SUGENO mit scharfen Konsequenzen entsprechen. *Additive unscharfe Systeme* [Kos92a, Kos92b, Kos94, DK96] benutzen ebenfalls das gleiche Inferenzverfahren, die Konsequenzen sind in diesem Fall aber wie bei uns unscharfe Mengen.

Die Deutung dieser Systeme als Ansätze für unscharfes Schließen ist bisher nur sehr vage, s. [JS93], weil sie über keine konkrete Semantik verfügen. Insbesondere sind dadurch die Inferenzverfahren rein intuitiv motiviert. Indem wir sie in unseren Ansatz zum unscharfen Schließen einbetten, können wir sie als regelbasierte Systeme auf der Basis einer konkreten Semantik interpretieren und für mehrstufiges unscharfes Schließen verallgemeinern. Neben dem reinen Erkenntnisgewinn ist dies ein wichtiger Fortschritt, weil wir konkrete Verfahren für den Entwurf solcher Systeme angeben können. Dazu gehört insbesondere die Definition unscharfer Begriffe. Bisher war in den meisten Fällen nicht klar, wie man in der Praxis zu den unscharfen Mengen als Antezedenzien kommt, da keinerlei Semantik definiert wurde. Umgekehrt stehen für unseren Ansatz nun Verfahren zur Verfügung, mit denen Regeln an Daten angepasst werden. Meist wird Gradientenabstieg verwendet, um die Regelgewichte zu optimieren, was sich in der Praxis als sehr erfolgreich erwiesen hat [Jan92, ZK99].

## 6.4 Parametermodell für Grundbegriffe

Die Implementierung der vorgestellten Verfahren dieser Arbeit ist im Großen und Ganzen unproblematisch. Auf der einen Seite müssen wir die Wissensbasis spei-

chern, die aus den Grundbegriffen und Regeln besteht. Andererseits benötigen wir eine Implementierung des Inferenzverfahrens, d. h. der Analyse unscharfer Mengen und der Inferenz selbst. Dabei haben wir uns zum Ziel gesetzt, dass sowohl der Rechenaufwand als auch der Speicheraufwand nur von der Anzahl der Grundbegriffe oder Regeln abhängen sollen, nicht aber von der Anzahl der Einzelheiten. Für den Rechenaufwand haben wir die Anforderung erfüllt, wie wir in Abschnitt 5.5 gezeigt haben. Der Speicheraufwand hingegen kann unangenehm groß werden. Das Problem liegt in der Darstellung der Grundbegriffe. Wir beschreiben daher in diesem Abschnitt, wie Grundbegriffe speichereffizient repräsentiert werden können, ohne die Überdeckungseigenschaft (4.3) zu verletzen.

### 6.4.1 Approximation von Grundbegriffen

Grundbegriffe werden durch Plausibilitätsdichten, d. h. Abbildungen  $\mathcal{X} \rightarrow [0, 1]$  definiert. Wir unterscheiden dabei zwei Fälle. Im ersten Fall wird die Abbildung nach Abschnitt 2.3 durch die vagen Charakteristika  $\Gamma(c)$  definiert. Für jeden Kontext  $c$  müssen wir eine scharfe Teilmenge von  $\mathcal{X}$  speichern. Stammen die vagen Charakteristika aus statistischen Erhebungen, dann kann die Anzahl der Kontexte sehr groß sein. Man denke nur an die Anzahl der Kunden einer Telefongesellschaft. Für jeden Grundbegriff hängt der Speicherbedarf von der Anzahl der Kontexte ab; im Beispiel von der Anzahl der Kunden. Im zweiten Fall liegen die Plausibilitätsdichten direkt als Abbildungen  $\mathcal{X} \rightarrow [0, 1]$  vor. Dieser Fall tritt z. B. ein, wenn wir die Plausibilitätsdichten mit dem Gruppierungsverfahren aus Abschnitt 6.1 berechnen. Sie werden dann durch ein Parametermodell dargestellt, das auf der Normalverteilung beruht. Der Speicherbedarf reduziert sich damit auf einen Parametersatz pro Grundbegriff. Diese Idee können wir aufgreifen, um die Plausibilitätsdichten aus dem ersten Fall darzustellen.

Statt die vagen Charakteristika zu speichern, ersetzen wir die Plausibilitätsdichte durch eine parametrisierte Funktion. Im Allgemeinen erhalten wir nur eine Approximation der Plausibilitätsdichte, was aber in vielen Fällen keine einschneidende Beschränkung ist. Zusätzlich zu der Speicherersparnis wird der Berechnungsaufwand reduziert, den wir für die Auswertung einer Plausibilitätsdichte benötigen. Ebenso wie der Speicheraufwand hängt dieser von der Anzahl der Kontexte ab. Bei einem Parametermodell müssen wir dagegen nur eine parametrisierte Funktion auswerten.

Da Plausibilitätsdichten Funktionen sind, lösen wir das Problem mit Techniken zur Funktionsapproximation. Wie in Abschnitt 6.3.2 betrachten wir als Approximatoren Linearkombinationen von Basisfunktionen. Ausgehend von einer Menge von Basisfunktionen  $\{A_1, A_2, \dots, A_n\}$  approximiert man eine Funktion  $pl$  durch  $pl(u) \approx \sum_i \alpha_i \cdot A_i(u)$ . Der Vorgang der Funktionsapproximation setzt sich aus zwei

Schritten zusammen: Zunächst wählt man adäquate Basisfunktionen, anschließend müssen mit einem Optimierungsverfahren die Parameter  $\alpha_i$  bestimmt werden. Die Wahl der Basisfunktionen beruht auf Wissen über die zu approximierende Funktion; die Numerik stellt eine Reihe von Basisfunktionen mit verschiedenen Eigenschaften zur Verfügung.

Bei der Optimierung der Parameter  $\alpha_i$  müssen wir in unserem Fall noch eine Nebenbedingung erfüllen: Die Approximationen der Grundbegriffe müssen die Überdeckungseigenschaft (4.3) einhalten; wie die Grundbegriffe selbst müssen sie eine unscharfe Partition des Universums bilden. Wir können also die Plausibilitätsdichten der Grundbegriffe nicht unabhängig voneinander approximieren.

Im Idealfall können wir jede Plausibilitätsdichte mit einer Basisfunktion ausreichend gut annähern. Falls die Approximationsgüte nicht ausreicht, benutzen wir mehrere Basisfunktionen für die Plausibilitätsdichte jedes Grundbegriffs  $\tilde{B}_i$ :

$$p_{\tilde{B}_i}(u) \approx \sum_{j=1}^n \alpha_{ij} \cdot A_j(u) \quad (6.21)$$

Um die Überdeckungseigenschaft sicherzustellen, genügt es, die beiden folgenden Bedingungen (6.22) und (6.23) zu erfüllen: Im Grenzfall, in dem jede Plausibilitätsdichte durch genau eine Basisfunktion  $A_i$  dargestellt wird, müssen wir genau die Überdeckungseigenschaft fordern; daher verlangen wir allgemein für alle  $u \in \mathcal{X}$

$$\sum_{j=1}^n A_j(u) = 1 \quad (6.22)$$

Unter dieser Voraussetzung ist dann folgende Bedingung für die Überdeckungseigenschaft hinreichend (vgl. Beweis von Satz 11):

$$\sum_{i=1}^k \alpha_{ij} = 1 \quad (6.23)$$

Bedingung (6.22) wird z. B. von *B-Splines* (Basis-Splines) erfüllt [dB78], die wegen ihrer guten Approximationseigenschaften weite Verbreitung in CAD-Anwendungen gefunden haben [HL92, Far93]. B-Splines sind stückweise definierte Polynome mit glatten Übergängen an den Stoßpunkten. Der Grad eines B-Splines wird durch den Grad der definierenden Polynome festgelegt. Wir werden uns in dieser Arbeit auf lineare B-Splines als Basisfunktionen beschränken, weil

- sie sehr einfach zu berechnen sind,
- sich in praktischen Anwendungen gezeigt hat, dass die Approximationsgüte ausreicht und

- die Parameter  $\alpha_{ij}$  leicht zu bestimmen sind, wobei Bedingung (6.23) erfüllt wird.

Ein linearer B-Spline ist eine *Dreieckfunktion*, die wir mit Formel (6.14) bereits in Abschnitt 6.3.2 definiert haben. Die Basisfunktionen  $A_j$  legen wir wie in 6.3.2 über Stützstellen  $u_j \in \mathcal{X}$  fest; statt von  $k$  gehen wir hier von  $n$  Basisfunktionen aus. Die Linearkombination  $\sum_{j=1}^n \alpha_{ij} \cdot A_j(u)$  von Dreieckfunktionen  $A_j$  interpoliert dann linear zwischen den Punkten  $(u_j, \alpha_{ij})$ ,  $j \in \{1, 2, \dots, n\}$ . Die Plausibilitätsdichten werden auf diese Weise durch Polygonzüge approximiert.

Wenn wir von einer Partitionierung von  $\mathcal{X}$  durch  $k$  Grundbegriffe  $\mathcal{P} = \{\tilde{B}_1, \tilde{B}_2, \dots, \tilde{B}_k\}$  ausgehen, die durch  $n$  Basisfunktionen  $A_j$  approximiert werden sollen, lautet das Verfahren für die Bestimmung der Gewichte  $\alpha_{ij}$  wie folgt:

### Verfahren für die Bestimmung der Parameter

1. Wähle  $n$  Stützpunkte  $u_1, u_2, \dots \in \mathcal{X}$  mit  $n \geq k$ .
2. Füge die Stützpunkte  $u_0 := u_1$  und  $u_{n+1} := u_n$  hinzu.
3. Definiere die Funktionen  $A_j$  durch  $(u_j^l, u_j^n, u_j^r) = (u_{j-1}, u_j, u_{j+1})$  als Dreieckfunktionen nach Formel (6.14).
4. Setze  $\alpha_{ij} := pl_{\tilde{B}_i}(u_j)$ .

Mit den Definitionen des Verfahrens gilt

**Satz 11** Die Funktionen  $pl_{\tilde{B}_i}^l(u) = \sum_{j=1}^n pl_{\tilde{B}_i}(u_j) \cdot A_j(u)$  bilden eine unscharfe Partition von  $\mathcal{X}$  und approximieren die Plausibilitätsdichten  $pl_{\tilde{B}_i}$  durch lineare Interpolation zwischen den Punkten  $(u_j, pl_{\tilde{B}_i}(u_j))$ ,  $j \in \{1, 2, \dots, n\}$ .

*Beweis:* Aus der Numerik wissen wir, dass die Interpolationsbehauptung richtig ist [SB94]. Um zu beweisen, dass die Funktionen  $pl_{\tilde{B}_i}^l$  eine unscharfe Partition von  $\mathcal{X}$  bilden, zeigen wir die Gültigkeit der Überdeckungseigenschaft (4.3). Sei  $u \in \mathcal{X}$ .

$$\begin{aligned}
 \sum_{i=1}^k pl_{\tilde{B}_i}^l(u) &= \sum_{i=1}^k \sum_{j=1}^n pl_{\tilde{B}_i}(u_j) \cdot A_j(u) \\
 &= \sum_{j=1}^n A_j(u) \sum_{i=1}^k pl_{\tilde{B}_i}(u_j) \\
 &= \sum_{j=1}^n A_j(u) \\
 &= 1
 \end{aligned}$$

□

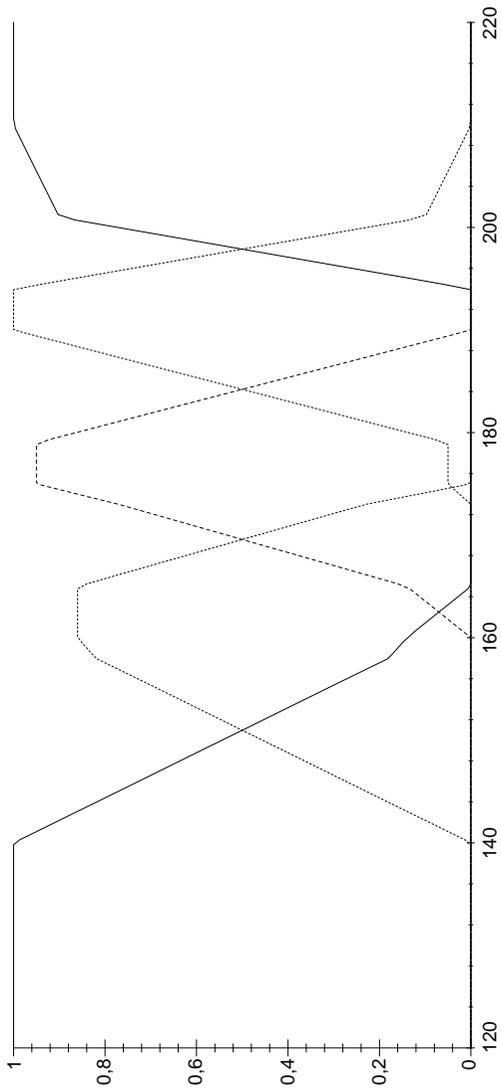


Abbildung 6.4: Approximationen der unscharfen Körpergrößen erwachsener Männer aus Abb. 4.1: *sehr klein*, *klein*, *normal*, *groß* und *sehr groß* (von links nach rechts)

Abbildung 6.4 zeigt die Approximation der Grundbegriffe *sehr klein*, *klein*, *normal*, *groß* und *sehr groß* aus Abschnitt 4.1 durch 13 Dreieckfunktionen. Wir erkennen deutlich die Glättung der ursprünglichen Plausibilitätsdichten durch die stückweise lineare Approximation, vgl. Abb. 4.1.

Ein allgemeines Problem bei Funktionsapproximation durch Interpolation ist die Wahl der Stützstellen, also Punkt 1 unseres Verfahrens. Die Freiheitsgrade sind die Anzahl und die Lage der Stützstellen. Bei einer überschaubaren Anzahl von Grundbegriffen wird man beides von Hand wählen. Die Approximationsgüte wird potenziell erhöht, wenn man die Anzahl der Stützstellen vergrößert. Damit wächst aber auch der Speicheraufwand. Man muss also eine vernünftige Balance zwischen diesen Größen finden. Wenn viele Grundbegriffe approximiert werden müssen, kann man für die Wahl der Stützstellen auf Optimierungsansätze zurückgreifen [ZK99, GRL00].

Die Approximation der Grundbegriffe durch Linearkombinationen von Basisfunktionen können wir als Verfeinerung der Partition  $\mathcal{P}$  von  $\mathcal{X}$  interpretieren. Da die Basisfunktionen  $A_j$  mit Gleichung (6.22) die Überdeckungseigenschaft erfüllen, können wir die  $A_j$  als Plausibilitätsdichten von Grundbegriffen  $\tilde{A}_j$  auffassen. Die Approximationen der Grundbegriffe  $\tilde{B}_i$  sind dann nichts anderes als gewichtete Kombinationen der Grundbegriffe  $\tilde{A}_j$ , wobei wir die Gewichte  $\alpha_{ij}$  als Begriffswahrscheinlichkeiten  $P_{\mathcal{P}}(\tilde{A}_j|\tilde{B}_i)$  deuten können. Im Grunde genommen haben wir eine Transformation von den Grundbegriffen  $\tilde{B}_i$  auf die  $\tilde{A}_j$  berechnet, wie wir es bereits in Abschnitt 5.2.2 bei mehrstufigem Schließen durchgeführt haben. Da die Anzahl der Grundbegriffe  $\tilde{A}_j$  größer oder gleich der Anzahl der  $\tilde{B}_i$  ist und die  $\tilde{A}_j$  so gewählt wurden, dass sie die  $\tilde{B}_i$  möglichst gut approximieren, stellt die Transformation eine Verfeinerung dar, gleichbedeutend einem Schritt auf eine Stufe feinerer Granularität. Das Parametermodell fügt sich auf diese Weise perfekt in unser Verständnis von der Darstellung unscharfer Informationen ein. Es ist nichts anderes als eine gewichtete Kombination speziell gewählter Grundbegriffe.

Die approximierten Grundbegriffe werden genauso verarbeitet wie die Originale, weder die Kombination von Grundbegriffen noch das Inferenzverfahren ändern sich. Als Beispiel haben wir die unscharfe Messung der Körpergröße von Herrn G. aus Abb. 4.5 zweimal nach Analyse-Ansatz 2 approximiert: einmal mit den Originalgrundbegriffen und einmal mit den approximierten Grundbegriffen. Abbildung 6.5 zeigt die Ergebnisse im Vergleich. Wir sehen, dass die Originalkombination gut durch die Approximation angenähert wird. Während die Originalgrundbegriffe *normal* und *groß* durch die Analyse mit den Gewichten 41,4% und 58,6% versehen werden, ergeben sich für die approximierten Grundbegriffe 41,7% und 58,3%. Die Abweichungen sind vernachlässigbar. Zur Erinnerung: die Umfrage bei der Synthese der Körpergröße von Herrn G. resultierte in 41% und 59%.

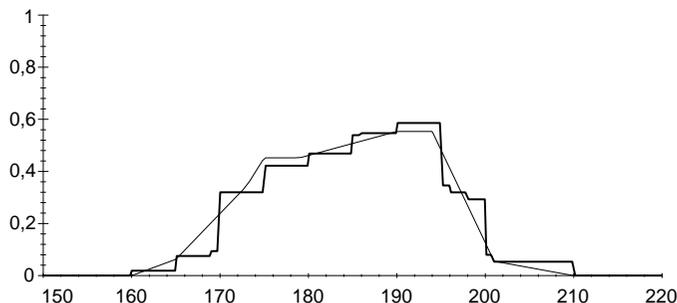


Abbildung 6.5: Approximation der Körpergröße von Herrn G. durch Kombination der Grundbegriffe (fett) und Kombination der approximierten Grundbegriffe (dünn)

## 6.4.2 Speicheraufwand

Der wesentliche Zweck der Approximation von Grundbegriffen besteht darin, den Speicheraufwand für die Repräsentation von Plausibilitätsdichten zu reduzieren. Wenn wir für jeden Grundbegriff die vagen Charakteristika speichern, so benötigen wir bei  $k$  Grundbegriffen und  $l$  vagen Charakteristika  $O(l \cdot k)$  Speicherplätze.

Approximieren wir die Grundbegriffe durch Basisfunktionen wie die Dreieckfunktionen, so müssen wir  $n \cdot k$  Gewichte  $\alpha_{ij}$  speichern. Wenn wir davon ausgehen, dass jeder Grundbegriff durch eine kleine Anzahl  $c$  von Basisfunktionen angenähert wird (typisch eins bis fünf), beträgt der Speicheraufwand  $O(c \cdot k^2) = O(k^2)$ .

Insgesamt können wir also mit einer Reduzierung des Speicherbedarfs von Faktor  $l/(c \cdot k)$  rechnen. Wie das Beispiel der Kunden einer Telefongesellschaft zeigt, kann  $l$  durchaus im Bereich von  $10^6$  liegen, während sich  $c \cdot k$  eher im Bereich unter 100 bewegt. Das Parametermodell würde in diesem Beispiel den Speicherbedarf um den Faktor  $10^4$  verbessern.

Der Aufwand für die Auswertung der Plausibilitätsdichte eines Grundbegriffs an einer Stelle wird ebenfalls reduziert: statt  $O(l)$  benötigen wir nur  $O(1)$  Operationen. Während wir für die Berechnung über das Kontextmodell eine Summe über  $l$  Kontexte bilden (vgl. (2.8) und (2.9)), werten wir bei der Approximation  $c$  Basisfunktionen mit konstantem Aufwand aus und addieren die Ergebnisse. Da  $c$  unabhängig von der Anzahl der Grundbegriffe ist, können wir den Rechenaufwand als konstant betrachten.

## 6.5 Zusammenfassung

In diesem Kapitel haben wir uns mit zwei Aspekten der Repräsentation unscharfen Wissens beschäftigt: erstens der mathematischen Formalisierung von Wissen, die von der Semantik des Wissens abhängt, und zweitens der implementierungstechnischen Seite der Repräsentation, die praktische Anforderungen hinsichtlich des Speicher- und Rechenaufwands erfüllen muss.

Wir haben mögliche Semantiken von Wissen klassifiziert und aus diesen Klassen Formen von Wissen herausgesucht, die sich in der Literatur für den praktischen Entwurf und den Einsatz regelbasierter unscharfer Systeme als sinnvoll erwiesen haben. Bei der Formalisierung unscharfer Begriffe handelt es sich um folgende Semantiken:

- Interpretation von Unschärfe als Unsicherheit: Diesen Ansatz haben wir bereits in Abschnitt 2.3.1 und in Kapitel 4 als Basis unserer Theorie ausführlich dargelegt.
- Interpretation von Unschärfe als Ähnlichkeit: Wir haben gezeigt, dass das weit verbreitete Gruppierungsverfahren von GATH/GEVA [GG89] direkt eine unscharfe Partition im Sinne dieser Arbeit erzeugt; unscharfe Grundbegriffe können wir damit als Gruppen ähnlicher Daten interpretieren. Insbesondere können wir unscharfe Partitionen direkt aus scharfen Daten generieren, anstatt wie in Abschnitt 4.1 eine unscharfe Partition als feingranulare Definition symbolischer Begriffe zu bilden.
- Darstellung scharfer Werte: Wir stumpfen scharfe Werte ab, indem wir sie durch die Grundbegriffe unscharf klassifizieren. Auf diese Weise können wir scharfe Werte semantisch fundiert und effizient mit grobgranularem Wissen verknüpfen.
- Die Darstellung von Unschärfe zweiter Ordnung haben wir in Kapitel 4 mit der Gewichtung unscharfer Begriffe diskutiert.

Bei möglichen Semantiken einer unscharfen Regel  $\tilde{A} \longrightarrow \tilde{B}$  gehen wir davon aus, dass sie dem Grundgedanken des Modus Ponens folgen. In Anwendungen haben sich folgende Semantiken für die Beschreibung unscharfer Regeln als sinnvoll erwiesen:

- Bedingte Wahrscheinlichkeiten: Diesen Ansatz haben wir als Basis unserer Theorie in Abschnitt 5.1 beschrieben.
- Positive und negative Regeln: Wir haben gezeigt, dass man positive Regeln (Unterstützung von Konsequenzen) und negative Regeln (Ausschluss von Konsequenzen) durch die logische Konjunktion und die Implikation darstellen kann. Konjunktion und Implikation wiederum können wir in un-

serer Theorie mit bedingten Wahrscheinlichkeiten modellieren. In praktischen Anwendungen werden positive und negative Regeln bereits erfolgreich kombiniert, um Expertenwissen kompakt und verständlich darzustellen [CDP96, Kie97, SSR00b].

- **Unschärfe Interpolation:** Regelbasierte unscharfe Systeme können wir als Abbildungen von unscharfen Eingaben auf unscharfe Ausgaben interpretieren. Als Verallgemeinerung klassischer Interpolationsverfahren aus der Numerik haben wir gezeigt, dass man unscharfe Regeln als unscharfe Stützpunkte einer unscharfen Abbildung interpretieren kann, zwischen denen unser Inferenzverfahren interpoliert. Wie die weite Verbreitung des Reglers von TAKAGI/SUGENO und von Neuro-Fuzzy-Systemen in technischen Anwendungen zeigt, entspricht die funktionale Beschreibung von Zusammenhängen viel eher der Denkweise eines Systemingenieurs als die formal logische Darstellung mit Regeln.

Insgesamt zeigt sich, dass unsere Theorie unscharfen Schließens so flexibel ist, dass sie wichtige Formen von Wissen, die bereits erfolgreich in der Praxis benutzt werden, in einem allgemeinen Ansatz zur Informationsverarbeitung vereinen kann. Dies ist ein wichtiger Fortschritt, denn bisher wurde auf die saubere Definition einer konkreten Semantik meist verzichtet, s. Kapitel 3, oder man hat sich wie in [SP98] auf eine Semantik spezialisiert, so dass viele Arten von Informationen nicht in das System eingebracht werden können.

Daraus ergeben sich zwei Konsequenzen. Dank der nun vorhandenen konkreten Semantik können wir aus unserer Arbeit Verfahren ableiten, die den Entwurf bekannter unscharfer Systeme transparenter machen und deren Wohldefiniertheit sicherstellen. Dazu gehören der Regler von TAKAGI/SUGENO und Neuro-Fuzzy-Systeme, die wir in unsere Theorie eingebettet haben. Andererseits bieten auch diese subsumierten Ansätze Entwurfs- und Adaptationstechniken, z. B. die Anpassung von Regelgewichten an Daten [Jan92, WM92, GRL00], die wir für den praktischen Einsatz unserer Verfahren nutzen können.

Hinsichtlich der implementierungstechnischen Seite der Repräsentation von Wissen haben wir gezeigt, dass lediglich die Darstellung unscharfer Begriffe sehr speicheraufwändig sein kann. Als Lösung haben wir ein Parametermodell vorgeschlagen, das einerseits den semantischen Anforderungen der Theorie gerecht wird und andererseits den Speicheraufwand um einen Faktor reduziert, der durchaus im Bereich von  $10^4$  liegen kann.

# Zusammenfassung und Ausblick

## 7.1 Zusammenfassung

Ziel der vorliegenden Arbeit war, ein solides theoretisches Fundament für regelbasiertes Schließen mit unscharfen Begriffen zu legen. Daran haben wir folgende Anforderungen geknüpft: Aus der Theorie sollen Verfahren resultieren, die mehrstufiges unscharfes Schließen ermöglichen, verständliche und exakte unscharfe Ergebnisse berechnen, effizient hinsichtlich Speicher- und Rechenaufwand sind und die Integration von Informationen verschiedener Semantiken erlauben.

Als Menschen sehen wir bei der Beschreibung von Objekten oder Zusammenhängen meist von Einzelheiten ab und benutzen oft unscharf definierte Begriffe. Trotzdem sind wir in der Lage, sinnvolle Schlüsse zu ziehen. Alltagsentscheidungen beruhen fast ausschließlich auf solchen Inferenzprozessen. Eine mathematische Theorie für das Schließen mit unscharfen Begriffen soll diese Vorgehensweise nachbilden, um sie rechnergestützt einsetzen zu können. Auf diese Weise können die Verfahren automatisiert ablaufen, wir können größere Regelwerke verarbeiten und erhalten deterministische Ergebnisse.

Ein zentrales Konzept unscharfen Schließens ist die Vergrößerung von Wissen. Je komplexer Systeme werden, desto stärker werden wir bei ihrer Modellierung vergrößern müssen, weil wir entweder keine präzisen Informationen besitzen oder eine genaue Modellierung zu teuer ist, sei es aus zeitlichen oder implementierungstechnischen Gründen. Eine wesentliche Rolle spielt auch die Verständlichkeit der Modellierung. Beschreibungen auf der Grundlage von Einzelheiten mögen exakt sein, sie tragen aber in vielen Fällen nicht zum Verständnis bei, weil sie zu detailliert sind. Indem wir Objekte und Zusammenhänge mit unscharfen Begriffen beschreiben, bewegen wir uns auf einer Abstraktionsstufe, die erstens unserem Verständnis entgegenkommt und zweitens oft der Genauigkeit des Wissens entspricht.

In einer Untersuchung der vorhandenen Verfahren zum mehrstufigen Schließen mit unscharfen Begriffen haben wir die Mängel aufgezeigt, die diese Verfahren hinsichtlich unserer Anforderungen aufweisen. Erstens werden Regeln zur Beschreibung kausaler Zusammenhänge zwar mit unscharfen Begriffen formuliert und sind damit verständlich; die unscharfen Schlüsse bewegen sich jedoch aus

dem Gerüst der Begriffe hinaus, da sie auf der Ebene der Einzelheiten definiert sind. Damit können wir unscharfe Schlüsse nicht mit den unscharfen Begriffen beschreiben, welche die Grundlage unseres Verständnisses bilden. Zweitens werden bei einem Großteil der Ansätze die Schlüsse durch Operationen auf den Einzelheiten berechnet, weshalb wir keinerlei Effizienzgewinn durch die Vergrößerung von Wissen verzeichnen können. Gleiches gilt für den Speicheraufwand. Drittens verfügen die meisten Ansätze nicht über eine konkrete Semantik für unscharfe Begriffe. Es werden i. A. formale Modelle mit vielen Freiheitsgraden angegeben, was aber solche abstrakten Semantiken in der Praxis konkret bedeuten, bleibt offen. Zudem berücksichtigen die Verfahren nicht, dass die Informationen in praktischen Anwendungen auf unterschiedlichen konkreten Semantiken beruhen und insofern heterogen sind. Dadurch wird die gemeinsame Verarbeitung heterogener Informationen schwierig oder die Ergebnisse sind zumindest fragwürdig.

Diese Unzulänglichkeiten bekannter Verfahren werden mit der Theorie zum Schließen mit unscharfen Begriffen beseitigt, die wir in den vorangehenden Kapiteln vorgestellt haben. Es gelingt uns, mehrstufiges unscharfes Schließen auf der symbolischen Ebene unscharfer Begriffe zu realisieren, ohne die Exaktheit der Ergebnisse auf der Ebene der Einzelheiten zu verlieren. Ein unscharfer Schluss wird dabei immer durch unscharfe Begriffe beschrieben und ist somit leicht verständlich. Rechen- und Speicheraufwand hängen im Wesentlichen nur von der Anzahl der unscharfen Begriffe und nicht mehr von der Anzahl der Einzelheiten ab.

Hinzu kommt, dass unser Verfahren auf einer konkreten Semantik für unscharfe Begriffe und Regeln beruht. In diesem Zusammenhang haben wir gezeigt, dass einige bekannte Verfahren aus der Praxis zur Gewinnung und Darstellung unscharfer Begriffe und Regeln in Instanzen unserer Semantik münden. Da diese Verfahren in bereits beschriebener Weise heterogene Informationen erzeugen, können wir in diesem Rahmen symbolisch mit unscharfen Begriffen schließen, ohne die Semantiken der Begriffe und Regeln berücksichtigen zu müssen. Dabei können wir erstmals die Wohldefiniertheit des Mechanismus und damit der unscharfen Schlüsse sicherstellen. Somit haben wir alle Anforderungen erfüllt, die wir an die Theorie gestellt haben.

Folgende Liste gibt nochmals einen Überblick über die Ideen, Techniken und Neuerungen unserer Theorie:

- Wir vergrößern ein Universum, d. h. eine Menge von Einzelheiten, durch *unscharfes Partitionieren* mit unscharfen Mengen. Dazu haben wir die Eigenschaften Vollständigkeit und Konsistenz scharfer Partitionen für den unscharfen Fall verallgemeinert und aus dieser Erweiterung ein Verfahren für die unscharfe Partitionierung abgeleitet. Die Elemente einer unscharfen Partition bezeichnen wir als *unscharfe Grundbegriffe*.

- Unscharfe Aussagen stellen wir auf der grobgranularen Ebene durch gewichtete Kombinationen der Grundbegriffe dar. Durch die *Synthese* wird eine solche symbolische Aussage als unscharfe Menge auf der feingranularen Ebene definiert. Auf der anderen Seite erlaubt die *Analyse*, eine beliebige unscharfe Menge in eine symbolische Aussage zu überführen und somit verständlich mit den Grundbegriffen darzustellen.

Damit haben wir erstmals eine wohldefinierte Zuordnung unscharfer Aussagen auf feingranularer und auf symbolischer Ebene hergestellt, die auf einer soliden, konkreten Semantik beruht; unscharfe Aussagen sind gleichzeitig exakt definiert und verständlich.

- Unscharfe Regeln interpretieren wir als Ausdruck des klassischen Modus Ponens. Konsistent zur Gewichtung von Grundbegriffen modellieren wir unscharfe Regeln als bedingte Wahrscheinlichkeiten auf den Grundbegriffen. Das Inferenzverfahren operiert auch bei mehrstufigen Regelbasen ausschließlich auf der symbolischen Ebene und ist damit sehr effizient im Gegensatz zu herkömmlichen Verfahren, die auf den Einzelheiten rechnen. Obwohl wir auf symbolischer Ebene rechnen, können wir die Exaktheit der Schlüsse auf feingranularer Ebene sicherstellen.

Ein unscharfer Schluss wird symbolisch als gewichtete Kombination von Grundbegriffen dargestellt und ist dadurch leicht verständlich. Die exakte Darstellung auf der Ebene der Einzelheiten können wir daraus mittels Synthese ableiten. Im Gegensatz dazu berechnen die meisten anderen Verfahren ausschließlich feingranulare Schlüsse, die auf symbolischer Ebene nicht interpretierbar sind.

- Der Rechenaufwand einer mehrstufigen Inferenz mit eindimensionalen Ein- und Ausgaben hängt im besten Fall – der gleichzeitig der häufigste in der Praxis ist – nur linear von der Anzahl der Regeln ab. Das schnellste Verfahren aus der Literatur benötigt einen Rechenaufwand, der von der quadrierten Anzahl der Regeln abhängt, der Aufwand der meisten anderen Ansätze verhält sich quadratisch mit der Anzahl der Einzelheiten und linear mit der Anzahl der Regelstufen. Wie gewünscht, reduziert sich also bei unserem Ansatz der Aufwand der Informationsverarbeitung mit dem Vergrößerungsgrad der Informationen.
- Wir haben Maße definiert, welche die Zuverlässigkeit symbolischer und feingranularer Schlüsse messen. Dank der konkreten Semantik unserer Verfahren sind diese Zuverlässigkeitswerte aussagekräftig und verständlich, was für die meisten anderen Verfahren nicht gilt, da sie auf abstrakten Semantiken beruhen.

- Wir haben gezeigt, dass wir Informationen verschiedener Semantiken mit unserer mathematischen Formalisierung darstellen und verarbeiten können. Dabei handelt es sich erstens um unscharfe Gruppen: Ähnliche Einzelheiten werden zu unscharfen Mengen gruppiert, die eine unscharfe Partition bilden, wodurch wir aus Daten unscharfe Partitionen in unserem Sinne generieren können. Zweitens haben wir ein Verfahren vorgestellt, mit dem scharfe Beobachtungen durch unscharfe Grundbegriffe dargestellt und somit effizient verarbeitet werden können. Darüber hinaus können wir mit unsicheren Informationen umgehen, die als Wahrscheinlichkeitsdichten gegeben sind.
- Bei der Definition von Regeln haben wir diverse Möglichkeiten vorgestellt, den Modus Ponens zu realisieren. Insbesondere haben wir die Bedeutung positiver und negativer Regeln aufgezeigt, diese mit der logischen Konjunktion und Implikation identifiziert und in unser wahrscheinlichkeitstheoretisches Modell integriert.

Außerdem haben wir gezeigt, dass man unser Inferenzverfahren als eine unscharfe Interpolation zwischen unscharfen Punkten deuten kann, die durch die Regeln repräsentiert werden. Diese Interpretation von Regeln entspricht der funktionalen Beschreibung von Zusammenhängen, die in den Ingenieurwissenschaften einer formal logischen vorgezogen wird.

- Da die Darstellung unscharfer Grundbegriffe bei großen Datenmengen sehr speicherintensiv sein kann, haben wir ein Parametermodell als Approximation der unscharfen Grundbegriffe vorgestellt. Der Speicheraufwand hängt durch diese Maßnahme nur noch von der Anzahl der Grundbegriffe ab, anstatt von der Anzahl der definierenden Daten.

Ein bedeutendes Nebenresultat der Arbeit ist, dass wir den Zusammenhang zwischen unscharfem Schließen und einer wichtigen Klasse von Interpolationsverfahren klären konnten. Interpolation erweist sich in klassischen Ingenieurwissenschaften wie der Elektrotechnik als wichtiger Ansatz für die Approximation von Funktionen wie Reglern oder Übertragungsfunktionen von Systemen. Wir haben gezeigt, dass weit verbreitete Verfahren wie RBF-Netzwerke, Neuro-Fuzzy-Ansätze und ein wichtiger Spezialfall des Reglers von TAKAGI/SUGENO, die Interpolation durch lineares Kombinieren von Basisfunktionen realisieren, Spezialfälle unserer Theorie darstellen. Auf diese Weise versehen wir sie mit einer konkreten Semantik und können sie als Verfahren für unscharfes Schließen anerkennen. Bisher war diese Deutung rein intuitiv motiviert. Neben der reinen Erkenntnis können wir dank dieser Deutung auch Verfahren zur Verfügung stellen, die beim Entwurf von Systemen auf Basis der genannten Ansätze helfen. Insbesondere handelt es sich dabei um die Verfahren zum unscharfen Partitionieren von Universen – die resultierenden Grundbegriffe entsprechen den Basisfunktionen der Interpolationsverfahren – und zum Festlegen der Regelgewichte.

Die Auswirkungen unserer Theorie auf praktische Probleme sind vielschichtig. Die Steigerung der Effizienz kann die Bearbeitung von Problemen ermöglichen, die bisher als zu komplex erschienen. Dabei denken wir insbesondere an Anwendungen aus dem Bereich Data-Mining, die wegen ihrer Problemgröße ohne eine Vergrößerung der Informationen kaum handhabbar sind.

Weiterhin stellen wir praktische Verfahren für den Entwurf regelbasierter unscharfer Systeme bereit. Wir haben verschiedene Möglichkeiten vorgestellt, um unscharfes Wissen zu modellieren und Regeln zu definieren. Die Verarbeitung der Informationen berücksichtigt die unterschiedlichen Semantiken und liefert dadurch vertrauenswürdige Ergebnisse.

Darüber hinaus sichert die konkrete Semantik der Theorie die Verständlichkeit der unscharfen Ergebnisse. Wir sind davon überzeugt, dass wir auf diese Weise zur Akzeptanz unscharfer Systeme beitragen können, die bisher wegen der schwächeren Semantik weniger Ansehen genossen haben als die Wahrscheinlichkeitsrechnung.

Für die praktische Umsetzung unserer Theorie haben wir prototypisch ein Werkzeug in MAPLE V implementiert, mit dem unscharfe Begriffe und Regelnetzwerke definiert und mehrstufige Inferenzen berechnet werden können. Es wurde erfolgreich benutzt, um unscharfe Regler zu entwerfen [SR98, SSR00a] und das Beispiel in Anhang A zu modellieren.

Unser Inferenzmechanismus wird bereits erfolgreich in einem Industrieprojekt mit ABB eingesetzt. Gegenstand dieser Kooperation sind entscheidungsunterstützende Systeme, mit deren Hilfe der Qualitätszustand technischer Anlagen beurteilt werden soll, um über Maßnahmen zur Qualitätssicherung zu entscheiden. Wie auch das Beispiel in Anhang A zeichnet sich diese Aufgabe dadurch aus, dass das Wissen über Objekte und Zusammenhänge unscharf und unvollständig ist und daher ein unscharfer Ansatz als natürliche Modellierung erscheint.

## 7.2 Ausblick

Den abschließenden Ausblick geben wir unter zwei Gesichtspunkten: Zunächst diskutieren wir Möglichkeiten, die vorgestellte Theorie zu erweitern, wodurch ihre Mächtigkeit und damit das Anwendungsspektrum vergrößert werden kann. Danach zeigen wir die Auswirkungen der vorgestellten Theorie auf andere Forschungsgebiete.

### 7.2.1 Erweiterung der Theorie

Unschärfe Beobachtungen stellen wir in unserem Ansatz durch gewichtete Kombinationen von Grundbegriffen dar. Das Basisverfahren der Analyse berechnet dabei u. U. negative Gewichte, die nicht als Wahrscheinlichkeiten interpretierbar sind und deshalb von uns nicht berücksichtigt werden.<sup>1</sup> Es ist aber durchaus wünschenswert, negative Gewichte zuzulassen, weil dadurch unscharfe Beobachtungen präziser durch Kombinationen der Grundbegriffe beschrieben werden können; wir vergrößern den Raum unscharfer Aussagen, die wir exakt mit den Grundbegriffen darstellen können. Allerdings müssten wir eine Möglichkeit finden, negative Gewichte semantisch in unseren Ansatz zu integrieren.

Eine mögliche Interpretation eines negativen Gewichtes wäre die eines *partiellen Vetos* gegen den zugehörigen unscharfen Begriff. Ein negatives Gewicht entspräche dann beispielsweise bei einer Umfrage den Stimmen gegen einen unscharfen Begriff und ein positives Gewicht den Stimmen dafür. Für jeden Begriff würden positives und negatives Gewicht so verrechnet, dass wir ein nichtnegatives Gewicht erhielten, das wir wiederum als Wahrscheinlichkeit interpretieren könnten. Falls sich diese Interpretation semantisch in unsere Theorie einbetten ließe, könnten wir auf diese Weise die Ausdrucksmächtigkeit unseres und anderer Ansätze merklich erhöhen.

In einigen Anwendungen ist es möglicherweise nicht angemessen, die Gewichte von Grundbegriffen und Regeln über Wahrscheinlichkeiten zu modellieren. Hier bietet sich eine Verallgemeinerung über Plausibilitätsmaße an. Man müsste dazu untersuchen, ob es wie bei Wahrscheinlichkeiten gelingen kann, den Zusammenhang zwischen der grobgranularen und der feingranularen Darstellung unscharfer Aussagen semantisch solide aufzubauen. Wie auch die Erweiterung durch negative Gewichte, würde dieser Ansatz die Ausdrucksmächtigkeit unserer Modellierung von Wissen erhöhen. Wir könnten insbesondere adäquat mit Unwissen umgehen, was bei Wahrscheinlichkeiten problematisch ist, wie wir in Abschnitt 6.3.1 gesehen haben.

---

<sup>1</sup>Dieses Problem ist in der Literatur auch im Zusammenhang mit anderen Verfahren bekannt [NK98].

Bei der Partitionierung von Universen durch Grundbegriffe haben wir verlangt, dass die Grundbegriffe disjunkt sind, um die Konsistenz der unscharfen Partition sicherzustellen. Damit setzen wir aber konsistentes Wissen voraus. Es müsste untersucht werden, ob es immer sinnvoll ist, vor der Modellierung Inkonsistenzen ausräumen, oder ob man in gewissen Fällen nicht besser inkonsistentes Wissen bewusst modellieren und verarbeiten sollte.

Wenn man die Forderung von Konsistenz fallen lässt, muss man den Inferenzmechanismus anpassen. An den Rechenvorschriften würde sich nichts ändern, allerdings müsste man zusätzliche Regeln berechnen, welche die gegenseitige Beeinflussung ähnlicher Regeln aufzunehmen hätten. Auf diese Weise könnte man inkonsistentes Wissen verarbeiten, ohne die Effizienz des Verfahrens und die Verständlichkeit der Ergebnisse zu gefährden.

Insgesamt sehen wir, dass die Theorie noch Raum für Erweiterungen bietet. Das Ziel muss dabei immer bleiben, das Anwendungsspektrum zu vergrößern, ohne die Semantik und damit die Wohldefiniertheit der Ergebnisse zu gefährden.

## 7.2.2 Auswirkungen auf andere Gebiete

Unser Inferenzmechanismus basiert auf einer wahrscheinlichkeitstheoretischen Erweiterung der Possibilitätstheorie. Diese Art, Regeln als bedingte Wahrscheinlichkeiten zu formulieren und zu verarbeiten, kennen wir in ähnlicher Form von BAYESSCHEN Netzen [Pea88, RN95]. Sie erweisen sich in den letzten Jahren als erfolgreicher Ansatz in den Bereichen Data-Mining und Diagnose. MICROSOFT RESEARCH versucht beispielsweise, das automatische Hilfesystem (*Office Assistant*) mit dieser Technik zu realisieren [HBH<sup>+</sup>98]. Die Grundidee von BAYESSCHEN Netzen besteht darin, eine mehrdimensionale Wahrscheinlichkeitsverteilung durch Strukturwissen über die Unabhängigkeiten von Variablen zu zerlegen. Auf diese Weise wird es ermöglicht, Inferenzen zu berechnen, die im allgemeinen Fall wegen exponentiellen Aufwands nicht abgeleitet werden können.

Bisher kann man in BAYESSCHEN Netzen nur mit scharfen Informationen rechnen. Da BAYESSCHE Netze aber wie unser Ansatz die Wahrscheinlichkeitsrechnung für den Inferenzmechanismus bemühen, könnten unsere Ergebnisse leicht übertragen werden. Eine Kombination der Ansätze würde in einem mächtigen Verfahren resultieren: Von BAYESSCHEN Netzen erhielten wir Techniken, um Abhängigkeitsstrukturen zu konstruieren und darin Inferenzen zu berechnen, von unserer Seite Verfahren zur Vergrößerung von Wissen über unscharfe Grundbegriffe und deren Verarbeitung. Als Ergebnis könnten wir effizient verständliche unscharfe Schlüsse aus unscharfem Wissen ziehen. Dabei wären wir nicht darauf angewiesen, Zusammenhänge regelbasiert zu formulieren, sondern könnten aus Daten allgemeinere statistische Relationen aufbauen.

Eine ganz andere Richtung könnten wir mit einer neuen, unscharfen Systemtheorie einschlagen. Wie zum Teil schon üblich in der Systemtheorie, können wir das Ein-/Ausgabeverhalten von Systemen mit unscharfen Regeln modellieren. Dabei bietet sich die Interpretation unseres Inferenzmechanismus als unscharfe Interpolation an. Untersuchungen und Aussagen über das Verhalten von Systemen, wie beispielsweise über ihre Stabilität, sind schwierig, wenn die Systeme nur unscharf beschrieben werden können.

Dank unserer symbolischen Darstellung können wir solche Untersuchungen auf der symbolischen, grobgranularen Ebene durchführen. Da eine Inferenz, welche die Ausgabe eines Systems berechnet, durch Matrixmultiplikationen dargestellt wird, können die Untersuchungen mithilfe linearer Algebra verwirklicht werden. Für Stabilitätsuntersuchungen beispielsweise könnte man sicherstellen, dass der Zustand des Systems eine unscharfe Menge wie „ungefähr null“ nicht verlässt. Der Bereich, in dem die Zugehörigkeitsfunktion der unscharfen Menge null ist, wird dann sicher nicht betreten.

Viele bekannte Verfahren zur Untersuchung von linearen Systemen könnte man vermutlich direkt übertragen, da lineare Systeme in Zustandsraumdarstellung die gleiche mathematische Form besitzen wie unsere Regeln. Der einzige Unterschied besteht darin, dass wir die Berechnungen auf einer größeren Stufe durchführen. Damit wären Untersuchungen möglich, die vorher mangels präzisen Wissens undenkbar erschienen.

Die Diversität der beiden Beispiele – BAYESSCHE Netze und unscharfe Systemtheorie – ist ein Indiz für die Breite der Anwendungsmöglichkeiten unserer Theorie. Dabei wird auch noch einmal unsere These unterstützt, dass es wichtig ist, unscharfes Schließen auf einer konkreten Semantik aufzubauen, die allgemein genug ist, um andere wichtige Semantiken zu subsumieren.

## Anhang A

---

# Beispiel: Betrugserkennung

Als Beispiel für das Schließen mit unscharfen Begriffen beschreiben wir in diesem Anhang ein regelbasiertes System, das betrugsverdächtige Versicherungsfälle erkennt. Alle Fälle, die das System als hinreichend betrugsverdächtig identifiziert, werden anschließend von Experten begutachtet. Wir gehen in diesem Beispiel von kleineren Versicherungsfällen aus, man denke etwa an eine Haftpflichtversicherung für Privatkunden. Das System haben wir mit unserem Werkzeug in MAPLE V erstellt.

Das Beispiel lehnt sich an eine reale Anwendung der Firma INFORM GmbH, Aachen, an, die für ein Unternehmen aus der Versicherungs- und Bankbranche ein unscharfes, regelbasiertes System zur Erkennung von Versicherungsbetrug entwickelt hat [vAJ00]. Das System von INFORM basiert auf dem Prinzip von Kombination/Projektion.

Da viele Hinweise auf einen Versicherungsbetrug unsicher oder vage sind, bietet es sich an, ein System zur Betrugserkennung unscharf zu modellieren. Es gibt zwar viele Indizien für einen Betrug, die auf harten Fakten beruhen, letztlich ist aber die Persönlichkeit des Kunden ausschlaggebend, die wir nur sehr eingeschränkt und indirekt an zugänglichen Daten ablesen können. So ist ein Kunde, der erst seit einem Monat versichert ist, einen hohen Betrag von der Versicherung fordert und finanziell ruiniert ist, prinzipiell sehr verdächtig. Ein ehrlicher Mensch ist aber vielleicht selbst in ausweglosen Situationen nicht zu einer Straftat zu bewegen. Der Mangel an Messbarkeit oder Zugänglichkeit der Kundenpersönlichkeit führt zu unscharfem Wissen über die Zusammenhänge von Kundendaten und Betrugsverdacht. Es liegt daher nahe, diese Zusammenhänge auch unscharf zu modellieren. Zwar stehen scharfe Daten aus bekannten Betrugsfällen zur Verfügung; da die Betrugsfälle aber vor ihrer Aufdeckung von Experten als verdächtig identifiziert werden müssen, befindet sich das Wissen ausschließlich bei den Experten. Bekannte Betrugsfälle können aber durchaus zur Validierung eines Systems benutzt werden.

In Abschnitt A.1 beschreiben wir die Struktur des regelbasierten Systems, die unscharfen Begriffe und die Regeln des Systems, die auf der Basis der unscharfen Begriffe formuliert sind. Anschließend zeigen wir in Abschnitt A.2 die Ergebnisse, die das unscharfe System in zwei Beispielfällen berechnet. Dabei diskutieren

wir verschiedene Möglichkeiten, die Entscheidung für oder gegen die weitere Untersuchung eines Falles zu fällen. Abschließend bewerten wir in Abschnitt A.3 unseren Ansatz im Vergleich mit anderen Verfahren wie dem Prinzip von Kombination/Projektion.

## A.1 Modellierung

Die Struktur des Systems sehen wir in Abb. A.1. Die Eingabegrößen, die Einflussfaktoren einer Entscheidung für oder gegen Betrugsverdacht, wurden in zwei Gruppen eingeteilt. Die Einflussfaktoren der Versicherung bilden die erste Gruppe, die der Bank die zweite Gruppe. Im Einzelnen handelt es sich auf Versicherungsseite um die Anzahl der Versicherungsfälle, die der Kunde in den letzten 12 Monaten eingereicht hat, den Betrag des aktuellen Versicherungsfalles und die bisherige Vertragsdauer. Aus diesen Größen wird geschlossen, wie stark der Kunde die Versicherung belastet. Auf Bankseite werden der gemittelte Kontostand der letzten 12 Monate und die Anzahl der Monate, während der das Konto in den vergangenen 12 Monaten überzogen war, zur Kreditwürdigkeit zusammengefasst. Aus der Versicherungsbelastung und der Kreditwürdigkeit wird schließlich auf die Betrugswahrscheinlichkeit geschlossen.

Wir nehmen an dieser Stelle an, dass die Bankdaten adäquat vorverarbeitet sind. Insbesondere diskutieren wir nicht, wie oft der Kontostand bestimmt und nach welcher Methode er gemittelt wird. Wir kommen auf die Mittelung des Kontostands allerdings in Abschnitt A.2 zurück, weil wir ihn auch unscharf messen können.

Das System wurde zweistufig aufgebaut, um die Kunden hinsichtlich ihrer Versicherungsbelastung und ihrer Kreditwürdigkeit getrennt bewerten zu können. Bei Betrugsverdacht kann dann unmittelbar nachvollzogen werden, ob die Versicherungsbelastung oder die Kreditwürdigkeit den wesentlichen Ausschlag für den Verdacht gegeben hat.

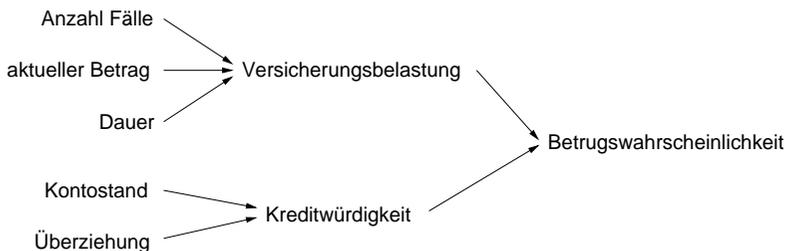


Abbildung A.1: Struktur des regelbasierten Systems

Die Wertebereiche der acht Größen haben wir unscharf partitioniert. Die Betrugswahrscheinlichkeit haben wir mit fünf unscharfen Grundbegriffen überdeckt (Abb. A.9), alle anderen Universen mit drei Grundbegriffen, wie die Abbildungen A.2–A.8 illustrieren. In den Abbildungen haben wir die Wertebereiche der Übersichtlichkeit wegen eingeschränkt. Werte, die außerhalb dieser Bereiche liegen, werden auf die Unter- bzw. Obergrenze abgebildet.

In der Praxis werden die unscharfen Partitionen durch die Experten des Unternehmens festgelegt. Die Anzahl der Grundbegriffe einer Partition entspricht der Anzahl der unscharfen Werte, welche die Experten für die entsprechende Größe unterscheiden möchten. Die Grundbegriffe selbst werden auf die gleiche Weise mittels Partitionierung durch die Experten definiert wie in Abschnitt 4.1 die unscharfen Körpergrößen: Jeder Experte legt eine scharfe Partition fest, aus den Partitionen aller Experten bilden wir schließlich die unscharfe Partition. Für unsere Illustration haben wir die Grundbegriffe durch unscharfe Mengen in Form von Trapez- und Dreieckfunktionen repräsentiert, die wir bereits in Abschnitt 6.4 für die Approximation von Plausibilitätsdichten benutzt haben.

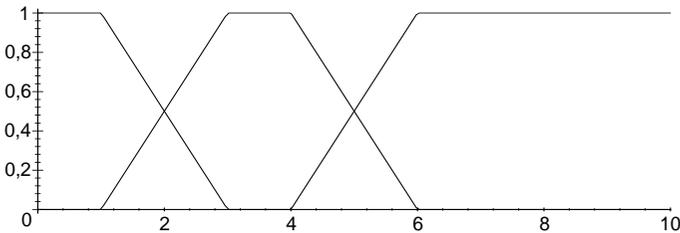


Abbildung A.2: Anzahl der Versicherungsfälle in den letzten 12 Monaten: *niedrig*, *mittel* und *hoch*

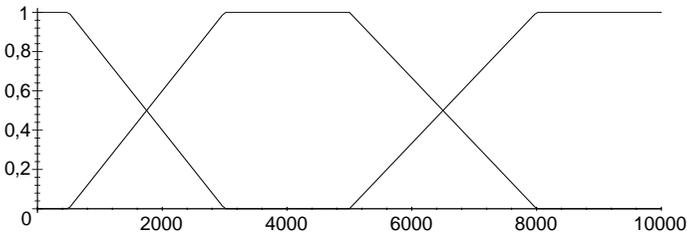


Abbildung A.3: Betrag des aktuellen Versicherungsfalles in DM: *niedrig*, *mittel* und *hoch*

Um die Regeln festzulegen, teilen wir das System in drei Blöcke. Im ersten Block schließen wir auf die Versicherungsbelastung, im zweiten auf die Kreditwürdigkeit und im dritten auf die Betrugswahrscheinlichkeit. Die Regeln des ersten Blocks haben die Form „Wenn die Anzahl der Fälle *niedrig* ist, der aktuelle Betrag *niedrig* ist

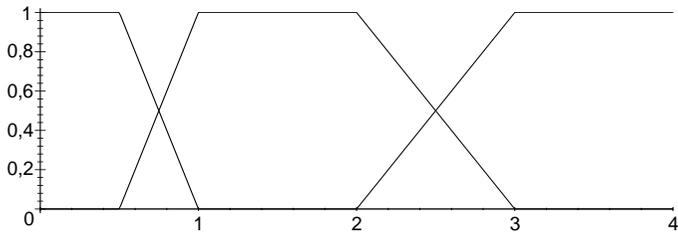


Abbildung A.4: Bisherige Vertragsdauer der Versicherung in Jahren: *kurz, mittel* und *lang*

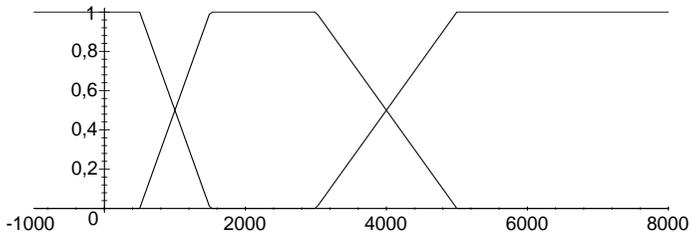


Abbildung A.5: Durchschnittlicher Kontostand während der letzten 12 Monate in DM: *niedrig, mittel* und *hoch*

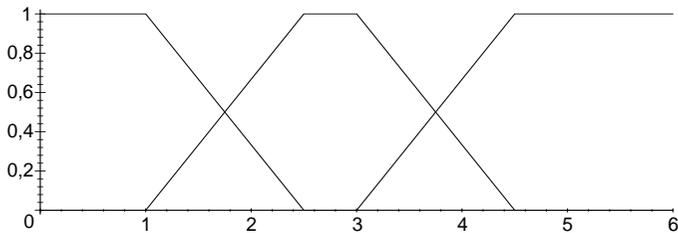


Abbildung A.6: Anzahl der Monate, während der das Konto in den letzten 12 Monaten überzogen war: *niedrig, mittel* und *hoch*

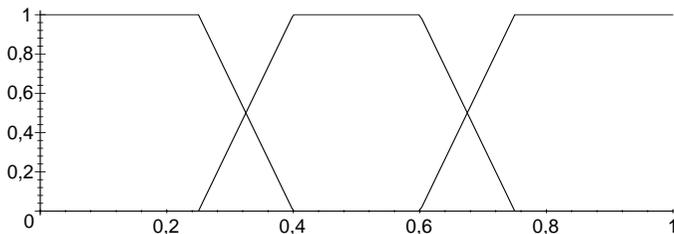
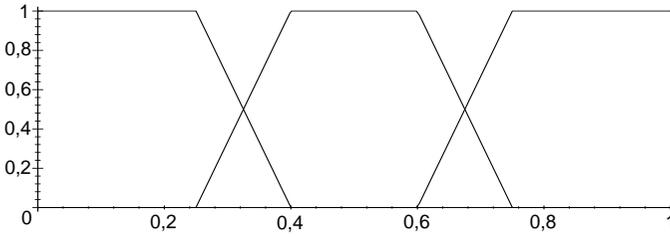
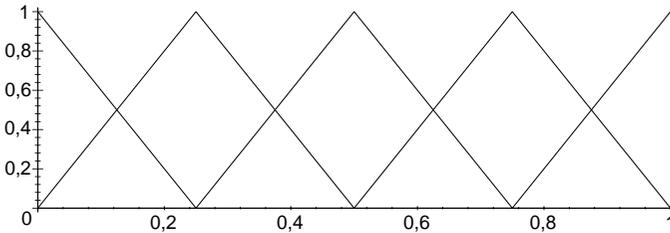


Abbildung A.7: Grad der Versicherungsbelastung auf der Skala  $[0, 1]$ : *gering, mittel* und *stark*

Abbildung A.8: Kreditwürdigkeit auf der Skala [0, 1]: *gering, mittel und stark*Abbildung A.9: Betrugswahrscheinlichkeit auf der Skala [0, 1]: *sehr klein, klein, mittel, groß und sehr groß*

und die Dauer *mittel* lang ist, dann ist die Versicherungsbelastung *gering*.“ Diese Regeln können noch mit einem Gewicht versehen werden, der bedingten Wahrscheinlichkeit  $P_p(\text{gering}|\text{niedrig,niedrig,mittel})$ , s. Abschnitt 5.1. Wir bestimmen die bedingten Wahrscheinlichkeiten, indem wir der Expertengruppe eine Kombination von Antezedenstermen wie (*niedrig,niedrig,mittel*) präsentieren und dann abstimmen lassen, ob die Versicherungsbelastung in diesem Fall *gering, mittel* oder *stark* ist. So könnten sich z. B. 70% der Befragten für *gering*, 30% für *mittel* und 0% für *stark* entscheiden. Die bedingten Wahrscheinlichkeiten wären entsprechend  $P_p(\text{gering}|\text{niedrig,niedrig,mittel}) = 0,7$ ,  $P_p(\text{mittel}|\text{niedrig,niedrig,mittel}) = 0,3$  und  $P_p(\text{stark}|\text{niedrig,niedrig,mittel}) = 0$ . Auf diese Weise gehen wir alle Kombinationen von Antezedenstermen durch. Im ersten Regelblock betrachten wir drei Eingangsgrößen mit je drei möglichen unscharfen Werten; wir benötigen also  $3^3 = 27$  Abstimmungen.

Für den zweiten und dritten Regelblock gehen wir entsprechend vor. Wir erhalten Regeln der Form „Wenn der Kontostand *niedrig* ist und der Kunde oft überzieht (Anzahl *hoch*), dann ist die Kreditwürdigkeit *gering*.“ bzw. „Wenn die Versicherungsbelastung *stark* und die Kreditwürdigkeit *gering* ist, dann ist die Betrugswahrscheinlichkeit *sehr groß*.“

Bei unseren Regeln verzichten wir der einfacheren Darstellung wegen auf Ge-

wichte. Die bedingten Wahrscheinlichkeiten sind damit eins oder null, was einem einstimmigen Votum der Experten entspricht. Die Regeln der drei Blöcke sind in den Tabellen A.1 – A.3 zusammengefasst.

| Dauer   |               | kurz          |              |  |
|---------|---------------|---------------|--------------|--|
| # Fälle | Betrag        |               |              |  |
|         | niedrig       | mittel        | hoch         |  |
| niedrig | <i>gering</i> | <i>mittel</i> | <i>stark</i> |  |
| mittel  | <i>stark</i>  | <i>stark</i>  | <i>stark</i> |  |
| hoch    | <i>stark</i>  | <i>stark</i>  | <i>stark</i> |  |

| Dauer   |               | mittel        |               |  |
|---------|---------------|---------------|---------------|--|
| # Fälle | Betrag        |               |               |  |
|         | niedrig       | mittel        | hoch          |  |
| niedrig | <i>gering</i> | <i>gering</i> | <i>mittel</i> |  |
| mittel  | <i>mittel</i> | <i>mittel</i> | <i>stark</i>  |  |
| hoch    | <i>stark</i>  | <i>stark</i>  | <i>stark</i>  |  |

| Dauer   |               | lang          |               |  |
|---------|---------------|---------------|---------------|--|
| # Fälle | Betrag        |               |               |  |
|         | niedrig       | mittel        | hoch          |  |
| niedrig | <i>gering</i> | <i>gering</i> | <i>gering</i> |  |
| mittel  | <i>gering</i> | <i>mittel</i> | <i>mittel</i> |  |
| hoch    | <i>mittel</i> | <i>stark</i>  | <i>stark</i>  |  |

Tabelle A.1: Regeln, die von Anzahl der Versicherungsfälle, Betrag des aktuellen Falls und Dauer der Kundschaft auf die Versicherungsbelastung schließen

| Überziehung | ∅ Kontostand  |               |               |
|-------------|---------------|---------------|---------------|
|             | niedrig       | mittel        | hoch          |
| niedrig     | <i>mittel</i> | <i>stark</i>  | <i>stark</i>  |
| mittel      | <i>gering</i> | <i>mittel</i> | <i>stark</i>  |
| hoch        | <i>gering</i> | <i>mittel</i> | <i>mittel</i> |

Tabelle A.2: Regeln, die von der Dauer, mit der das Bankkonto überzogen wurde, und dem Kontostand auf die Kreditwürdigkeit schließen

## A.2 Ergebnisse

Zur Demonstration des Systems betrachten wir zwei Versicherungsfälle. Da alle Daten scharf sind, unterziehen wir sie zunächst einer natürlichen Abstimmung

| Kreditwürdigkeit | Versicherungsbelastung |               |              |
|------------------|------------------------|---------------|--------------|
|                  | gering                 | mittel        | stark        |
| gering           | <i>klein</i>           | <i>groß</i>   | <i>sgroß</i> |
| mittel           | <i>klein</i>           | <i>mittel</i> | <i>groß</i>  |
| stark            | <i>sklein</i>          | <i>mittel</i> | <i>groß</i>  |

Tabelle A.3: Regeln, die von der Kreditwürdigkeit und der Versicherungsbelastung auf die Betrugswahrscheinlichkeit schließen; *klein* und *sgroß* bedeuten *sehr klein* und *sehr groß*.

nach Abschnitt 6.2. In den Tabellen A.4 und A.5 sehen wir die scharfen Daten der beiden Fälle und das Ergebnis ihrer Abstumpfung als Kombination unscharfer Grundbegriffe.

Den Kontostand können wir alternativ zu einer scharfen Mittelung über die letzten 12 Monate mit anschließender Abstumpfung auch direkt unscharf repräsentieren. Bei einer Mittelung gehen viele Informationen über den Kontostand verloren, der über die Zeitdauer von 12 Monaten betrachtet unscharf ist. Je nach Verfahren (arithmetisches, gestutztes, winsorisiertes Mittel usw.) werden unterschiedliche Aspekte der Unschärfe modelliert. Die Variation des Kontostands über die 12 Monate geht bei jeder Mittelung verloren. Als Alternative kann man ein Histogramm als Approximation für die Wahrscheinlichkeitsdichte des Kontostands über die 12 Monate aufstellen, so dass alle Informationen erhalten bleiben. Die Wahrscheinlichkeitsdichte stellt als spezielle Plausibilitätsdichte eine unscharfe Information im Sinne unserer Arbeit dar, die wir mittels einer Analyse nach Abschnitt 4.2.3 als Kombination der Grundbegriffe beschreiben können. Damit befinden wir uns am gleichen Punkt wie nach der Abstumpfung eines scharfen Wertes.

|        |          | Anzahl der Fälle        | aktueller Betrag        | Dauer                |
|--------|----------|-------------------------|-------------------------|----------------------|
| Fall 1 | scharf   | 2                       | DM 1000                 | 2 Jahre              |
|        | unscharf | 0,5 niedrig, 0,5 mittel | 0,8 niedrig, 0,2 mittel | mittel               |
| Fall 2 | scharf   | 3                       | DM 5000                 | 0,7 Jahre            |
|        | unscharf | mittel                  | mittel                  | 0,6 kurz, 0,4 mittel |

Tabelle A.4: Die Versicherungsdaten von zwei Fällen: scharf und durch natürliche Abstumpfung mit unscharfen Grundbegriffen beschrieben

Die nun unscharfen Werte der Eingabegrößen verknüpfen wir mit den Regeln und berechnen durch Inferenz die Versicherungsbelastung, die Kreditwürdigkeit und schließlich die Betrugswahrscheinlichkeit. Die Schlüsse unseres unscharfen Systems liegen auf grobgranularer und auf feingranularer Ebene vor. Die grobgranularen Schlüsse sehen wir in den Tabellen A.6 und A.7. Wir lesen ein Ergebnis wie

|        |          | ∅ Kontostand            | Überziehung             |
|--------|----------|-------------------------|-------------------------|
| Fall 1 | scharf   | DM 1000                 | 1,5 Monate              |
|        | unscharf | 0,5 niedrig, 0,5 mittel | 2/3 niedrig, 1/3 mittel |
| Fall 2 | scharf   | DM 600                  | 4 Monate                |
|        | unscharf | 0,9 niedrig, 0,1 mittel | 1/3 mittel, 2/3 hoch    |

Tabelle A.5: Die Bankdaten von zwei Fällen: scharf und durch natürliche Abstumpfung mit unscharfen Grundbegriffen beschrieben

die Betrugswahrscheinlichkeit in Tab. A.6 als „Die Betrugswahrscheinlichkeit ist mit 17% Wahrscheinlichkeit *sehr klein*, mit 33% *klein*, mit 42% *mittel*, mit 8% *groß* und mit 0% *sehr groß*.“

Der Kunde aus Fall 1 belastet seine Versicherung eher gering, seine Kreditwürdigkeit ist mittel, was zusammen in einer unverdächtigen Betrugswahrscheinlichkeit resultiert, die mittel bis klein ist. Der zweite Kunde belastet die Versicherung stark bis mittel, ist nicht kreditwürdig und wird daher mit sehr großer bis großer Wahrscheinlichkeit als Betrüger eingeschätzt.

Aus diesen grobgranularen Schlüssen können wir durch Synthese feingranulare Schlüsse berechnen. Als Ergebnis erhalten wir unscharfe Mengen in Form von Plausibilitätsdichten, welche die groben Aussagen auf feingranularer Ebene definieren. Die Abbildungen A.10–A.12 zeigen die Plausibilitätsdichten der Schlüsse in Fall 1, die Abbildungen A.13–A.15 die Schlüsse in Fall 2. An einer Plausibilitätsdichte wie der Funktion in Abb. A.12 können wir ablesen, wie hoch die Plausibilität ist, dass die Betrugswahrscheinlichkeit  $u$  Prozent beträgt. Während also ein grobgranularer Schluss die Betrugswahrscheinlichkeit auf einer intuitiv verständlichen, symbolischen Ebene beschreibt, definiert die feingranulare Darstellung die genaue Bedeutung der symbolischen Aussage.

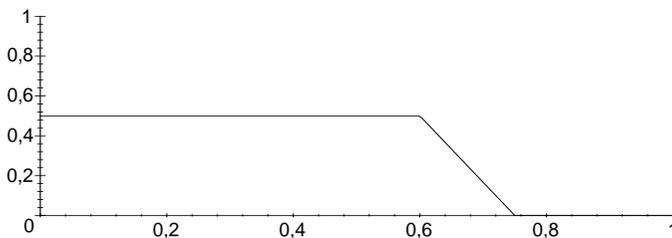


Abbildung A.10: Unscharfe Versicherungsbelastung in Fall 1

Um die Entscheidung zu treffen, ob ein Versicherungsfall wegen Betrugsverdachts näher untersucht werden muss, sind grundsätzlich zwei Ansätze denkbar. Einerseits können wir durch Schärfung der unscharfen Menge in Abb. A.12 bzw.

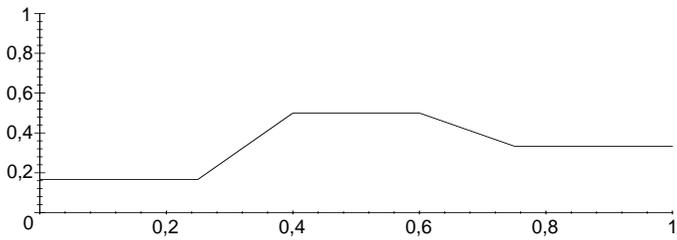


Abbildung A.11: Unscharfe Kreditwürdigkeit in Fall 1

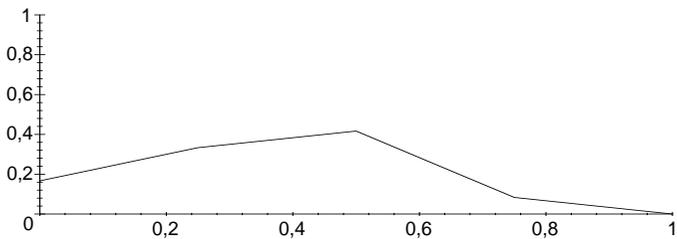


Abbildung A.12: Unscharfe Betrugswahrscheinlichkeit in Fall 1

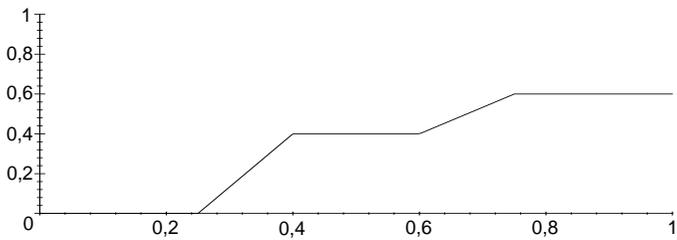


Abbildung A.13: Unscharfe Versicherungsbelastung in Fall 2

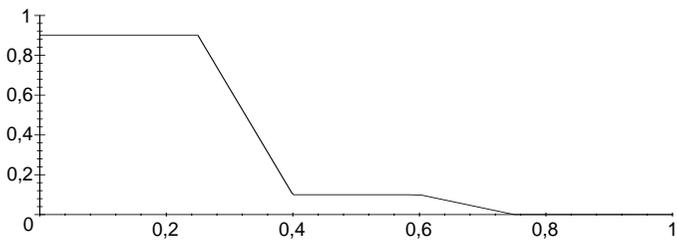


Abbildung A.14: Unscharfe Kreditwürdigkeit in Fall 2

|                           | Fall 1 |        |        |       |       |
|---------------------------|--------|--------|--------|-------|-------|
| Versicherungsbelastung    |        | 50%    | 50%    |       | 0%    |
|                           |        | gering | mittel |       | stark |
| Kreditwürdigkeit          |        | 17%    | 50%    | 33%   |       |
|                           |        | gering | mittel | stark |       |
| Betrugswahrscheinlichkeit | 17%    | 33%    | 42%    | 8%    | 0%    |
|                           | sklein | klein  | mittel | groß  | sgroß |

Tabelle A.6: Grobgranulare Schlüsse in Fall 1

Abb. A.15 eine scharfe Betrugswahrscheinlichkeit berechnen und dann mithilfe eines Schwellenwertes entscheiden, ob der Fall ruhen soll oder nicht. Wählt man die Maximummethode für die Schärfung, so berechnet man eine geschärfte Betrugswahrscheinlichkeit mit maximaler Plausibilität. Wie in Abschnitt 5.4 beschrieben, entspricht die maximale Plausibilität dem Maß  $Z_f$  für die Zuverlässigkeit des Schlusses. Das Maß gibt an, wieviel Prozent der Experten diese scharfe Betrugswahrscheinlichkeit für plausibel halten.

Alternativ können wir die Entscheidung pro oder kontra einer weiteren Untersuchung des Falles auch direkt auf die Wahrscheinlichkeiten der Grundbegriffe

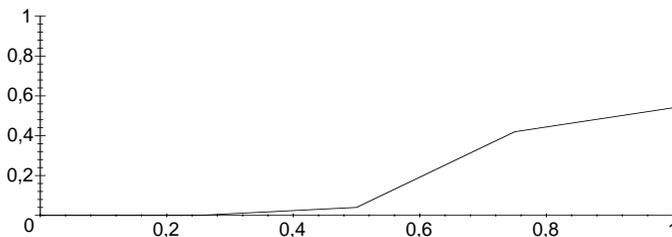


Abbildung A.15: Unscharfe Betrugswahrscheinlichkeit in Fall 2

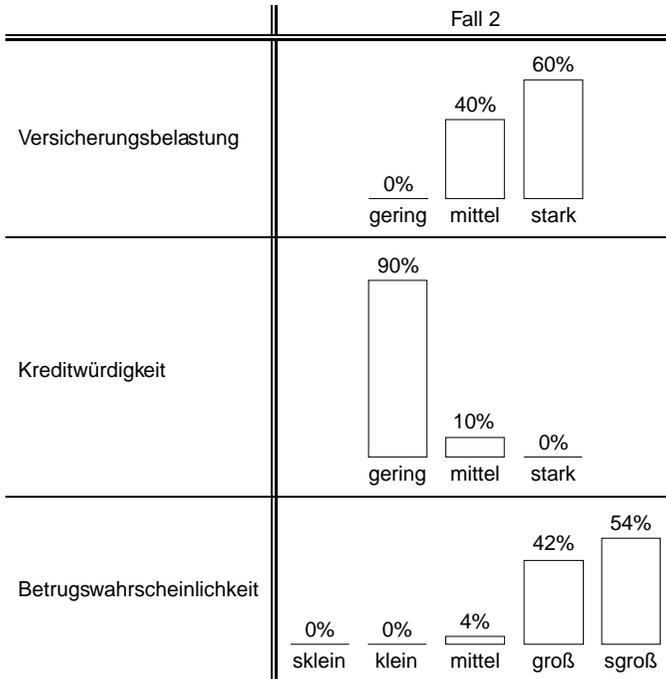


Tabelle A.7: Grobgranulare Schlüsse in Fall 2

stützen. Wir können z. B. einen Schwellenwert  $s$  für die Wahrscheinlichkeit des Grundbegriffs *sehr große* Betrugswahrscheinlichkeit definieren. Wenn dieser überschritten wird, sind mehr als  $100 \cdot s\%$  der Experten der Meinung, dass mit sehr großer Wahrscheinlichkeit ein Betrug vorliegt, und der Fall wird genauer untersucht. Dabei entspricht die Wahrscheinlichkeit des Grundbegriffs *sehr groß* der Zuverlässigkeit  $Z_g$  des Schlusses, s. Abschnitt 5.4.

Welche der beiden Vorgehensweisen in der Praxis gewählt wird, hängt im Wesentlichen von ihrer Akzeptanz innerhalb des Unternehmens ab, in dem das System zur Betrugserkennung eingesetzt wird. Die erste Methode führt zu scharfen Betrugswahrscheinlichkeiten, die ein exaktes Ergebnis suggerieren. Dass dieser Wert in Wirklichkeit sehr unsicher ist – z. B. könnte ein stark abweichender Wert fast ebenso plausibel sein – wird in der Praxis meist verschwiegen. Eine Entscheidung nach dem zweiten Verfahren auf Basis der Grundbegriffe ist ehrlicher, weil man sich die Unsicherheit bewusst eingesteht, und leichter zu verstehen, da man die Schlüsse über die Regeln direkt nachvollziehen kann. Allerdings könnte eine Entscheidung auf Basis eines unscharfen Begriffes bei Laien den Eindruck erwecken, dass sie

nicht fundiert ist, weil der unscharfe Begriff als Aussage zu vage erscheint. Aus Sicht der Theorie sind solche Zweifel aber unbegründet, weil grobgranularer und feingranularer Schluss in beide Richtungen eindeutig ineinander überführt werden können.

Unabhängig von der Entscheidungsmethode muss man einen Schwellenwert festlegen, der die verdächtigen von den unverdächtigen Fällen trennt. In der Praxis hängt dieser letztlich von der Arbeitskapazität der Experten ab, welche die verdächtigen Fälle untersuchen müssen, denn mit dem Schwellenwert steuern wir direkt die Anzahl der verdächtigen Fälle. Man wird den Schwellenwert dementsprechend so einstellen, dass die Anzahl der potentiellen Betrugsfälle von den Experten bewältigt werden kann.

|   |  | Fall 1 |       |        |      |       |
|---|--|--------|-------|--------|------|-------|
| Beitrag der<br>Versicherungsbelastung                     |  |        |       |        |      |       |
| (Unwissen über<br>Kreditwürdigkeit)                       |  | 17%    | 33%   | 33%    | 17%  | 0%    |
|   |  | sklein | klein | mittel | groß | sgroß |
| Beitrag der<br>Kreditwürdigkeit                           |  |        |       |        |      |       |
| (Unwissen über<br>Versicherungsbelastung)                 |  | 11%    | 22%   | 28%    | 33%  | 6%    |
|   |  | sklein | klein | mittel | groß | sgroß |
| Versicherungsbelastung<br>und Kreditwürdigkeit<br>bekannt |  |        |       |        |      |       |
|   |  | 17%    | 33%   | 42%    | 8%   | 0%    |
|   |  | sklein | klein | mittel | groß | sgroß |

Tabelle A.8: Grobgranulare Betrugswahrscheinlichkeiten in Fall 1 bei Unwissen über Versicherungsbelastung bzw. Kreditwürdigkeit

Um die Entscheidung für oder gegen die weitere Untersuchung eines Versicherungsfalls besser zu verstehen und zu rechtfertigen, können wir den Einfluss der Versicherungsbelastung und der Kreditwürdigkeit auf den Schluss untersuchen. Wir betrachten zunächst den Beitrag der Versicherungsbelastung zu den unscharfen Betrugswahrscheinlichkeiten. Dazu nehmen wir an, dass keine Informationen über die Kreditwürdigkeit vorliegen. Formal bedeutet dies, dass die Kreditwürdigkeit gleichverteilt ist:  $1/3$  gering,  $1/3$  mittel,  $1/3$  stark. Wir benutzen

die gleichverteilte Kreditwürdigkeit als unscharfe Eingabe der zweiten Regelstufe und schließen mit der Versicherungsbelastung auf die Betrugswahrscheinlichkeit. Entsprechend gehen wir vor, wenn wir den Anteil der Kreditwürdigkeit an der Betrugswahrscheinlichkeit bestimmen.

Das Ergebnis in Fall 1 sehen wir in Tabelle A.8. Während wir aus der Versicherungsbelastung, also bei Unwissen der Kreditwürdigkeit, eine mittlere oder kleine Betrugswahrscheinlichkeit ableiten, ist die Betrugswahrscheinlichkeit aus Sicht der Kreditwürdigkeit groß bis mittel. Zusammengenommen ist die Wahrscheinlichkeit einer mittleren Betrugswahrscheinlichkeit am größten, wie wir in der dritten Zeile in Tabelle A.8 sehen, die wir bereits aus Tab. A.6 kennen. Den größeren Anteil dazu liefert offenbar die Versicherungsbelastung. Insbesondere ist auch die Versicherungsbelastung der Grund dafür, dass eine sehr große Betrugswahrscheinlichkeit ausgeschlossen werden kann.

Falls ein Kunde bei dem Unternehmen zwar versichert aber kein Bankkunde ist, geht man nach dem gleichen Schema vor. Wir haben keine Informationen über die Kreditwürdigkeit und müssen daher eine Gleichverteilung annehmen. Über die Betrugswahrscheinlichkeit wird dann nur aufgrund der Versicherungsbelastung entschieden.

### A.3 Bewertung

Ein wesentlicher Vorteil unseres Ansatzes zum Schließen mit unscharfen Begriffen gegenüber anderen Ansätzen ist, dass unscharfe Schlüsse gleichzeitig auf grobgranularer Ebene und auf feingranularer Ebene vorliegen. Bei anderen Ansätzen wie dem Prinzip von Kombination/Projektion ist der Schluss immer eine unscharfe Menge auf feingranularer Ebene. Mit der unscharfen Menge selbst kann man in Anwendungen aber meist nichts anfangen. Wir können sie nicht mit den Grundbegriffen veranschaulichen, welche die Konsequenzen der Regeln bilden. Damit verstehen wir den feingranularen Schluss auf grobgranularer Ebene nicht. Stattdessen werden die unscharfe Menge geschärft und der scharfe Wert weiter verarbeitet. Im Beispiel haben wir auf diese Weise eine scharfe Betrugswahrscheinlichkeit bestimmt und daraus geschlossen, ob der Versicherungsfall betrugsverdächtig ist oder nicht. Für viele Anwendungen ist ein scharfer Schluss ausreichend oder wie bei unscharfen Regeln sogar zwingend. Er suggeriert aber eine Genauigkeit, die nicht gegeben ist.

Ein Schluss auf grobgranularer Ebene ist aus zwei Gründen wichtig. Erstens können wir statt eines geschärften, feingranularen Schlusses den grobgranularen Schluss für die weitere Verarbeitung benutzen. Wir haben im Beispiel gesehen, wie wir auf der Basis einer grobgranularen Aussage über den Betrugsverdacht

entscheiden können. Wir gestehen in diesem Fall die Unsicherheit der Aussage ein, die wir bei der geschärften Betrugswahrscheinlichkeit unterschlagen. Ob man mit grobgranularen oder geschärften Schlüssen arbeitet, hängt von der Anwendung und der Akzeptanz der einen und der anderen Möglichkeit ab. Bei anderen Verfahren zum unscharfen Schließen ist diese Wahlmöglichkeit nicht gegeben.

Ein grobgranularer Schluss ist zweitens wichtig, weil er leicht verständlich ist. Wir beschreiben ihn durch die Grundbegriffe, die wir auch in den Regeln verwendet haben. Damit ist ein Schluss über die Regeln unmittelbar nachvollziehbar.

Wegen der konkreten Semantik unseres Ansatzes können wir die Zuverlässigkeit eines Schlusses mit einer messbaren Größe belegen. Im Beispiel der Betrugserkennung entspricht sie der relativen Anzahl der Experten, die diesen Schluss unterstützen. Andere Ansätze erlauben zwar auch solche Angaben, so wird wie bei uns oft der Zugehörigkeitsgrad des geschärften Schlusses benutzt, mangels einer konkreten Semantik können wir diesem Wert aber keine Bedeutung zuordnen.

Insgesamt sehen wir am Beispiel der Betrugserkennung, dass unser Ansatz zu Ergebnissen führt, die leichter verständlich sind, mehr Flexibilität bei der Auswertung bieten und deren Zuverlässigkeit messbar ist.

## Anhang B

---

# Mathematischer Anhang

## B.1 Maßtheorie

In diesem Abschnitt stellen wir jene elementaren Begriffe der Wahrscheinlichkeitstheorie zusammen, die in der vorliegenden Arbeit benötigt werden. Für ausführliche Darstellungen sei auf Lehrbücher wie [Hin72, LW85] verwiesen.

**Definition 8** Sei  $\Omega$  eine nichtleere Menge. Ein System  $\mathcal{A}$  von Teilmengen von  $\Omega$  heißt  $\sigma$ -Algebra über  $\Omega$ , falls

$$(i) \quad \Omega \in \mathcal{A}$$

$$(ii) \quad \text{Für } A \in \mathcal{A} \text{ gilt } A^c \in \mathcal{A}.$$

$$(iii) \quad \text{Aus } \forall i \in \mathbb{N} : A_i \in \mathcal{A} \text{ folgt } \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}.$$

In der Arbeit wählen wir die Potenzmenge  $2^\Omega$ , die Menge aller Teilmengen von  $\Omega$ , als  $\sigma$ -Algebra über  $\Omega$ .

**Definition 9** Sei  $\Omega$  eine nichtleere Menge und  $\mathcal{A}$  eine  $\sigma$ -Algebra über  $\Omega$ . Eine Abbildung  $P: \mathcal{A} \rightarrow [0, 1]$  heißt **Wahrscheinlichkeitsmaß**, wenn gilt:

$$(Pi) \quad \forall A \in \mathcal{A} : P(A) \geq 0$$

$$(Pii) \quad P(\Omega) = 1$$

$$(Piii) \quad P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) \text{ für paarweise disjunkte Mengen } A_i \in \mathcal{A}$$

Das Paar  $(\Omega, \mathcal{A})$  heißt **Messraum**, das Tripel  $(\Omega, \mathcal{A}, P)$  heißt **Maßraum**.

Die Eigenschaften (Pi)–(Piii) eines Wahrscheinlichkeitsmaßes werden als Axiome von **KOLMOGOROFF**<sup>1</sup> bezeichnet. Dabei ist Axiom (Pi) schon in der Anforderung enthalten, dass ein Wahrscheinlichkeitsmaß in das Einheitsintervall abbildet. Aus (Pii) und (Piii) folgt mit  $P(\emptyset) = 0$  eine Grenzbedingung, die in der Literatur oft als Axiom angeführt wird.

---

<sup>1</sup>benannt nach dem russischen Mathematiker A. N. Kolmogoroff (geb. 1903)

**Definition 10** Sei  $\Omega$  eine nichtleere Menge und  $\mathcal{A}$  eine  $\sigma$ -Algebra über  $\Omega$ . Eine Abbildung  $U : \mathcal{A} \rightarrow [0, 1]$  heißt **unscharfes Maß**, wenn gilt:

$$(Ui) \quad U(\Omega) = 1 \text{ und } U(\emptyset) = 0$$

$$(Uii) \quad A, B \in \mathcal{A} \text{ mit } A \subseteq B \Rightarrow U(A) \leq U(B)$$

(Uiii) Für jede monotone Mengenfamilie  $\{A_i \in \mathcal{A} \mid \forall i : A_i \subseteq A_{i+1} \vee \forall i : A_i \supseteq A_{i+1}\}$  gilt  $\lim_{i \rightarrow \infty} U(A_i) = U(\lim_{i \rightarrow \infty} A_i)$ .

Ein Wahrscheinlichkeitsmaß ist ein spezielles unscharfes Maß, denn aus (Piii) folgen (Uii) und (Uiii). Mehr Informationen über unscharfe Maßtheorie findet man z. B. in [KF88].

**Definition 11** Sei  $(\Omega, \mathcal{A}, P)$  ein Maßraum,  $A, B \in \mathcal{A}$ . Gilt  $P(A) > 0$  so heißt

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \quad (\text{B.1})$$

die **bedingte Wahrscheinlichkeit** von  $B$  unter der Bedingung  $A$ .

**Definition 12** Sei  $(\Omega, \mathcal{A}, P)$  ein Maßraum.

- Eine Abbildung  $X : \Omega \rightarrow \mathbb{R}$  heißt **Zufallsvariable** über  $(\Omega, \mathcal{A}, P)$ , falls

$$\{\omega \in \Omega \mid X(\omega) \in I\} \in \mathcal{A} \quad (\text{B.2})$$

für alle konvexen Mengen  $I \subseteq \mathbb{R}$  gilt.

- Die Abbildung  $F : \mathbb{R} \rightarrow [0, 1]$  mit

$$F(x) = P(X \leq x) = P(\{\omega \in \Omega \mid X(\omega) \leq x\}), \quad x \in \mathbb{R} \quad (\text{B.3})$$

heißt **Verteilungsfunktion** der Zufallsvariablen  $X$ .

- Eine Zufallsvariable  $X$  heißt **stetig verteilt mit der Dichte**  $f$ , falls sich ihre Verteilungsfunktion  $F$  durch eine nichtnegative Funktion  $f : \mathbb{R} \rightarrow \mathbb{R}$  folgendermaßen schreiben lässt:

$$F(x) = \int_{-\infty}^x f(t) dt, \quad x \in \mathbb{R} \quad (\text{B.4})$$

- Ist  $X$  eine stetig verteilte Zufallsvariable mit der Dichte  $f$ , so heißt

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx \quad (\text{B.5})$$

**Erwartungswert** von  $X$ , falls  $\int_{-\infty}^{\infty} |x| \cdot f(x) dx < \infty$ .

- Ist  $X$  eine diskrete Zufallsvariable, die die Werte  $x_1, x_2, \dots$  annehmen kann, so heißt

$$E(X) = \sum_i x_i \cdot P(X = x_i) \quad (\text{B.6})$$

**Erwartungswert** von  $X$ , falls  $\sum_i |x_i| \cdot P(X = x_i)$  konvergiert.

- Ist  $X$  eine Zufallsvariable, für die sowohl  $E(X)$  als auch  $E([X - E(X)]^2)$  existieren, so heißt

$$\text{Var}(X) = E([X - E(X)]^2) \quad (\text{B.7})$$

die **Varianz** von  $X$ .

- Sei  $(X, Y)$  eine zweidimensionale Zufallsvariable. Die Varianzen  $\text{Var}(X)$  und  $\text{Var}(Y)$  mögen existieren und positiv sein. Dann heißt die Größe

$$\text{Cov}(X, Y) = E([X - E(X)] \cdot [Y - E(Y)]) = E(X \cdot Y) - E(X)E(Y) \quad (\text{B.8})$$

die **Kovarianz** von  $X$  und  $Y$ .

## B.2 Funktionsapproximation

In diesem Anhang fassen wir kurz einige Ergebnisse zusammen, die wir für die Analyse unscharfer Mengen benötigen. Bei der Analyse wird eine gegebene Funktion, die Zugehörigkeitsfunktion der unscharfen Menge, durch eine Linearkombination von anderen Funktionen, den Zugehörigkeitsfunktionen der Grundbegriffe, approximiert. Für diesen Zweck benutzen wir Verfahren zur Funktionsapproximation in Prä-Hilberträumen. Unsere Darstellung folgt im Wesentlichen dem Lehrbuch [HH92], in dem auch die Beweise nachgelesen werden können.

Sei  $V$  ein Vektorraum mit innerem Produkt  $\langle f, g \rangle$  und der daraus induzierten Norm  $\|f\| := \langle f, f \rangle^{1/2}$ .  $V$  nennen wir dann *Prä-Hilbertraum*<sup>2</sup>. In unseren Anwendungen werden wir normalerweise den Raum  $V = C([a, b])$  der stetigen Funktionen auf dem abgeschlossenen Intervall  $[a, b] \subseteq \mathbb{R}$  betrachten. Als inneres Produkt benutzen wir

$$\langle f, g \rangle = \int_a^b f(x)g(x)dx, \quad (\text{B.9})$$

---

<sup>2</sup>Prä deshalb, weil er nicht vollständig sein muss; benannt nach dem großen deutschen Mathematiker DAVID HILBERT (1862–1943).

woraus sich als Norm

$$\|f\|_2 = \left( \int_a^b f(x)^2 dx \right)^{1/2} \quad (\text{B.10})$$

ergibt. Gegeben sei nun eine Funktion  $f \in V$ . Wir möchten sie durch eine Linearkombination von linear unabhängigen Funktionen  $g_1, g_2, \dots, g_k \in V$  approximieren:

$$f(x) \approx f'(x) = \sum_{i=1}^k \alpha_i \cdot g_i(x) \quad (\text{B.11})$$

Die Approximation soll in dem Sinne möglichst gut sein, dass  $\|f - f'\|_2$  als Abstand zwischen  $f$  und  $f'$  minimal ist. Wir suchen also die Bestapproximation von  $f$  in dem linearen Unterraum  $U = \text{span}(g_1, g_2, \dots, g_k) \subseteq V$ . Für die Bestapproximation gilt

**Satz 12 (Charakterisierungssatz)**  $f'$  ist genau dann Bestapproximation aus  $U$  an  $f \in V$ , wenn  $\langle f - f', g \rangle = 0$  für alle  $g \in U$  gilt.

Die Gewichte  $\alpha_i$  der Bestapproximation erhalten wir durch

**Satz 13 (Normalgleichungen)** Die Gewichte  $\alpha_i$  der Bestapproximation  $f'$  lösen die Normalgleichungen

$$\sum_{i=1}^k \alpha_i \langle g_i, g_j \rangle = \langle f, g_j \rangle \quad (\text{B.12})$$

für  $1 \leq j \leq k$ . Sie sind eindeutig bestimmt.

In der Praxis wird man oft auf die diskrete Version der Funktionapproximation zurückgreifen, die *Methode der kleinsten Quadrate*. Dabei sind  $n$  Wertepaare  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$  mit paarweise verschiedenen Stützstellen  $x_j$  gegeben. Wir suchen eine Linearkombination von linear unabhängigen Funktionen  $g_1, g_2, \dots, g_k$ , welche die Werte  $y_j$  an den Stellen  $x_j$  möglichst gut annähert:

$$f(x_j) \approx f'(x_j) = \sum_{i=1}^k \alpha_i \cdot g_i(x_j) \quad (\text{B.13})$$

Wir gehen diese Aufgabe genauso an wie bei der kontinuierlichen Approximation und definieren einen geeigneten Prä-Hilbertraum. Wir wählen den euklidischen Raum  $V = \mathbb{R}^n$  mit innerem Produkt

$$\langle \vec{u}, \vec{v} \rangle = \sum_{j=1}^n u_j \cdot v_j \quad (\text{B.14})$$

für  $\vec{u} = (u_1, u_2, \dots, u_n)^t, \vec{v} = (v_1, v_2, \dots, v_n)^t \in \mathbb{R}^n$ . Für die Norm ergibt sich

$$\|\vec{u}\| = \|\vec{u}\|_2 = \left( \sum_{j=1}^n u_j \cdot v_j \right)^{1/2} \quad (\text{B.15})$$

Seien nun  $\vec{y} = (y_1, y_2, \dots, y_n)^t, \vec{g}_i = (g_i(x_1), g_i(x_2), \dots, g_i(x_n))^t$  und  $\vec{f}^t = \sum_{i=1}^k \alpha_i \vec{g}_i$ . Dann lautet die Approximationsaufgabe, die Gewichte  $\alpha_i$  so zu bestimmen, dass der Abstand  $\|\vec{y} - \vec{f}^t\|_2$  minimal ist. Wie beim kontinuierlichen Problem betrachten wir für die Lösung der Aufgabe die Normalgleichungen.

**Satz 14 (Diskrete Normalgleichungen)** *Die Gewichte  $\alpha_i$  der Bestapproximation  $f^t$  lösen die Normalgleichungen*

$$\sum_{i=1}^k \alpha_i \langle \vec{g}_i, \vec{g}_j \rangle = \langle \vec{y}, \vec{g}_j \rangle \quad (\text{B.16})$$

für  $1 \leq j \leq k$ . Die Lösung ist genau dann eindeutig, wenn die Vektoren  $\vec{g}_1, \vec{g}_2, \dots, \vec{g}_k$  linear unabhängig sind.

Unabhängig davon, ob wir die Approximationsaufgabe diskret oder kontinuierlich stellen, müssen wir ein lineares Gleichungssystem lösen, um die gesuchten Gewichte  $\alpha_i$  zu bestimmen.



---

# Literaturverzeichnis

- [Ari96] C. ARIMAN: *Unscharfes Schließen: Theorie und Praxis*. Diplomarbeit, Universität Karlsruhe (TH), 1996.
- [Bez73] J. C. BEZDEK: *Fuzzy Mathematics in Pattern Classification*. Dissertation, Cornell University, 1973.
- [Bez81] J. C. BEZDEK: *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, 1981.
- [Bez92] J. C. BEZDEK: *Fuzzy models for pattern recognition : methods that search for structures in data – a selected reprint volume*. IEEE Press, 1992.
- [BH94] M. BROWN und C. J. HARRIS: *Neurofuzzy adaptive modelling and control*. Prentice Hall, 1994.
- [BH95] M. BROWN und C. J. HARRIS: *A perspective and critique of adaptive neurofuzzy systems used in modelling and control applications*. Intern. J. Neural Systems, 6(2):197–220, 1995.
- [BMP95] J. F. BALDWIN, T. P. MARTIN und B. W. PILSWORTH: *Fril: fuzzy and evidential reasoning in artificial intelligence*. Wiley, New York, 1995.
- [Bor00] C. BORGELT: *Data Mining with Graphical Models*. Dissertation, Otto-von-Guericke-Universität Magdeburg, 2000.
- [CDP96] D. CAYRAC, D. DUBOIS und H. PRADE: *Handling Uncertainty with Possibility Theory and Fuzzy Sets in a Satellite Fault Diagnosis Application*. IEEE Transactions on Fuzzy Systems, 4(3):251–269, 1996.
- [CFP82] M. CAYROL, H. FARRENY und H. PRADE: *Fuzzy pattern matching*. Kybernetes, 11:103–116, 1982.
- [CP96] H.-P. CHEN und T.-M. PARNG: *A new approach of multi-stage fuzzy logic inference*. Fuzzy Sets and Systems, 78:51–72, 1996.
- [dB78] C. DE BOOR: *A Practical Guide to Splines*. Springer-Verlag, 1978.
- [DH73] R. DUDA und P. HART: *Pattern Classification and Scene Analysis*. Wiley, New York, 1973.

- [DHR93] D. DRIANKOV, H. HELLENDORF und M. REINFRANK: *An Introduction to Fuzzy Control*. Springer, Berlin, 1993.
- [DK96] J. A. DICKERSON und B. KOSKO: *Fuzzy Function Approximation with Ellipsoidal Rules*. IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics, 26(4):542–560, 1996.
- [DMP93] D. DUBOIS, S. MORAL und H. PRADE: *A Semantics for Possibility Theory based on Likelihoods*. Technischer Bericht, CEC-ESPRIT III BRA 6156 DRUMS II, 1993. Annual Report.
- [DP83] D. DUBOIS und H. PRADE: *On distances between fuzzy points and their use for plausible reasoning*. In: *Proc. of IEEE Intern. Conference on Cybernetics and Society*, S. 300–303, Bombay, India, 1983.
- [DP88] D. DUBOIS und H. PRADE: *Possibility theory: an approach to computerized processing of uncertainty*. Plenum Press, New York, 1988.
- [DP91a] D. DUBOIS und H. PRADE: *Fuzzy sets in approximate reasoning, Part 1: Inference with possibility distributions*. Fuzzy Sets and Systems, 40:143–202, 1991.
- [DP91b] D. DUBOIS und H. PRADE: *Fuzzy sets in approximate reasoning, Part 2: Logical Approaches*. Fuzzy Sets and Systems, 40:203–244, 1991.
- [DP97] D. DUBOIS und H. PRADE: *The three semantics of fuzzy sets*. Fuzzy Sets and Systems, 90:141–150, 1997.
- [Far93] G. E. FARIN: *Curves and surfaces for computer aided geometric design: a practical guide*. Academic Press, Boston, 3. Aufl., 1993.
- [Geb92] J. GEBHARDT: *Ein integrierender Ansatz zur possibilistischen Inferenz in wissensbasierten Systemen*. Dissertation, TH Braunschweig, 1992.
- [GG89] I. GATH und A. B. GEVA: *Unsupervised Optimal Fuzzy Clustering*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 11(7):773–781, 1989.
- [GK79] D. E. GUSTAFSON und W. C. KESSEL: *Fuzzy Clustering with a Fuzzy Covariance Matrix*. In: *Proc. IEEE CDC*, S. 761–766, San Diego, CA, 1979.
- [GK91] J. GEBHARDT und R. KRUSE: *The context model—a uniform approach to vagueness and uncertainty*. In: *Proc. 4th IFSA Congress: Computer, Management and System Science*, S. 82–85, Bruxelles, 1991.

- [GK93a] J. GEBHARDT und R. KRUSE: *The context model—an integrating view of vagueness and uncertainty*. Intern. Journal of Approximate Reasoning, 9:283–314, 1993.
- [GK93b] J. GEBHARDT und R. KRUSE: *A New Approach to Semantic Aspects of Possibilistic Reasoning*. In: M. CLARKE, R. KRUSE und S. MORAL (Hrsg.): *Symbolic and Quantitative Approaches to Reasoning and Uncertainty, ECSQARU '93 (LNCS 747)*, S. 151–159. Springer-Verlag, 1993.
- [Got89] S. GOTTWALD: *Mehrwertige Logik (Many-valued logic) – Eine Einführung in Theorie und Anwendungen*. Akademie-Verlag, Berlin, 1989.
- [GRL00] A. GRAUEL, I. RENNERS und L. A. LUDWIG: *Knowledge Exploration for Toxicity Prediction by Using Genetic Optimized B-Spline Networks*. In: *Proc. 6th Intern. Symposium on Artificial Intelligence and Mathematics*, Fort Lauderdale (USA), 2000.
- [Haj99] P. HAJEK: *Metamathematics of Fuzzy Logic*. Kluwer Academic Publications, 1999.
- [HBH<sup>+</sup>98] E. HORVITZ, J. BREESE, D. HECKERMAN, D. HOVEL und K. ROMMELSE: *The Lumiere Project: Bayesian User Modeling for Inferring the Goals and Needs of Software Users*. In: *Proceedings of Conf. on Uncertainty in Artificial Intelligence (UAI-98)*, 1998.
- [HH92] G. HÄMMERLIN und K.-H. HOFFMANN: *Numerische Mathematik*. Springer-Verlag, 3. Aufl., 1992.
- [HHG96] K. HEESCHE, W. HAUPTMANN und K. GOSER: *A neural net architecture with functional equivalence to fuzzy systems adaptive complexity*. In: *EUFIT '96*, S. 772–776, Aachen, 1996.
- [Hin72] K. HINDERER: *Grundbegriffe der Wahrscheinlichkeitstheorie*. Springer Verlag, Berlin, 1972.
- [His88] E. HISDAL: *Are grades of membership probabilities?* Fuzzy Sets and Systems, 25:325–348, 1988.
- [HL92] J. HOSCHEK und D. LASSER: *Grundlagen der geometrischen Datenverarbeitung*. Teubner, Stuttgart, 2. Aufl., 1992.
- [Höp99] F. HÖPNER: *Fuzzy cluster analysis : methods for classification, data analysis and image recognition*. Wiley, 1999.

- [Jan92] J.-S. R. JANG: *Adaptive Network Fuzzy Inference System*. Dissertation, University of California, Berkeley, 1992.
- [Jan93] J.-S. R. JANG: *ANFIS: Adaptive-network-based fuzzy inference systems*. IEEE Transactions on Systems, Man, and Cybernetics, 23(3):665–685, 1993.
- [JS93] J.-S. R. JANG und C.-T. SUN: *Functional Equivalence between Radial Basis Function Networks and Fuzzy Inference Systems*. IEEE Trans. on Neural Networks, 4(1):156–159, 1993.
- [KF88] G. J. KLIR und T. A. FOLGER: *Fuzzy Sets, Uncertainty and Information*. Prentice Hall, Englewood Cliffs, 1988.
- [KGK95] R. KRUSE, J. GEBHARDT und F. KLAWONN: *Fuzzy-Systeme*. Teubner, Stuttgart, 2. Aufl., 1995.
- [KH93] L. T. KÓCZY und K. HIROTA: *Approximate reasoning by linear rule interpolation and general approximation*. Intern. Journal of Approximate Reasoning, 9:197–225, 1993.
- [Kie97] H. KIENDL: *Fuzzy control methodenorientiert*. Oldenbourg Verlag, 1997.
- [KK93] F. KLAWONN und R. KRUSE: *Equality relations as a basis for fuzzy control*. Fuzzy Sets and Systems, 54:147–156, 1993.
- [KM95] H. M. KIM und J. MENDEL: *Fuzzy Basis Functions: Comparisons with Other Basis Functions*. IEEE Transactions on Fuzzy Systems, S. 158–168, 1995.
- [Kos92a] B. KOSKO: *Fuzzy associative memory systems*. In: A. KANDEL (Hrsg.): *Fuzzy expert systems*, S. 135–164. CRC Press Inc., 1992.
- [Kos92b] B. KOSKO: *Neural networks and fuzzy systems: A dynamical systems approach to machine intelligence*. Prentice Hall, Englewood Cliffs, N.J., 1992.
- [Kos94] B. KOSKO: *Fuzzy Systems as universal approximators*. IEEE Transactions On Computers, 43(11):1329–1333, 1994.
- [LH95] J. LEICHTFRIED und M. HEISS: *Ein kennfeldorientiertes Konzept für Fuzzy-Regler*. Automatisierungstechnik, 43:31–40, 1995.
- [LTT98] S. LEHMKE, K.-H. TEMME und H. THIELE: *Reducing the number of inference steps for multiple-stage fuzzy if-then rule bases*. Technischer Bericht, Universität Dortmund, 1998.

- [LW85] J. LEHN und H. WEGMANN: *Einführung in die Statistik*. Teubner Studienbücher, 1985.
- [MA75] E. H. MAMDANI und S. ASSILIAN: *An experiment in linguistic synthesis with a fuzzy logic controller*. Intern. Journal of Man–Machine Studies, 7, 1975.
- [Mat75] G. MATHERON: *Random Sets and Integral Geometry*. Wiley, New York, 1975.
- [MC89] R. MARTIN-CLOUAIRE: *Semantics and Computation of the Generalized Modus Ponens: The Long Paper*. Intern. Journal of Approximate Reasoning, 3:195–217, 1989.
- [Mik99] R. MIKUT: *Modellgestützte on-line Stabilitätsüberwachung komplexer Systeme auf der Basis unscharfer Ljapunov-Funktionen*. Dissertation, Universität Karlsruhe, VDI-Fortschritt-Bericht Nr. 757, Reihe 8, VDI-Verlag, 1999.
- [MS89] P. MAGREZ und P. SMETS: *Fuzzy Modus Ponens: A New Model Suitable for Applications in Knowledge-Based Systems*. Intern. Journal of Intelligent Systems, 4:181–200, 1989.
- [NK98] D. NAUCK und R. KRUSE: *How the Learning of Rule Weights Affects the Interpretability of Fuzzy Systems*. In: *Proc. IEEE Intern. Conference on Fuzzy Systems (FUZZ-IEEE'98)*, S. 1235–1240, Anchorage, Alaska, 1998.
- [NK99] D. NAUCK und R. KRUSE: *Neuro-Fuzzy systems for function approximation*. Fuzzy Sets and Systems, 101:261–271, 1999.
- [Pea88] J. PEARL: *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann, 1988.
- [Ped91] W. PEDRYCZ: *Processing in relational structures: Fuzzy relational equations*. Fuzzy Sets and Systems, 40:77–106, 1991.
- [PG90] T. POGGIO und F. GIROSI: *Networks for approximation and learning*. Proc. of the IEEE, 78(9):1481–1497, 1990.
- [Pow87] M. J. D. POWELL: *Radial basis functions for multivariable interpolation: A review*. In: J. C. MASON und M. G. COX (Hrsg.): *Algorithms for Approximation*, S. 143–167. Oxford University Press, 1987.
- [RN95] S. RUSSELL und P. NORVIG: *Artificial Intelligence, A Modern Approach*. Prentice Hall, New Jersey, 1995.

- [SB94] J. STOER und R. BULIRSCH: *Numerische Mathematik 1*. Springer-Verlag, 1994.
- [SH98] P. SUBAŠIĆ und K. HIROTA: *Similarity rules and gradual rules for analogical and interpolative reasoning with imprecise data*. *Fuzzy Sets and Systems*, 96(1):53–75, 1998.
- [Sha76] G. SHAFER: *A Mathematical Theory of Evidence*. Princeton University Press, London, 1976.
- [SK00] K. SCHMID und K. KREBS: *Modellierung mit dynamischen Fuzzy-Systemen unter Berücksichtigung von Vorwissen*. In: *Computational Intelligence im industriellen Einsatz*, Band 1526, S. 37–42. VDI-Verlag, 2000.
- [SP98] R. SUN und T. PETERSON: *A subsymbolic+symbolic model for learning sequential navigation*. In: *Fifth Intern. Conference of the Society for Adaptive Behavior (SAB'98)*, Zürich, 1998. MIT Press.
- [Spo94] M. SPOTT: *Unscharfes Schließen mit komplexen Wissensbasen*. Diplomarbeit, Universität Karlsruhe (TH), Juli 1994.
- [Spo99] M. SPOTT: *A theory of possibility distributions*. *Fuzzy Sets and Systems*, 102(2):135–155, 1999.
- [SR98] M. SPOTT und M. RIEDMILLER: *Improving a priori control knowledge by reinforcement learning*. In: *Proceedings in Artificial Intelligence: Fuzzy-Neuro-Systems '98*, S. 146–153, München, 1998. Infix-Verlag.
- [SSLR99] R. SCHOKNECHT, M. SPOTT, F. LIEKWEG und M. RIEDMILLER: *Search Space Reduction for Strategy Learning in Sequential Decision Processes*. In: *Proc. of Iconip99*, Perth, Australia, 1999.
- [SSR00a] R. SCHOKNECHT, M. SPOTT und M. RIEDMILLER: *Design of self-learning controllers using Fynesse*. In: T. FURUHASHI, S. TANO und H.A. JACOBSEN (Hrsg.): *Deep Fusion of Computational and Symbolic Processing*. Physica-Verlag, 2000.
- [SSR00b] R. SCHOKNECHT, M. SPOTT und M. RIEDMILLER: *FYNESSE: A new architecture for sequential decision problems*. In: *Computational Intelligence im industriellen Einsatz*, Band 1526, S. 109–118. VDI-Verlag, 2000.

- [SW96] M. SPOTT und J. WEISBROD: *A new approach to the adaptation of fuzzy relations*. In: *Proc. of EUFIT'96*, Band 2, S. 782–786, Aachen, 1996.
- [Tho95] S. F. THOMAS: *Fuzziness and Probability*. ACG Press, 1995.
- [TS84] T. TAKAGI und M. SUGENO: *Derivation of fuzzy control rules from human operator's control applications*. In: E. SANCHEZ (Hrsg.): *IFAC Proceedings-Series*, 6, S. 55–60, Oxford, 1984.
- [vAJ00] C. VON ALTROCK und D. JUNG: *Fuzzy Logic Applications*. URL: <http://www.fuzzytech.de>, September 2000.
- [Wan94] L. WANG: *Adaptive Fuzzy Systems and Control*. Prentice Hall, 1994.
- [Wei95a] J. WEISBROD: *Fuzzy Control Revisited — Why Is It Working?* In: P. P. WANG (Hrsg.): *Advances in Fuzzy Theory and Technology, Vol. III*, S. 219–244. Bookwrights, Durham (NC), 1995.
- [Wei95b] J. WEISBROD: *Unschärfe Schließen*. Dissertation, Universität Karlsruhe (TH), 1995.
- [Wei96] J. WEISBROD: *A Combined Approach to Fuzzy Reasoning*. In: *Proc. of EUFIT'96*, Band 1, S. 554–557, Aachen, 1996.
- [WM92] L. WANG und J.M. MENDEL: *Fuzzy Basis Functions, Universal Approximation, and Orthogonal Least-Squares Learning*. *IEEE Transactions on Neural Networks*, 3(5), 1992.
- [WS00] J. WEISBROD und M. SPOTT: *Conditional constraints, implication based rules, and possibilistic rule bases: Are they any good?* Technischer Bericht 2000-23, Universität Karlsruhe, Fakultät für Informatik, 2000.
- [YT97] D. S. YEUNG und E. C. C. TSANG: *Weighted fuzzy production rules*. *Fuzzy Sets and Systems*, 88:299–313, 1997.
- [Zad65] L. A. ZADEH: *Fuzzy Sets*. *Information and Control*, 8:338–353, 1965.
- [Zad73] L. A. ZADEH: *Outline of a new approach to the analysis of complex systems and decision processes*. *IEEE Trans. on System, Man, and Cybernetics*, 3:28–44, 1973.
- [Zad75] L. A. ZADEH: *The concept of a linguistic variable and its application to approximate reasoning*. *Information Sciences*, 8,9:199–249, 301–357, 43–80, 1975.

- [Zad78] L. A. ZADEH: *Fuzzy sets as a basis for a theory of possibility*. Fuzzy Sets and Systems, 1:3–28, 1978.
- [Zad79] L. A. ZADEH: *A theory of approximate reasoning*. In: J.E. HAYES, D. MITCHIE und L.I. MIKULICH (Hrsg.): *Machine Intelligence*, S. 149–194. Wiley, New York, 1979.
- [Zad96] L. A. ZADEH: *Fuzzy logic = computing with words*. IEEE Transactions on Fuzzy Systems, 4(2):103–111, 1996.
- [Zad97] L. A. ZADEH: *Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic*. Fuzzy Sets and Systems, 90:111–127, 1997.
- [ZK96] J. ZHANG und A. KNOLL: *Constructing Fuzzy Controllers with B-Spline Models*. In: *FUZZ-IEEE '96*, New Orleans, USA, 1996.
- [ZK99] J. ZHANG und A. KNOLL: *Designing fuzzy controllers by rapid learning*. Fuzzy Sets and Systems, 101(2):287–301, 1999.
- [ZS94] X.-J. ZENG und M. G. SINGH: *Approximation Theory of Fuzzy Systems—SISO Case*. IEEE Transactions on Fuzzy Systems, 2(2):162–176, 1994.

---

# Eigene Veröffentlichungen

- [Spo94] M. SPOTT: *Unscharfes Schließen mit komplexen Wissensbasen*. Diplomarbeit, Universität Karlsruhe (TH), Juli 1994.
- [SW96] M. SPOTT und J. WEISBROD: *A new approach to the adaptation of fuzzy relations*. In: *Proc. of EUFIT'96*, Band 2, S. 782–786, Aachen, 1996.
- [RSW97] M. RIEDMILLER, M. SPOTT und J. WEISBROD: *First Results on the Application of the Fynesse Control Architecture*. In: *IEEE 1997 Intern. Aerospace Conference*, Band 2, S. 421–434, Aspen, USA, 1997.
- [GSWR97] G. GOOS, M. SPOTT, J. WEISBROD und M. RIEDMILLER: *Interpretation und Adaption unscharfer Relationen innerhalb einer hybriden selbstlernenden Steuerungsarchitektur*. In: *Proceedings in Artificial Intelligence: Fuzzy-Neuro-Systeme '97*, S. 332–339, Soest, 1997. Infix-Verlag.
- [BS97] E. BOLTEN und M. SPOTT: *Fuzzy Rule Extraction from Fuzzy Relations*. In: *Proc. of EUFIT'97*, S. 1019–1023, Aachen, 1997.
- [SR98] M. SPOTT und M. RIEDMILLER: *Improving a priori control knowledge by reinforcement learning*. In: *Proceedings in Artificial Intelligence: Fuzzy-Neuro-Systems '98*, S. 146–153, München, 1998. Infix-Verlag.
- [Spo98] M. SPOTT: *Using classic approximation techniques for approximate reasoning*. In: *FUZZ-IEEE '98*, S. 909–914, Anchorage, USA, 1998.
- [SSR<sup>+</sup>98] R. SCHOKNECHT, M. SPOTT, M. RIEDMILLER, G. GOOS und W. MENZEL: *Selbständiges Lernen von Regelungen mit Fynesse*. In: *Anwenderseminar Fuzzy Karlsruhe '98*. IAR Karlsruhe, 1998.
- [Spo99] M. SPOTT: *A theory of possibility distributions*. *Fuzzy Sets and Systems*, 102(2):135–155, 1999.
- [GSR99] T. GAUL, M. SPOTT, M. RIEDMILLER und R. SCHOKNECHT: *Fuzzy-Neuro-Controlled Verified Instruction Scheduler*. In: *Proc. of NAFIPS99*, S. 869–873, New York, 1999.

- [SSR99] M. SPOTT, R. SCHOKNECHT und M. RIEDMILLER: *Approaches for the integration of a priori knowledge into an autonomously learning control architecture*. In: *Proc. of EUFIT'99*, Aachen, 1999.
- [SSLR99] R. SCHOKNECHT, M. SPOTT, F. LIEKWEG und M. RIEDMILLER: *Search Space Reduction for Strategy Learning in Sequential Decision Processes*. In: *Proc. of Iconip99*, Perth, Australia, 1999.
- [SSR00b] R. SCHOKNECHT, M. SPOTT und M. RIEDMILLER: *FYNESSE: A new architecture for sequential decision problems*. In: *Computational Intelligence im industriellen Einsatz*, Band 1526, S. 109–118. VDI-Verlag, 2000.
- [RSW99] M. RIEDMILLER, M. SPOTT und J. WEISBROD: *Fynesse: A hybrid architecture for selflearning control*. In: I. CLOETE und J. ZURADA (Hrsg.): *Knowledge-Based Neurocomputing*. MIT Press, 1999.
- [SSR00a] R. SCHOKNECHT, M. SPOTT und M. RIEDMILLER: *Design of self-learning controllers using Fynesse*. In: T. FURUHASHI, S. TANO und H.A. JACOBSEN (Hrsg.): *Deep Fusion of Computational and Symbolic Processing*. Physica-Verlag, 2000.
- [WS00] J. WEISBROD und M. SPOTT: *Conditional constraints, implication based rules, and possibilistic rule bases: Are they any good?* Technischer Bericht 2000-23, Universität Karlsruhe, Fakultät für Informatik, 2000.