

17 Fuzzy-Methoden in der Datenanalyse

Christian Borgelt¹, Jörg Gebhardt² und Rudolf Kruse³

Otto-von-Guericke-Universität Magdeburg
Institut für Wissens- und Sprachverarbeitung
Universitätsplatz 2, 39106 Magdeburg

Kurzfassung. Durch den Einsatz von Fuzzy-Methoden ergeben sich neue Möglichkeiten bei der Analyse impräziser, vager und unscharfer Daten. In diesem Aufsatz geben wir zunächst einen Überblick über Methoden, die eine statistische Analyse von unscharfen Daten (Fuzzy-Daten) erlauben. Wir betrachten dann den Spezialfall mengenwertiger Daten, der in jüngster Zeit in einigen Bereichen der explorativen Datenanalyse eine wichtige Rolle spielt. Derartige Daten lassen sich gut durch das Lernen possibilistischer graphischer Modelle bearbeiten. Wir skizzieren die Theorie dieser Technik und zeigen seine erfolgreiche Anwendung an einem Beispiel.

17.1 Fuzzy-Datenanalyse

Das Gebiet der Datenanalyse entwickelt sich rasch, da wegen der Menge derzeit erfaßter und der Auswertung harrender Daten einfach zu handhabende und leicht zu verstehende Analysetechniken in vielen Bereichen der Industrie, der Medizin etc. benötigt werden. Die *Fuzzy-Datenanalyse* erfreut sich wachsender Beliebtheit, weil

¹ Christian Borgelt studierte an der Technischen Universität Braunschweig Informatik und Physik. Nach dem Diplom in Informatik 1995 arbeitete er als Angestellter der TU Braunschweig im Forschungszentrum Ulm der Daimler-Benz AG in der Arbeitsgruppe Maschinelles Lernen und Data Mining. Seit 1996 ist er Doktorand an der Universität Magdeburg. Zur Zeit beschäftigt er sich mit dem Lernen von probabilistischen und possibilistischen Schlußfolgerungsnetzen, Entscheidungsbäumen und Assoziationsregeln aus Daten.

² Jörg Gebhardt studierte an der Technischen Universität Braunschweig mit Diplomen in Mathematik und Informatik. In seiner Anstellung als wissenschaftlicher Mitarbeiter promovierte er 1992 am Institut für Betriebssysteme und Rechnerverbund und habilitierte sich 1997 für Informatik. Seit Oktober 1998 ist er als freier Unternehmensberater für die Volkswagen AG im Konzernprojekt „EPL : Planung von Fahrzeugen, Eigenschaften und Teilen (K-BKM, BKM)“ tätig. Seine wichtigsten Forschungsgebiete sind Possibilitätstheorie, Data Mining, graphische Modellierung, Data Fusion und Fuzzy-Systeme.

³ Rudolf Kruse studierte an der Technischen Universität Braunschweig Mathematik mit Nebenfach Informatik. Er promovierte 1980 im Rahmen seiner Tätigkeit als wissenschaftlicher Mitarbeiter am Institut für Stochastik und habilitierte sich 1984 für Mathematik. Anschließend arbeitete er in einem Forschungsvertrag mit der Fraunhofergesellschaft über Expertensysteme. 1986 wurde er Professor für Informatik an der TU Braunschweig. Seit 1996 hat er den Lehrstuhl für Praktische Informatik (Neuro-Fuzzy-Systeme) an der Universität Magdeburg inne. Seine Hauptforschungsgebiete sind zur Zeit Neuro- und Fuzzy-Systeme und ihre Anwendung im Bereich des Data Mining.

die im Bereich der Fuzzy-Systeme entwickelten Methoden oft besonders leicht zu verstehen und daher einfach anzuwenden sind — wie ja schon die Entwicklung und der Einsatz der Fuzzy-Regelungstechnik zeigte.

Leider stimmt diese Einschätzung für die Fuzzy-Datenanalyse nur zum Teil, wie wir in diesem Aufsatz zeigen. (Fuzzy-Daten)-Analysen sind in Wahrheit ziemlich kompliziert, da mit ihnen versucht wird, statistische Aussagen aus unscharfen Daten abzuleiten. Nun ist aber schon die statistische Analyse scharfer Daten nicht immer einfach. Berücksichtigt man Unschärfe wie Impräzision und Vagheit, führt dies oft zu ziemlich komplizierten Modellen. Einige Resultate aus diesem Bereich werden im Abschnitt 17.3 beschrieben.

Einfacher sind Fuzzy-(Datenanalysen) durchzuführen, denn hier hat man es meist mit gewöhnlichen (scharfen, höchstens mengenwertigen) Daten zu tun, die mit (einfachen) Fuzzy-Methoden analysiert werden. Zu diesem Bereich gehören z.B. die Fuzzy-Clusteranalyse [2, 26] und das Lernen (possibilistischer) graphischer Modelle (auch (possibilistische) Schlußfolgerungsnetze genannt) aus Daten [14]. In diesem Beitrag beschränken wir uns darauf, letzteres zu skizzieren und das Anwendungspotential dieser Methode für den Bereich der explorativen Datenanalyse aufzuzeigen (siehe Abschnitt 17.5). Im Abschnitt 17.6 wird anhand eines Beispiels der erfolgreiche Einsatz dieser neuen Methode demonstriert.

17.2 Fuzzy-Mengen und ihre Interpretation

In der Praxis — sowohl in der Industrie als auch in der Medizin und in anderen Bereichen — liegen Informationen oft nur in unscharfer sprachlicher Form vor, z.B. in Form von Aussagen wie „Der Patient W. hat hohen Blutdruck.“ Solche Aussagen bezeichnet man auch als *linguistische Daten*. Obwohl Menschen i.a. wenig Probleme haben, mit solchen Daten umzugehen (d.h. die in natürlichsprachlichen Aussagen steckende Information zu verstehen und z.B. beim Treffen von Entscheidungen auszunutzen), ist die mathematische Analyse linguistischer Daten ziemlich kompliziert. Das liegt vor allem daran, daß viele Begriffe natürlicher Sprachen vage sind (Was bedeutet „hoher Blutdruck“ genau?), das klassische statistische Instrumentarium aber auf die Untersuchung präziser Zahlenwerte ausgerichtet ist.

Ein gangbarer Weg, die Schwierigkeiten zu überwinden, besteht darin, sprachliche Ausdrücke wie „hoher Blutdruck“ u.ä. in geeignete Fuzzy-Mengen zu „übersetzen“. Der sprachliche Ausdruck wird dazu durch eine unscharfe Menge (Fuzzy-Menge) z.B. reeller Zahlen dargestellt, wobei „unscharf“ bzw. „fuzzy“ bedeutet, daß es zwar einige Zahlenwerte gibt, die sicher zu der Menge gehören und andere, die sicher nicht zu ihr gehören, dazwischen aber eine „penumbra“ (ein „Halbschatten“) von Zahlenwerten liegt, die nur zu einem gewissen Grade der Menge angehören. Dieser „Halbschatten“ soll die unscharfen Grenzen (der Anwendbarkeit) eines sprachlichen Ausdrucks widerspiegeln.

Natürlich hängt die Wahl der Fuzzy-Menge vom Kontext ab — z.B. wird

„warm“ im Zusammenhang mit Speisen durch eine andere Fuzzy-Menge von Temperaturwerten dargestellt werden müssen als im Zusammenhang mit Kraftfahrzeug-Motoren. Aber auch bei gegebenem Kontext muß die Semantik der Fuzzy-Menge genau festgelegt werden, da zunächst nicht klar ist, was „graduelle Zugehörigkeit“ zu einer Menge bedeuten soll. Wir betrachten in diesem Aufsatz zwei verschiedene Interpretationen von Fuzzy-Mengen [21], nämlich die *physikalische* und die *epistemische* Interpretation.

Im ersten Fall fassen wir die Fuzzy-Menge als Beschreibung eines realen unscharfen Objektes oder Zustandes auf. D.h. die durch die Fuzzy-Menge dargestellte Unschärfe wird als Eigenschaft der physikalischen Realität aufgefaßt und nicht als Eigenschaft etwa der Sprache, die eine scharfe Realität nur unscharf beschreibt. Diese Sichtweise findet man z.B. bei der Analyse von Grauwertbildern: Graustufen zwischen Schwarz und Weiß zeigen in dieser Interpretation unscharfe Objektgrenzen an und sind nicht eine unscharfe Beschreibung scharfer Objektgrenzen. Ein anschauliches Beispiel ist der Begriff „Nase“, der durch eine Fuzzy-Menge von Raumpunkten (oder, bei einem Bild, von Punkten in der Ebene) beschrieben werden könnte. Die Nase eines Menschen geht über in seine Wangen, seine Stirn, etc. Es gibt keine (natürliche) scharfe Grenze, die diese Gesichtspartien voneinander trennt — nicht einmal bei einem ganz bestimmten Menschen, geschweige denn bei allen Menschen in gleicher Weise. In dieser (der physikalischen) Deutung wird eine Fuzzy-Menge als unveränderliches mathematisches Objekt aufgefaßt, da ja die Unschärfe der Realität entspringt und sich folglich nicht z.B. durch zusätzliche Informationen (gewonnen etwa durch genauere Beobachtungen) verringern läßt.

Im zweiten Fall, der epistemischen Interpretation, werden Fuzzy-Mengen benutzt, um Unsicherheit über einen (scharfen) wahren Wert zu modellieren, etwa, weil es die Beobachtungs- und Meßmöglichkeiten nicht zulassen, den wahren Wert eindeutig zu bestimmen. Die durch die Fuzzy-Menge dargestellte Unschärfe spiegelt hier die Unvollkommenheit unserer Erkenntnis der Realität wider (daher *epistemische* Interpretation) und ist nicht Eigenschaft dieser Realität. Diese gilt vielmehr als scharf, d.h. als auf einen bestimmten Zustand festgelegt. So hat z.B. ein Patient zu einem gegebenen Zeitpunkt einen bestimmten Blutdruck. Wir könnten ihn im Prinzip genau messen. Steht uns aber kein Meßgerät zur Verfügung und müssen wir uns daher mit indirekten Anzeichen begnügen (z.B. Rötung des Gesichts o.ä.), so bleibt uns nur übrig, uns mit einer Aussage wie „Der Patient W. hat hohen Blutdruck.“ zu bescheiden. Aber auch eine solche Aussage schränkt die Menge möglicher Werte für den tatsächlichen Blutdruck ein; er läßt einige Werte mehr, andere weniger plausibel erscheinen. In dieser Interpretation ist eine Fuzzy-Menge daher eine Quantifizierung der Möglichkeit verschiedener Zustände oder Ereignisse. Sie ist folglich kein unveränderliches mathematisches Objekt, sondern wird durch neue Informationen, etwa die Beobachtung weiterer indirekter Anzeichen, verändert. In der epistemischen Interpretation werden Fuzzy-Mengen auch (für Wahrscheinlichkeitstheoretiker u.U. etwas irreführend) als *Possibilitätsverteilungen* bezeichnet [56, 9]. In der Tat konkurrieren sie in dieser Interpretation mit einer wahrscheinlichkeitstheoretischen Beschreibung, da ja Wahrscheinlichkeitsverteilungen ebenfalls Quantifizierungen der Möglichkeit von Zuständen oder Ereignissen sind.

Natürlich gibt es Beziehungen zwischen den beiden Interpretationen. So induziert eine Fuzzy-Menge in physikalischer Interpretation eine (sinnvollerweise formal identische) Possibilitätsverteilung. Diese Possibilitätsverteilung gibt an, wie plausibel bestimmte Werte sind, wenn der wahre Wert durch den sprachlichen Ausdruck beschrieben wird, der in die Fuzzy-Menge „übersetzt“ wurde. Z.B. beschreibt die durch die Fuzzy-Menge zum Begriff „Nase“ induzierte Possibilitätsverteilung, wie plausibel es ist, daß in der Aussage „Er hat einen Fleck auf der Nase.“ mit dem Ausdruck „auf der Nase“ bestimmte Raumpunkte im Gesicht eines Menschen gemeint sind. Denn der Fleck hat natürlich eine bestimmte Lage, die aber durch den Ausdruck „auf der Nase“ unscharf beschrieben wird.

Umgekehrt kann man einen Begriff zur Bezeichnung eines (physikalischen) Objektes oder Zustandes definieren, indem man eine Possibilitätsverteilung zum Konzept erhebt. Ein Arzt könnte etwa über ein Meßgerät verfügen, daß ihm eine nicht ganz exakte Beobachtung eines physiologischen Merkmals erlaubt. Er stellt nun fest, daß ein bestimmtes Meßergebnis mit einer bestimmten Krankheit verbunden ist. Daher definiert er ein entsprechendes Symptom, das die Krankheit anzeigt. Dieses Symptom wird nun durch eine Fuzzy-Menge dargestellt, die sich aus der von dem Meßgerät gelieferten Possibilitätsverteilung ergibt (wir hatten ja angenommen, daß es keine exakte Messung des physiologischen Zustandes erlaubt).

17.3 Statistik mit unscharfen Daten

Der Unterschied zwischen den im vorstehenden Abschnitt besprochenen Interpretationen (den man ja auch schon bei der Interpretation von gewöhnlichen Mengen findet) spielt im folgenden eine wichtige Rolle: Will man Fuzzy-Daten statistisch analysieren, so muß man natürlich die Semantik der Fuzzy-Mengen berücksichtigen.

Verwendet man die physikalische Interpretation, so muß man eine Statistik aufbauen, in der statt Zahlen Fuzzy-Mengen zur Beschreibung von Merkmalen verwendet werden. In diesem Fall werden Zufallsprozesse modelliert, durch die z.B. eine Fuzzy-Menge aus einer Menge möglicher Fuzzy-Mengen ausgewählt wird, genauso, wie bei einer gewöhnlichen Zufallsvariable durch einen Zufallsprozeß ein Zahlenwert aus ihrem Wertebereich ausgewählt wird. Man kann folglich im Prinzip die Methoden der klassischen Wahrscheinlichkeitstheorie und Statistik nutzen. Allerdings sind diese Methoden so zu erweitern, daß man z.B. Erwartungswerte, Varianzen etc. für Zufallsvariablen berechnen kann, deren Werte Fuzzy-Mengen sind. Hat man etwa Fuzzy-Mengen über Raumpunkte, die jeweils die Nase einer bestimmten Person beschreiben, und hat man eine Wahrscheinlichkeitsverteilung über den Personen, so möchte man als Erwartungswert eine Fuzzy-Menge erhalten, die eine Art „Durchschnittsnase“ beschreibt.¹ Der Schlüssel zur Erweiterung der Wahrscheinlichkeitstheorie auf Nichtstandarddaten wie Fuzzy-Mengen ist gewöhnlich der Beweis eines

¹ Man beachte, daß eine solche Erwartungswertbildung kein Widerspruch zur Unveränderlichkeit der Fuzzy-Mengen in der physikalischen Interpretation ist, denn die sich als Erwartungswert ergebende Fuzzy-Menge braucht kein reales Objekt zu beschreiben. Beim Würfeln mit einem

„starken Gesetzes der großen Zahlen“. Kann man dieses Gesetz zeigen, so gelingt oft auch der Beweis einer Verallgemeinerung weiterer aus der klassischen Statistik bekannter Sätze.

Eine Erweiterung der Statistik auf mengenwertige Daten gibt es schon seit längerem [40, 29, 50]. Die Daten sind hier (scharfe) Mengen, die Zufälligkeit ihrer Auswahl wird über einen gewöhnlichen Wahrscheinlichkeitsraum (Ω, \mathcal{S}, P) und eine Abbildung $X : \Omega \rightarrow 2^D$ modelliert, wobei D eine geeignete Menge ist, deren Teilmengen als Daten auftreten können. Die Abbildung X wird als *zufällige Menge (random set)* bezeichnet; formal ist sie nichts anderes als eine mengenwertige Zufallsvariable. Ein starkes Gesetz der großen Zahlen wurde für zufällige Mengen schon früh bewiesen [1]. Einen Überblick über neuere Resultate in diesem Bereich gibt [23].

Möchte man nun statt einfacher Mengen möglicher Werte unscharfe Beschreibungen in Form von Fuzzy-Mengen verwenden, so kann man auf die o.g. Arbeiten zurückgreifen und die Ergebnisse von Mengen auf Fuzzy-Mengen erweitern. Viele solcher sogenannten „Fuzzifizierungen“ (Verallgemeinerungen auf Fuzzy-Mengen) wurden von 1975 bis 1985 für unterschiedliche Bereiche der Mathematik — nicht nur für die Statistik — vorgenommen. Die Fuzzifizierung des starken Gesetzes der großen Zahlen für zufällige Mengen wurde von [46] vorgeschlagen und dann in [30] bewiesen. Die zugrundeliegende Theorie ist ziemlich kompliziert, da die Beweistechniken auf Banach-Räume zurückgreifen müssen. Hauptproblem ist die Wahl einer geeigneten Metrik (Abstandsfunktion) für Fuzzy-Mengen. Oft wird ein verallgemeinerter Hausdorff-Abstand benutzt.

Verwendet man statt der physikalischen die epistemische Interpretation von Fuzzy-Mengen, so muß man einen anderen Ansatz wählen, da hier der Wert eines Merkmals nach wie vor ein scharfer Wert ist (wie in der klassischen Statistik) und die Fuzzy-Menge einen Anteil an der Beschreibung der Unsicherheit über seinen Wert hat. Hier geht man davon aus, daß ein gewöhnlicher Zufallsprozeß vorliegt, der wie üblich durch eine Zufallsvariable $U : \Omega \rightarrow \mathbb{R}$ beschrieben werden kann. Man nimmt jedoch an, daß die Werte, die diese Zufallsvariable annimmt, nicht mit der nötigen Präzision beobachtet werden können. Der Beobachter erhält nur Beschreibungen in Form von Fuzzy-Mengen (genauer: Possibilitätsverteilungen). Diese Situation wird durch eine Abbildung $X : \Omega \rightarrow \mathcal{F}(\mathbb{R})$ (wobei $\mathcal{F}(\mathbb{R})$ die Menge aller Fuzzy-Mengen über den reellen Zahlen ist) modelliert. X ist eine Zufallsvariable, deren Werte Fuzzy-Mengen sind, die jeweils eine unscharfe Beobachtung beschreiben. Sie wird auch als *Fuzzy-Zufallsvariable* bezeichnet. In [38] wird die obige Sichtweise erstmals in der Literatur bekannt gemacht. Methodisch ist sie ein Unsicherheitskonzept zweiter Ordnung, da hier die beiden Unsicherheitskonzepte Zufälligkeit (Auswahl des wahren Wertes) und Möglichkeit (unscharfe Beschreibung seiner Beobachtung) gleichzeitig benutzt werden.

Die Idee einer Verallgemeinerung der Statistik auf Fuzzy-Daten in der epistemischen Interpretation besteht darin, eine Fuzzy-Zufallsvariable $X : \Omega \rightarrow \mathcal{F}(\mathbb{R})$ als (unscharfe) Beobachtung einer nicht zugänglichen Zufallsvariablen, des sogenannten Originals U_0 von X , anzusehen. Jede Zufallsvariable $U \in \mathcal{U}$ ist damit als mögliches

gewöhnlichen Würfel ist der Erwartungswert 3.5 der Augenzahl auch kein mögliches Ereignis.

Original von X in Betracht zu ziehen. Die Möglichkeit, daß U Original von X ist, wird mit dem Grad $\inf_{\omega \in \Omega} \{(X(\omega))(U(\omega))\}$ bewertet. (Die Wahl des inf-Operators läßt sich mit dem Prinzip der minimalen Spezifität aus der Possibilitätstheorie begründen [37].) Man erhält so eine Fuzzy-Menge Orig_X auf der Menge \mathcal{U} der Zufallsvariablen

$$\text{Orig}_X : \mathcal{U} \rightarrow [0, 1], \quad U \mapsto \inf_{\omega \in \Omega} \{(X(\omega))(U(\omega))\},$$

die das unscharfe Wissen vollständig beschreibt. (Hier zeigen sich Beziehungen zur Theorie der zufälligen Mengen: Für jedes $\alpha > 0$ ist die α -Niveaumenge von Orig_X gleich der Menge der sogenannten Selektoren der zufälligen Menge $X_\alpha : \Omega \rightarrow 2^R$, $X_\alpha(\omega) = [X(\omega)]_\alpha$ [34].)

Mit Hilfe des bereits erwähnten Prinzips der minimalen Spezifität kann man verschiedene Kernaussagen der Fuzzy-Mengentheorie wie z.B. das Extensionsprinzip ableiten. Mit Hilfe dieses Prinzips wiederum lassen sich Begriffe der klassischen Statistik auf Fuzzy-Mengen verallgemeinern. So kann man z.B. die Abbildung $E : \mathcal{U} \rightarrow \mathbb{R}$, die jeder Zufallsvariablen (im Falle der Existenz) ihren Erwartungswert zuordnet, mit dem Extensionsprinzip auf Fuzzy-Teilmengen von \mathcal{U} erweitern. Für die Fuzzy-Mengen Orig_X erhält man

$$E(\text{Orig}_X)(t) = \sup_{U \in \mathcal{U}: E(U)=t} \left\{ \inf_{\omega \in \Omega} \{(X(\omega))(U(\omega))\} \right\}.$$

Diese Fuzzy-Menge wird als Erwartungswert der Fuzzy-Zufallsvariablen X bezeichnet.

Das starke Gesetz der großen Zahlen für Fuzzy-Zufallsvariablen wurde in [31, 32] bewiesen. Mittlerweile gibt es eine ganze Reihe von Varianten dieses Gesetzes wie z.B. [42, 41, 45] sowie Verallgemeinerungen des zentralen Grenzwertsatzes, des Theorems von Glivenko-Cartelli, Parameterschätzungen und Fuzzy-Tests. Diese Methoden sind in einem Programmsystem namens SOLD (Statistics On Linguistic Data) verfügbar, das im Rahmen einer Kooperation der TU Braunschweig mit der Siemens AG in München implementiert wurde [33, 35].

Allgemein kann man aus der Untersuchung der in diesem Abschnitt angesprochenen Verfahren den Schluß ziehen, daß es sehr rechenaufwendig ist, wenn man versucht, zwei unterschiedliche Ansätze zur Beschreibung unvollkommener Informationen — etwa die Wahrscheinlichkeitstheorie zur Beschreibung von Unsicherheit und die Fuzzy-Mengentheorie zur Beschreibung von Vagheit — parallel zu benutzen. Das gleiche gilt auch für vergleichbare in der Literatur bekannt gewordene Ansätze wie die possibilistischen Mengen [25, 7], das GoM-Modell [43], die Fuzzy-Informationssysteme [52, 22, 21] und die Bayesschen Analysen von Fuzzy-Daten [53]. Viele nützliche Hinweise für die Anwendung der genannten Ansätze findet man in den Monographien [2, 3].

Da, wie gesagt, die parallele Verwendung zweier unterschiedlicher Ansätze zur Modellierung unvollkommener Informationen rechenaufwendig ist, sind Ansätze gefragt, die in der Lage sind, unterschiedliche Aspekte unvollkommener Information in einem gemeinsamen Kalkül darzustellen. Im Falle von Fuzzy-Mengen in der physikalischen Interpretation sind solche Versuche sicher problematisch, bei Fuzzy-Mengen

in der epistemischen Interpretation jedoch leicht möglich: Die Possibilitätstheorie stellt hier einen geeigneten Rahmen bereit. Zwar ließen sich im Prinzip wohl auch alle auftretenden Probleme mit rein probabilistischen Methoden behandeln, doch erscheint die Possibilitätstheorie wegen ihrer Einfachheit oft als sinnvolle Alternative. Das gilt insbesondere in einigen Bereichen der explorativen Datenanalyse (Data Mining) [55], wo man schon aus Komplexitätsgründen oft mit Heuristiken arbeiten muß, so daß eventuelle Verluste gegenüber einer wahrscheinlichkeitstheoretischen Behandlung vernachlässigt werden können.

17.4 Possibilitätsverteilungen und ihre Interpretation

Wie bereits oben gesagt (siehe Abschnitt 17.2), dienen Possibilitätsverteilungen — wie auch die Wahrscheinlichkeitsverteilungen der klassischen Wahrscheinlichkeitstheorie — der Modellierung von Unsicherheit über einen wahren, aber unbekanntem Zustand der Wirklichkeit. Z.B. kann man die Unsicherheit über die (tatsächliche) Körpergröße einer bestimmten, als „groß“ bezeichneten Person P. im Kontext mitteleuropäischer erwachsener Männer darstellen, indem man eine Fuzzy-Menge μ über dem Intervall $[0, 250]$ (cm) als Interpretation des vagen Begriffes „groß“ wählt. Diese kann dann (epistemisch) so gedeutet werden: Wenn jemand sagt, P. sei „groß“, so bringt er zum Ausdruck, daß für jede Körpergröße x aus der Referenzmenge $[0, 250]$ durch den Wert $\mu(x)$ der Möglichkeitsgrad angegeben wird, mit dem P. eine Körpergröße von x cm hat. $\mu(x) = 0$ bedeutet, daß es unmöglich ist, daß P. x cm mißt (wenn man ihn als „groß“ bezeichnet), während $\mu(x) = 1$ bedeutet, daß die Möglichkeit, daß P. x cm groß ist, in keiner Weise eingeschränkt werden kann. Für das obige Beispiel wird man etwa für die Person P. eine Körpergröße von weniger als 160 cm als unmöglich und eine Körpergröße von mehr als 190 cm als uneingeschränkt möglich ansehen (wenn P. als „groß“ bezeichnet wird). Folglich ist $\mu(x) = 0$ für $x \in [0, 160)$ und $\mu(x) = 1$ für $x \in (190, 250]$. Für das Intervall $[160, 190]$ wird man dagegen Zwischenwerte bevorzugen, also Möglichkeitsgrade $\mu(x)$ aus dem offenen Intervall $(0, 1)$ wählen.

Dieser intuitive Zugang zum Begriff der Possibilitätsverteilung ist natürlich nicht ausreichend, um eine präzise Semantik von Possibilitätsverteilungen festzulegen. Insbesondere ist noch unklar, was ein Möglichkeitsgrad zwischen 0 und 1 genau bedeuten soll, da ja der Begriff „möglich“ im alltäglichen, natürlichsprachlichen Gebrauch zweiwertig ist: Entweder etwas ist möglich oder es ist nicht möglich. Ähnlich wie beim Wahrscheinlichkeitsbegriff gibt es daher verschiedene Ansätze zur Festlegung einer Semantik von Possibilitätsverteilungen, auf die wir im folgenden kurz eingehen.

Zuvor bemerken wir jedoch, daß man dem Semantikproblem — wenigstens mathematisch — auch durch einen axiomatischen Zugang über die Einführung subadditiver Maße entgegen gehen kann [9, 10] — genauso wie die Wahrscheinlichkeitstheorie dem Semantikproblem durch die Kolmogoroffsche Axiomatisierung des Wahrscheinlichkeitsbegriffs als additives Maß ausweicht. Daß man in der Possibilitätstheorie

nicht-additive Maße benutzt, liegt im wesentlichen daran, daß man durch Possibilitätsverteilungen Unsicherheit und Impräzision gleichzeitig modelliert, während eine Wahrscheinlichkeitsverteilung nur Unsicherheit bei präzisen Daten beschreibt, wie dies z.B. durch die Elementarität der Werte von Zufallsvariablen zum Ausdruck gebracht wird. Daß eine Possibilitätsverteilung auch Impräzision modelliert, läßt sich sehr gut anhand einer (analog zu dem obigen Beispiel) über dem Intervall $[0, 250]$ (cm) definierten Possibilitätsverteilung μ sehen, die für $x \in [180, 190]$ den Wert 1 annimmt, und sonst 0 ist. Dient die Verteilung μ dazu, die Körpergröße einer Person P. anzugeben, so bedeutet dies, daß es für uneingeschränkt möglich gehalten wird, daß P. zwischen 180 cm und 190 cm mißt, alle anderen Körperhöhen jedoch ausgeschlossen sind. Die Körpergröße wird folglich impräzise (durch ein Intervall möglicher Werte) beschrieben.

Wenden wir uns nun der Semantik von Possibilitätsverteilungen zu. Neben der bereits erwähnten epistemischen Interpretation von Fuzzy-Mengen als Possibilitätsverteilungen [56] ist ein weiterer vielversprechender Weg zu einer Semantik von Possibilitätsverteilungen ihre Interpretation als informationskomprimierte Darstellung einer Datenbank mit Unsicherheit behafteter, eventuell auch impräziser (mengenwertiger) Fallbeispiele. Die unscharfe Beschreibung der Körpergröße von P. könnte man dann etwa auf eine Menge von Betrachtungskontexten beziehen (z.B. Vergleich mit anderen (ungefähr) „gleich großen“ Personen, deren Körpergrößen man durch scharf begrenzte Intervalle angeben kann), um dann mit Hilfe einer Wahrscheinlichkeitsverteilung über diesen Betrachtungskontexten deren Auswahlwahrscheinlichkeit zur (impräzisen) Beschreibung der Körpergröße von P. zu quantifizieren. Für die formale Darstellung empfiehlt sich hier die Verwendung von Zufallsmengen mit der zusätzlichen Semantik der oben erwähnten Betrachtungskontexte. Possibilitätsverteilungen sind bei dieser Sichtweise als (nicht-normalisierte) Ein-Punkt-Überdeckungen von (generalisierten) Zufallsmengen interpretierbar [44, 24]. Dies führt im sogenannten Kontextmodell [15] auf einen sehr vielversprechenden semantischen Rahmen [14, 20]. In diesem Rahmen läßt sich beispielsweise das aus der Theorie der Fuzzy-Mengen bekannte, in seinen Anwendungen jedoch eher für die Possibilitätstheorie typische Extensionsprinzip als einzig sinnvolle, mit der obigen Semantik vereinbare Erweiterung mengenwertiger Operationen zu entsprechenden Operationen auf Possibilitätsverteilungen nachweisen [14].

Mengenwertige Abbildungen auf Wahrscheinlichkeitsräumen zu verwenden, um unsichere und impräzise Daten handhaben zu können, ist eine seit langem bekannte Vorgehensweise [8, 51, 28]. Aus diesem Grund haben sich historisch auch noch andere Sichtweisen von Possibilitätsverteilungen etablieren können, z.B. Possibilitätsverteilungen als Konturfunktionen konsonanter Belief-Funktionen [49] und als Falling Shadows im Sinne der mengenwertigen Statistik [54]. Darüber hinaus gibt es auch noch die — für diesen Aufsatz allerdings weniger wichtigen — Arbeiten, in denen die Possibilitätstheorie als qualitative Unsicherheitstheorie im Rahmen der possibilistischen Logik gesehen wird [11]. Trotz der auf den ersten Blick sehr unterschiedlichen Darstellungsweisen sind Beziehungen zwischen den logischen und numerischen Kalkülen herstellbar. Sie beruhen im wesentlichen darauf, daß Possibilitätsgrade auch bei den numerischen, auf Beschreibungen mit Zufallsmengen

basierenden Ansätzen wegen der beim Übergang zu Possibilitätsverteilungen auftretenden Informationskompression im Gegensatz zu Wahrscheinlichkeiten keine Unsicherheitsarithmetik mehr zulassen. Stattdessen muß man sich auf Operationen beschränken, die mit der Ordnungsstruktur der Menge zulässiger Possibilitätsgrade vereinbar sind. Diese Erkenntnis brachte in der jüngsten Forschung die Entwicklung einer von der Arithmetik der Entscheidungstheorie von [48] losgelöste possibilistische Entscheidungstheorie mit eigener Axiomatik hervor [12], die in letzter Konsequenz die Possibilitätstheorie als eigenständige normative Theorie zur Modellierung von Unsicherheit und Impräzision etablieren wird.

17.5 Possibilistische graphische Modelle

Sollen Informationen in einem wissensbasierten System dargestellt und verarbeitet werden, so ist vor allem die Entwicklung eines geeigneten formalen und semantischen Rahmens zur Handhabung mit Unsicherheit behafteter und ggf. impräziser Daten wichtig [36]. Dieser Rahmen wird sich natürlich stark an der von dem wissensbasierten System zu lösenden Aufgabe orientieren müssen. Eine in Anwendungen sehr häufig vorkommende Aufgabe besteht nun in folgendem: Jeder Zustand eines zu modellierenden Weltausschnitts sei durch ein Tupel von Werten endlich vieler Merkmale (Attribute, Variablen) beschreibbar. Gegeben *generisches* (d.h. bereichsspezifisches) *Wissens* über die Abhängigkeiten zwischen den Merkmalen und *Evidenzwissen* über die in einer konkreten Situation vorliegenden bzw. (bei unscharfer Beobachtung) möglichen Merkmalswerte soll der wahre, aber unbekannt Zustand des Weltausschnitts so genau wie möglich beschrieben werden. In der medizinischen Diagnose ist z.B. der zu modellierende Weltausschnitt der Gesundheitszustand eines Patienten. Er wird durch mehrere zueinander in Beziehung stehende Merkmale wie Krankheiten, Symptome und bestimmte Meßgrößen (Blutdruck, Puls, Körpertemperatur etc.) beschrieben. Das generische Wissen besteht in medizinischen Regeln (z.B. über den Zusammenhang von Krankheiten und Symptomen), Erfahrungswerten von Ärzten und Datenbanken von Fallbeispielen (z.B. Krankengeschichten). Das Evidenzwissen besteht aus einer (unvollständigen) Menge von Merkmalswerten des zu behandelnden Patienten, die durch eine Befragung und Untersuchung dieses Patienten bekannt werden. Das Ziel der Diagnose ist, den Krankheitszustand des Patienten aus dem generischen und dem Evidenzwissen so genau wie möglich zu bestimmen.

Sind die Abhängigkeiten zwischen den Merkmalen zwar mit Unsicherheit behaftet, aber doch so beschaffen, daß sie eine Zerlegung auf der Basis eines geeigneten Konzeptes *bedingter Unabhängigkeit* der Merkmale zulassen, so kann man zum Aufbau der Abhängigkeitsstruktur der Merkmale (Erwerb des generischen Wissens) und zum effizienten Schlußfolgern innerhalb der gelernten Struktur (Ausnutzen des Evidenzwissens, um den tatsächlichen Zustand möglichst genau zu bestimmen) sehr gut Methoden der *graphischen Modellierung* anwenden [6]. Graphische Modelle werden in so unterschiedlichen Bereichen wie Diagnostik, Expertensysteme, Planungs-

systeme, Datenanalyse und Regelung verwendet.

Ein graphisches Modell zur Darstellung unsicheren Wissens besteht aus zwei Teilen: einem *qualitativen* und einem *quantitativen*. Der qualitative (strukturelle) Teil ist ein Graph (daher der Name *graphische Modelle*), zum Beispiel ein gerichteter azyklischer Graph (directed acyclic graph, DAG) oder ein ungerichteter Graph (undirected graph, UG). Dieser Graph beschreibt auf eindeutige Weise die bedingten Unabhängigkeiten, die zwischen den durch Knoten dargestellten Merkmalen bestehen. Der quantitative Teil eines graphischen Modells ist eine Familie von Verteilungsfunktionen auf Unterräumen des vieldimensionalen Wertebereiches der betrachteten Merkmale. Für gerichtete azyklische Graphen werden bedingte Verteilungsfunktionen benutzt, die die Unsicherheit über die Werte eines Merkmals in Abhängigkeit von den möglichen Belegungen aller im Graphen vorhandenen direkten Vorgänger dieses Merkmals beschreiben. In der medizinischen Diagnostik könnte es sich z.B. um eine bedingte Wahrscheinlichkeitsverteilung über das Auftreten eines bestimmten Symptoms handeln, gegeben das Vorliegen einer Krankheit und/oder eines bestimmten physiologischen Zustandes. Für ungerichtete Graphen werden (unbedingte) Marginalverteilungen auf einer von dem Graphen induzierten Hypergraphstruktur verwendet, die jeweils auf dem gemeinsamen Wertebereich der in einer Hyperkante enthaltenen Merkmale definiert werden. Hier würde z.B. eine Wahrscheinlichkeitsverteilung über das gemeinsame Auftreten bzw. Nichtauftreten einer Krankheit und bestimmter Symptome angegeben. Schlußfolgerungen werden in graphischen Modellen gezogen, indem man die Belegungen einzelner anwendungsabhängig zugänglicher Merkmale (das Evidenzwissen) nutzt, um die Verteilungen einzuschränken und so die revidierten Marginalverteilungen aller im Graphen vorkommenden Merkmale zu berechnen. Aus den Eigenschaften eines Patienten (Geschlecht, Alter etc.) und seinen Symptomen werden so z.B. Wahrscheinlichkeitsverteilungen über möglichen Krankheiten bestimmt.

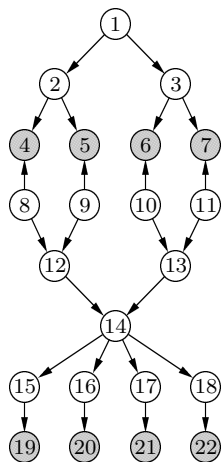
Am weitesten entwickelt für unsicheres Schließen in wissensbasierten Systemen sind probabilistische Netze (Bayes-Netze und Markov-Netze), für die Entwicklungsumgebungen wie z.B. HUGIN erhältlich sind [39, 27]. Bayes-Netze und Markov-Netze beruhen auf probabilistischen Modellen (d.h. bedingten oder unbedingten Wahrscheinlichkeitsverteilungen — wie in den obigen Beispielen angedeutet) zur Darstellung des generischen Wissens, das folglich zwar mit Unsicherheit behaftet sein darf, aber präzise sein muß (siehe oben). In der industriellen Praxis wird jedoch häufig darauf hingewiesen, daß die gemeinsame Modellierung von Unsicherheit und Impräzision in vielen Anwendungen die angemessene Vorgehensweise sei. Dies führt bei probabilistischen Ansätzen dazu, daß Familien von Wahrscheinlichkeitsverteilungen betrachtet werden müssen, was im allgemeinen erhebliche Komplexitätsprobleme für das Schlußfolgern in graphischen Modellen mit sich bringt. Wegen dieser Komplexitätsprobleme bei impräzisen (mengenwertigen) Daten sollten auch solche Modellierungen berücksichtigt werden, die eine geeignete vereinfachende Informationskompression verwenden und so approximatives Schließen ermöglichen. Hier bietet sich die Possibilitätstheorie an. Ähnlich wie in der Regelungstechnik ein Fuzzy-Regler als Werkzeug zur (informationskomprimierten) Interpolation zwischen Punkten in vagen Umgebungen aufgefaßt werden kann, kann ein possibilistisches

graphisches Modell als Werkzeug zum (informationskomprimierten) approximativen Schließen unter Unsicherheit und Impräzision in wissensbasierten Systemen gesehen werden.

Bevor ein graphisches Modell zum Schlußfolgern über konkrete Situationen eingesetzt werden kann, muß es natürlich erstellt werden, d.h. das generische Wissen über den zu modellierenden Weltausschnitt muß durch einen passenden Unabhängigkeitsgraphen und die zugehörige Familie von Wahrscheinlichkeitsverteilungen dargestellt werden. Diese notwendige Vorarbeit ist nicht zu unterschätzen. Teils kann sie von Bereichsexperten übernommen werden, doch ist sie oft mühselig und langwierig. Günstiger ist es, wenn man auf automatische Verfahren zurückgreifen kann, die eine Datenbank von Fallbeispielen (wie sie im Zuge der durch die moderne Hard- und Softwareentwicklung immer besseren Datenverwaltung oft zur Verfügung stehen) benutzen, um ein graphisches Modell zu konstruieren. Sind die Fallbeispiele präzise, wird man aus ihnen eine Wahrscheinlichkeitsverteilung schätzen und zwangsläufig zu probabilistischen graphischen Modellen gelangen. Liegt jedoch Impräzision vor — die sich dadurch äußert, daß eine höhere Anzahl von Fallbeispielen nicht durch einzelne Datentupel, sondern nur durch Mengen möglicher Datentupel beschrieben werden können — empfiehlt sich der Übergang zu possibilistischen graphischen Modellen. Die Gesamtheit der Fallbeispiele kann dann, wie im vorhergehenden Abschnitt beschrieben, als (generalisierte) Zufallsmenge gesehen und informationskomprimiert als Possibilitätsverteilung dargestellt werden.

Um possibilistische graphische Modelle auf die beschriebene Weise einsetzen zu können, müssen theoretische Grundlagen und Algorithmen erarbeitet werden, wie sie in den letzten fünfzehn Jahren für die probabilistischen graphischen Modelle geschaffen wurden [39, 6]. Die erste durchgängige Entwicklung einer Theorie possibilistischer graphischer Modelle ist in [14] zu finden. Sie umfaßt die Beschreibung einer auf dem Kontextmodell (vgl. den vorhergehenden Abschnitt) beruhenden Semantik von Possibilitätsverteilungen, die zu ihrer konsistenten Definition einer (bedingten) possibilistischen Unabhängigkeit mit passendem Faktorisierungskriterium, hierauf aufbauend die adäquate Definition und Interpretation bedingter possibilistischer Unabhängigkeitsgraphen sowie die Formulierung eines Satzes zur Zerlegung von Possibilitätsverteilungen, der zu dem aus der Wahrscheinlichkeitstheorie bekannten Hammersley-Clifford-Theorem analog ist.

Diese Theorie ermöglicht es nicht nur, effiziente Propagationsalgorithmen für das Schließen in possibilistischen graphischen Modellen mit Hyperbaumstruktur zu entwickeln, sondern die gelieferten Inferenzergebnisse auch zur Entscheidungsfindung in Anwendungsproblemen auszunutzen. Neben diesen für die Entwicklung von wissensbasierten Systemen wichtigen Ergebnissen ist eine aus der Sicht der Datenanalyse (Data Mining) interessante Konsequenz die Möglichkeit, possibilistische graphische Modelle aus Stichproben impräziser, voneinander unabhängiger Fallbeispiele zu erlernen. Die Lernaufgabe besteht darin, aus einer gewählten Klasse graphischer Modelle diejenige Zerlegung der aus den Fallbeispielen schätzbaren gemeinsamen Possibilitätsverteilung der Variablen des betrachteten Weltausschnitts zu berechnen, die die gemeinsame Verteilung am besten annähert. Um die Güte eines konstruierten Modells bewerten zu können, benötigt man ein Maß der Nichtspezifi-



- 1 – parental error
- 2 – dam correct?
- 3 – sire correct?
- 4 – stated dam ph.gr. 1
- 5 – stated dam ph.gr. 2
- 6 – stated sire ph.gr. 1
- 7 – stated sire ph.gr. 2
- 8 – true dam ph.gr. 1
- 9 – true dam ph.gr. 2
- 10 – true sire ph.gr. 1
- 11 – true sire ph.gr. 2
- 12 – offspring ph.gr. 1
- 13 – offspring ph.gr. 2
- 14 – offspring genotype
- 15 – factor 40
- 16 – factor 41
- 17 – factor 42
- 18 – factor 43
- 19 – lysis 40
- 20 – lysis 41
- 21 – lysis 42
- 22 – lysis 43

Die grauen Knoten entsprechen beobachtbaren Merkmalen. Knoten 1 kann entfernt werden, um die Konstruktion des Cliquesbaumes zu vereinfachen.

Abbildung 17.1 Von Rinderexperten entworfenes Bayessches Netz zur Blutgruppenbestimmung bei dänischen Jersey-Rindern.

| | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|----|----|----|----|----|----|----|----|----|----|------|---|---|---|---|---|---|---|---|
| n | y | y | f1 | v2 | f1 | v2 | f1 | v2 | f1 | v2 | v2 | v2 | v2v2 | n | y | n | y | 0 | 6 | 0 | 6 |
| n | y | y | f1 | v2 | ** | ** | f1 | v2 | ** | ** | ** | ** | f1v2 | y | y | n | y | 7 | 6 | 0 | 7 |
| n | y | y | f1 | v2 | f1 | f1 | f1 | v2 | f1 | f1 | f1 | f1 | f1f1 | y | y | n | n | 7 | 7 | 0 | 0 |
| n | y | y | f1 | v2 | f1 | f1 | f1 | v2 | f1 | f1 | f1 | f1 | f1f1 | y | y | n | n | 7 | 7 | 0 | 0 |
| n | y | y | f1 | v2 | f1 | v1 | f1 | v2 | f1 | v1 | v2 | f1 | f1v2 | y | y | n | y | 7 | 7 | 0 | 7 |
| n | y | y | f1 | f1 | ** | ** | f1 | f1 | ** | ** | f1 | f1 | f1f1 | y | y | n | n | 6 | 6 | 0 | 0 |
| n | y | y | f1 | v1 | ** | ** | f1 | v1 | ** | ** | v1 | v2 | v1v2 | n | y | y | y | 0 | 5 | 4 | 5 |
| n | y | y | f1 | v2 | f1 | v1 | f1 | v2 | f1 | v1 | f1 | v1 | f1v1 | y | y | y | y | 7 | 7 | 6 | 7 |

Tabelle 17.1 Ein Auszug aus der Datenbank zur Blutgruppenbestimmung bei dänischen Jersey Rindern.

zität der durch eine Possibilitätsverteilung dargestellten Information. [18, 19] zeigen, daß ein auf der Hartley-Information beruhendes possibilistisches Informationsmaße zur Entwicklung eines effizienten Lernverfahrens geeignet ist. Außerdem kann man die Ideen verschiedener probabilistischer Maße auf den possibilistischen Fall übertragen und so weitere Maße erhalten [4, 5]. Die sich ergebenden Lernverfahren sind sowohl als Erweiterung des an der TU Braunschweig entwickelten Programms POSSINFER (POSSibilistic INFERENCE) [17] und des an der TU Braunschweig und der Universität Magdeburg in Zusammenarbeit mit dem Forschungszentrum Ulm der Daimler-Benz AG entwickelten Programmprototyps INES (Induktion von NetzwerkStrukturen aus Daten), mit dem sich auch probabilistische Netze aus Daten lernen lassen, implementiert [4, 5].

17.6 Anwendung possibilistischer Netze

Auch wenn eine gute Theorie das praktischste Ding ist, das man haben kann, so muß sich doch jede Theorie in einem Test an der Wirklichkeit bewähren. Als Beispiel für eine Anwendung der Theorie des Lernens possibilistischer graphischer Modelle aus Daten wählen wir hier das Problem der Blutgruppenbestimmung bei dänischen Jersey Rindern [47]. Zur Lösung dieses Problems wurde von Rinderexperten ein Bayessches Netz (probabilistisches Schlußfolgerungsnetz) entworfen (siehe Abbildung 17.1). Außerdem gibt es eine Datenbank von 500 Beispielfällen (ein Auszug aus dieser Datenbank ist in Tabelle 17.1 gezeigt). Aus dieser Datenbank gilt es, ein entsprechendes Schlußfolgerungsnetz zu lernen, wobei das von den Rinderexperten entworfene Bayessche Netz als Qualitätsmaßstab dient.

Das Problem, das sich bei der Verwendung dieser Datenbank stellt, ist, daß sie eine relativ hohe Zahl unbekannter Werte enthält — nur knapp über die Hälfte aller Tupel sind vollständig (man sieht dies bereits an dem in Tabelle 17.1 gezeigten Ausschnitt: die Sterne markieren unbekannte Werte). Unbekannte Werte stellen nun aber impräzise Information dar, denn der wahre Wert kann ein beliebiger aus dem zugehörigen Wertebereich sein. Ein unvollständiges Tupel ist folglich eine impräzise Beschreibung der Verhältnisse in einem konkreten Fall, da es als Menge möglicher vollständiger Tupel gedeutet werden kann. Diese Impräzision muß zusätzlich zu der Unsicherheitsinformation, die man aus der relativen Häufigkeit verschiedener Merkmalswertkombinationen in der Datenbank erhält, berücksichtigt werden. Noch schwieriger wird das Problem natürlich, wenn über ein Merkmal gewisse Information vorliegt, also die Menge möglicher Werte eingeschränkt werden kann, jedoch nicht auf *einen* Wert. Für rein probabilistische Ansätze bringt das nicht geringe Schwierigkeiten mit sich, wenn man auch auf keine prinzipielle Unmöglichkeit stößt. Gewöhnlich besteht die einzig praktikable Lösung darin, einfach alle Tupel zu entfernen, in denen impräzise (mengenwertige) Information vorliegt. Damit wird im vorliegenden Beispiel allerdings viel Information verschenkt, obwohl es immer noch möglich ist, durch ein Lernverfahren für Bayessche Netze die Netzstruktur zu bestimmen. Dies liegt jedoch i.w. daran, daß die Abhängigkeiten relativ stark sind und daher die verbleibenden vollständigen Tupel für die Netzbestimmung ausreichen.

Verwendet man dagegen Lernverfahren für possibilistische graphische Modelle, so tritt das Problem, wie mit den unbekanntenen Werten umzugehen sei, gar nicht auf, da die Possibilitätstheorie ja besonders zur Behandlung mengenwertiger Information geeignet ist. Folglich brauchen keine Tupel entfernt oder in besonderer Weise behandelt zu werden. Ein Test des gelernten possibilistischen Netzwerks zeigt, daß seine Qualität mit der eines gelernten Bayesschen Netzes vergleichbar ist. Wir können also schließen, daß possibilistische Methoden in der Datenanalyse eine beachtenswerte Alternative zu den etablierten probabilistischen Methoden darstellt.

Literaturverzeichnis

- [1] Z. Artstein und R.A. Vitale. A Strong Law of Large Numbers for Random Compact Sets. *Ann. Probability*, 3:879–882, 1975
- [2] H. Bandemer und W. Näther. *Fuzzy Data Analysis, Series B: Mathematical and Statistical Methods* Kluwer, Dordrecht, Niederlande 1992
- [3] H. Bandemer. *Ratschläge zum mathematischen Umgang mit Ungewißheit — Reasonable Computing*. Teubner, Stuttgart 1997
- [4] C. Borgelt and R. Kruse. Evaluation Measures for Learning Probabilistic and Possibilistic Networks. *Proc. 6th IEEE Int. Conf. on Fuzzy Systems (FUZZ-IEEE'97)*, Vol. 2, 1034–1038, Barcelona, Spanien 1997
- [5] C. Borgelt and R. Kruse. Some Experimental Results on Learning Probabilistic and Possibilistic Networks with Different Evaluation Measures. *Proc. 1st Int. Joint Conference on Qualitative and Quantitative Practical Reasoning (ECSQARU/FAPR'97)*, 71–85, Springer, Berlin 1997
- [6] E. Castillo, J.M. Gutierrez und A.S. Hadi. *Expert Systems and Probabilistic Network Models*. Serie Monographs in Computer Science, Springer, New York, 1997
- [7] E. Czogala und K. Hirota. *Probabilistic Sets: Fuzzy and Stochastic Approach to Decision Control and Recognition Processes*. Verlag TÜV Rheinland, Köln 1986
- [8] A.P. Dempster. Upper and Lower Probabilities Induced by a Multivalued Mapping. *Ann. Math. Stat.*, 38:325–339, 1967
- [9] D. Dubois und H. Prade. *Possibility Theory*. Plenum Press, New York, 1988
- [10] D. Dubois und H. Prade. Fuzzy Sets in Approximate Reasoning, Part 1: Inference with Possibility Distributions. *Fuzzy Sets and Systems* 40:143–202, 1991
- [11] D. Dubois und H. Prade. Belief Change and Possibility Theory. In: P. Gärdenfors (Hrsg.): *Belief Revision*, 141–182, Cambridge University Press, Cambridge, 1992
- [12] D. Dubois, H. Prade und R. Sabbadin. Qualitative Decision Theory with Sugeno Integrals. *Proc. 14th Conf. on Uncertainty in Artificial Intelligence*. Madison, Wisconsin, 1998
- [13] J. Gebhardt und R. Kruse. Some New Aspects of Testing Hypotheses in Fuzzy Statistics. *Proc. NAFIPS'90*, 185–187. Toronto, Kanada, 1990
- [14] J. Gebhardt. *Learning From Data: Possibilistic Graphical Models*. Habilitationsschrift, TU Braunschweig, 1997.
- [15] J. Gebhardt und R. Kruse. The Context Model — An Integrating View of Vagueness and Uncertainty. *Int. J. of Approximate Reasoning*, 9:283–314, 1993

-
- [16] J. Gebhardt und R. Kruse. A New Approach to Semantic Aspects of Possibilistic Reasoning. In: M. Clarke, R. Kruse and S. Moral (Hrsg.): *Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, 151–160. Lecture Notes in Computer Science 747, Springer, Berlin 1993
- [17] J. Gebhardt und R. Kruse. POSSINFER — A Software Tool for Possibilistic Inference. In: D. Dubois, H. Prade und R. Yager (Hrsg.): *Fuzzy Set Methods in Information Engineering: A Guided Tour of Applications*, J. Wiley & Sons, New York, 1996
- [18] J. Gebhardt und R. Kruse. Tightest Hypertree Decompositions of Multivariate Possibility Distributions. *Proc. Int. Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems*, 923–928, Granada, Spanien 1996
- [19] J. Gebhardt und R. Kruse. Automated Construction of Possibilistic Networks from Data. *Journal of Applied Mathematics and Computer Science*, 6:101-136, 1996
- [20] J. Gebhardt und R. Kruse. Parallel Combination of Information Sources. In: D. Gabbay und P. Smets (Hrsg.): *Handbook of Defeasible Reasoning and Uncertainty Management Systems*, Vol. 3, 393–439, Kluwer, Dordrecht, Niederlande 1998
- [21] J. Gebhardt, M.A. Gil und R. Kruse. Fuzzy Set Theoretic Methods in Statistics. In: R. Slowinski (Hrsg.): *Fuzzy Sets in Decision Analysis, Operations Research and Statistics*, 311–347. The Handbook of Fuzzy Sets Series, Kluwer, Boston 1998
- [22] M.A. Gil, N. Corral und M.R. Casals. The Likelihood Ratio Test for Goodness of Fit with Fuzzy Experimental Observations. *IEEE Transactions on Systems, Man and Cybernetics*, 19:771–779, 1989
- [23] J. Gontsias, R.P.S. Mahler und H.T. Nguyen. *Random Sets — Theory and Applications*. Springer, New York 1997
- [24] K. Hestir, H.T. Nguyen und G.S. Rogers. A Random Set Formalism for Evidential Reasoning. In: I.R. Goodman, M.M. Gupta, H.T. Nguyen und G.S. Rogers (Hrsg.): *Conditional Logic in Expert Systems*, 309-344, North-Holland, Amsterdam, Niederlande 1991
- [25] K. Hirota. Concepts of Probabilistic Sets. *Fuzzy Sets and Systems*, 5:31–46.
- [26] F. Höppner, F. Klawonn und R. Kruse. *Fuzzy-Clusteranalyse — Verfahren für die Bilderkennung, Klassifikation und Datenanalyse*. Vieweg, Wiesbaden, 1997
- [27] F.V. Jensen. *An Introduction to Bayesian Networks*. Springer, New York 1996
- [28] J. Kampé de Fériet. Interpretation of Membership Functions of Fuzzy Sets in Terms of Plausibility and Belief. In: M.M. Gupta und E. Sanchez (Hrsg.): *Fuzzy Information and Decision Processes*, 13-98, North-Holland, Amsterdam, Niederlande 1982
- [29] D.G. Kendall. Foundations of a Theory of Random Sets. In: E.F. Harding und D.G. Kendall (Hrsg.): *Stochastic Geometry*, 322–376. J. Wiley & Sons, Chichester 1974

- [30] E.P. Klement, M.L. Puri und D.A. Ralescu. Limit Theorems for Fuzzy Random Variables. *Proc. of the Royal Society of London — Series A*, 19:171–182, 1986
- [31] R. Kruse. The Strong Law of Large Numbers for Fuzzy Random Variables. *Information Sciences* 12:53–57, 1982
- [32] R. Kruse. *Schätzfunktionen für Parameter von unscharfen Zufallsvariablen*. Habilitationsschrift, TU Braunschweig 1983
- [33] R. Kruse. On a Software Tool for Statistics with Linguistic Data. *Fuzzy Sets and Systems*, 24:377–383, 1987
- [34] R. Kruse and K.D. Meyer. *Statistics with Vague Data*. Reidel, Dordrecht, Niederlande, 1987
- [35] R. Kruse und J. Gebhardt. On a Dialog System for Modelling and Statistical Analysis of Linguistic Data. *Proc. IFSA Congress*, 157–160. Seattle 1989
- [36] R. Kruse, E. Schwecke und J. Heinsohn. *Uncertainty and Vagueness in Knowledge-based Systems*. Serie Artificial Intelligence, Springer, Berlin 1991
- [37] R. Kruse, J. Gebhardt und F. Klawonn. *Foundations of Fuzzy Systems*. J. Wiley & Sons, New York, 1994.
- [38] H. Kwakernaak. Fuzzy Random Variables Part 1: Definitions and Theorems. *Information Sciences*, 15:1–15, 1978
- [39] S.L. Lauritzen. *Graphical Models*. Oxford University Press, Oxford, England 1997
- [40] G. Matheron. *Random Sets and Integral Geometry*. J. Wiley & Sons, New York 1975
- [41] K.D. Meyer. *Grenzwertsätze zum Schätzen von Parametern unscharfer Zufallsvariablen*. Dissertation, TU Braunschweig 1986
- [42] M. Miyakoshi und M. Shimbo. A Strong Law of Large Numbers for Fuzzy Random Variables. *Fuzzy Sets and Systems*, 12:133–142, 1984
- [43] K.G. Manton, M.A. Woodburg und H.D. Tolley. *Statistical Applications Using Fuzzy Sets*. J. Wiley & Sons, New York 1994
- [44] H.T. Nguyen. On Random Sets and Belief Functions. *Journal of Mathematical Analysis and Applications*, 65:431-542, 1978
- [45] F. Proske. *Grenzwertsätze für Fuzzy-Zufallsvariablen unter dem Gesichtspunkt der Wahrscheinlichkeitstheorie auf inseparablen Halbgruppen*. Dissertation, Universität Ulm, 1997
- [46] D.A. Ralescu. Fuzzy Logic and Statistical Estimation. *Proc. 2nd World Conf. on Mathematics at the Services of Man*, 1982
- [47] L.K. Rasmussen. *Blood Group Determination of Danish Jersey Cattle in the F-blood Group System*. Dina Research Report no. 8, 1992
- [48] L.J. Savage. *The Foundations of Statistics*. J. Wiley & Sons, New York, 1954
- [49] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton 1976
- [50] D. Stoyan, W.S. Kendall und J. Mecke. *Stochastic Geometry and Its Applications*. J. Wiley & Sons, Chichester, England 1987

-
- [51] V. Strassen. Meßfehler und Information. *Zeitschrift Wahrscheinlichkeitstheorie und verwandte Gebiete*, 2:273-305, 1964
- [52] H. Tanaka, T. Okuda und K. Asai. Fuzzy Information and Decision in Statistical Model. In: M.M. Gupta, R.K. Ragade und R.R. Yager (Hrsg.): *Advances in Fuzzy Sets Theory and Applications*, 303-320. North Holland, Amsterdam, Niederlande 1979
- [53] R. Viertl. Is It Necessary to Develop a Fuzzy Bayesian Inference? *Probability and Bayesian Statistics*, 471-475. Plenum Press, New York 1987
- [54] P.Z. Wang. From the Fuzzy Statistics to the Falling Random Subsets. In: P.P. Wang (Hrsg.): *Advances on Fuzzy Sets, Possibility and Applications*, 81-96, Plenum Press, New York 1983
- [55] O. Wolkenhauer. *Possibility Theory with Applications to Data Analysis*. UM-IST Control Systems Center Series 5, Research Studies Press/John Wiley & Sons, Somerset, England, 1998
- [56] L.A. Zadeh. Fuzzy Sets as a Basis for a Theory of Possibility. *Fuzzy Sets and Systems*, 1:3-28, 1978