Uncertainty Management

Alfs T. Berztiss[1,2]

[1] University of Pittsburgh, Department of Computer Science, Pittsburgh PA 1526 USA
[2] SYSLAB, University of Stockholm, Sweden
e-mail: alpha@cs.pitt.edu

Abstract. The purpose of this fairly nontechnical introduction to uncertainty management is to identify various forms of uncertainty, and to survey methods for managing some of these uncertainties. Our emphasis is on topics that may not be familiar to software engineers or, to a lesser extent, to knowledge engineers. These topics include Bayesian estimation, fuzziness, time Petri nets, rough sets, belief and evidence, and possibility theory. Uncertainty management has been studied in the contexts of information systems and of artificial intelligence. We attempt to present a balanced view of the contributions from both areas.
Keywords: belief, fuzziness, possibility, probability, rough set, time Petri net, uncertainty.

1 Introduction

Most decisions in life have to be made under uncertainty. This applies also to decisions made by software systems. But uncertainty can be managed, that is, it can be allowed for in a systematic and disciplined way. The purpose of this chapter is to survey approaches to the management of uncertainty. Section 2 is a catalog of different types of uncertainty. Section 3 is a brief survey of basic discrete probability and stochastic inference, including Bayesian inference. Fuzzy sets are introduced in Sec. 4, and fuzzy logic in Sec. 5. In Sec. 6 we outline the operation of a fuzzy control system, including defuzzification.

Time Petri nets allow the introduction of uncertainty into the time aspects of real-time processes. An application of time Petri nets is discussed in Sec. 7. Similarity gives rise to a particular type of uncertainty — given a set of objects that are to be grouped into different classes, what criteria should be used to put a particular object into one of the classes? The theory of rough sets, which is looked at in Sec. 8, is one way of dealing with this problem. Section 9 is concerned with belief, and Sec. 10 with possibility theory. In the interests of readability we have avoided sprinkling the text with references, putting nearly all bibliographic pointers into Sec. 11.

2 Types of uncertainty

In [1], which is probably the best reference on various aspects of uncertainty, uncertainty issues are partitioned into those that deal with fuzziness and those that deal with ambiguity. Fuzziness is taken to deal with lack of distinctions. Ambiguity problems are further partitioned into discord problems and nonspecificity. Discord is a disagreement when a choice between different alternatives is possible; nonspecificity relates to situations in which some such possible alternatives are not taken into account. In [2] we set up a finer classification scheme, which we present here in a somewhat expanded form. For quite a few of the uncertainty types there is no definite procedure for dealing with them, but awareness of their existence will at least cause us to make allowances for them.

Inconsistency. This arises when conflicting solutions to a problem are being advanced, notably in dietary recommendations and in economics. For example, both lowering and raising taxes may be suggested as a cure for reducing budget deficits. Software requirements can state that costs are to be low, and reliability is to be high — satisfaction of one of these objectives normally means that the other cannot be satisfied.

Permanent exceptions. Emus and penguins are exceptions to "All birds fly". This problem is easily solved by a more refined classification of the entities under consideration.

Temporary exceptions. Suppose that all vice-presidents are to have offices on a particular floor of the head office buiding. If Ms. Smith has been made a vice-president, but has not yet made a move to this floor, we have a temporary exception.

Limited validity. Although Jack may not own a car, there are times when he "has" a car, borrowed from his parents. It is not clear to what extent one "owns" a heavily mortgaged house.

Multiple options. People often have several addresses, used for different purposes. There may be different options on how to fly from place A to place B — cost or convenience may decide which of them is selected.

Nondeterminism. If a software system interacts with its environment the behavior of the system is not fully predictable from knowledge of its initial state.

Obscurity. A general trend in data may be obscured by temporal variations, e.g., monthly income of a ski resort.

Fuzziness. We designate Mr. Smith as tall, but tallness is not a well-defined property. We interpret a property as fuzzy if a precise measurement of this property can be obtained in principle. Examples of fuzzy terms: cold, old, loud.

Vagueness. In [1] vagueness is considered a sub-category of fuzziness. We consider vagueness to be distinct from fuzziness. In contrast to fuzzy terms, we call those terms vague for which no measurement process can exist. In "People feel uncomfortable when it is hot" the term "hot" is fuzzy, but "uncomfortable" is vague — we have no dependable way of measuring discomfort. Part of the research on uncertainty should be aimed at reducing vagueness by developing new measurement processes.

Faults. An electrical utility meter breaks down. In some circumstances it would help to know when this happened.

Either/or uncertainty. If this meter registers the same value over a period of time, it may be broken, but not necessarily. The rumor "Company X will show a loss this quarter" may be true or false, but we do not know which.

Cause/effect uncertainty. An effect is observed that could be due to different causes, singly or jointly. This is quite common in medical diagnosis. Analysis methods based on fault trees and Bayesian networks have been developed to deal with such situations.

No knowledge as negation. In logic programming, if the fact X is not in its knowledge base, not(X) is assumed true. For example, if the knowledge base does not contain "Mr. Khachaturian is married," and this cannot be derived from, say, "The husband of Mrs. Khachaturian is Mr. Khachaturian," the answer to the query "Is Mr. Khachaturian married?" will be "no".

Null values. This problem has been extensively studied by the data base community. It has two aspects. One is representational: when a value is missing, how should this be indicated? A missing value may exist, but be unknown (John's age), not exist at all (temperature data for a given place and date because no reading was taken), or be inapplicable (the name of the spouse of an unmarried person). Different null-markers may be used to differentiate between these three types of null values, but at times there may be uncertainty as to which marker is appropriate. The other aspect relates to query answering. If the data base does not contain data needed to respond to a query, an approximate answer may be given. Here the uncertainty relates to similarity: how are we to determine what approximate answer is appropriate?

Interpretation uncertainty. At one time the author worked in an insurance office where three different interpretations of "age" were in effect, namely age last birthday, age nearest birthday, and age next birthday. In the United States a weather forecast may state that tomorrow there is a 60% chance of precipitation. What exactly does this mean?

Rounding. This is interpretation uncertainty as it applies to numerical data. Is USD 5,000,000 an exact amount or is it a rounded estimate? As the purchase price for a mansion it is probably exact; as damage estimate due to flooding it is likely to be a rounded figure.

Noisy data. An experimentally determined value is nearly always uncertain to some degree, e.g., the speed of light. Also, data may become polluted during transmission, requiring the use of error-detecting codes.

Computational round-off. As an illustration, if one keeps adding up floating-point representations of the number 1, at some point the sum no longer changes. Different results are obtained depending on whether a series is added from the small values first or the large values first.

Computational result uncertainty. This can take two forms. First, the program is correct, but the result it gives need not be the required solution. The aim of nonlinear optimization is to find the maximum of an objective function. A two-variable function $f(x, y)$ can be pictured as a three-dimensional hilly terrain. Hill climbing methods will find the top of a hill, i.e., a solution, but there is no guarantee that this is the highest hill. The problem becomes more and more difficult as the number of variables goes up. Under the second form we have no way of telling whether a program is correct. The whole point of many programs is to arrive at results that cannot be obtained in any other way. Hence there is really no way of testing such a program, and a result will be accepted as long as it "looks reasonable".

Partial knowledge. Six prisoners escape. Their names are known. We also know that they have split into two groups. However, we may not know who is in each group, or even the size of each group.

Trends. An upward trend has been observed in global temperatures. Is this due to chance or is there an underlying reason for it?

Context-dependence. Many terms cannot be fully understood unless the context is known in which they arise. For example, a warm day near the Arctic Circle is likely to be cooler than a cool day near the Equator. Since the concepts "warm" and "cold" are fuzzy, the example shows that care must be taken in the interpretation of fuzzy terms. Although we take the boiling point of water to be 100°C, at high elevations it is actually lower.

3 Probabilistic concepts

Probability theory was the earliest attempt to deal with uncertainty in a disciplined quantitative manner. Here we can do no more than give a very brief review of basic discrete probability. In particular, we introduce Bayesian estimation, which is being applied extensively in artificial intelligence to deal with uncertainty.

Consider an experiment that can have k possible outcomes or events a_1, a_2, ..., a_k. This experiment is performed n times, and the counts of observed outcomes are n_1, n_2, ..., n_k. Then $p(a_i) = n_i/n$ is the observed probability of outcome a_i. We have

$$\sum_{i=1}^{k} p(a_i) = \sum_{i=1}^{k} n_i/n = \frac{1}{n}\sum_{i=1}^{k} n_i = 1.$$

The probability of the disjunction of outcomes a_1, a_2, ..., a_t is the sum of their respective probabilities; the probability of the conjunction of the outcomes is the product of their probabilities:

$$\sum_{i=1}^{t} p(a_i); \quad \prod_{i=1}^{t} p(a_i).$$

Thus the probability of the throw of a die resulting in a 3 or a 6 is $1/6 + 1/6 = 1/3$, and the probability of the first throw of a die resulting in a 3 and the second throw in a 6 is $1/6 \times 1/6 = 1/36$.

Conditional probability determines the probability of an event, say A, given that another event, say B, has occurred:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

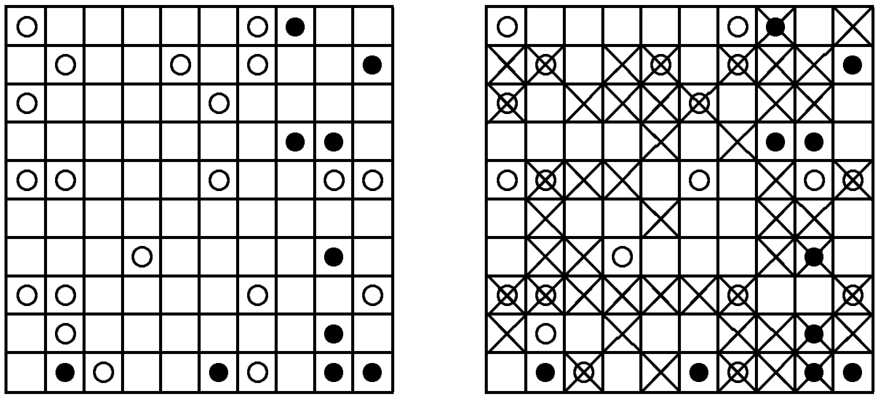

Figure 1: An experiment in estimation

Example: In unit testing of 100 code units 20 were found to have syntax errors, 10 to have semantic errors, and 6 to have both. If we select a unit that we know to have a syntax error, then the probability that it also has a semantic error is $0.06/0.20 = 0.30$.

The joint probability distribution of random variables X and Y is

$$f(x, y) = P(X = x, Y = y),$$

i.e., $f(x, y)$ is the probability that both x and y occur at the same time. From this we obtain marginal distributions:

$$g(x) = \sum_{y} f(x, y), \quad h(y) = \sum_{x} f(x, y).$$

Marginal distributions arise in Bayesian estimation.

Consider a 10×10 grid, in which 20 white and 10 black balls are randomly distributed as shown in Fig.1. Now select a random sample of 50 of the 100 squares, marked with crosses in the the diagram on the right side of Fig.1. These squares contain 13 white and 4 black balls, although, based on the distribution, we expect to get 10 and 5 balls, respectively. Instead of probabilities 0.20 and 0.10, we get experimental estimates 0.26 and 0.08. This suggests that we also need measures for the confidence we can put on the estimates, e.g., the ranges in which the true numbers of the balls are expected to lie with a probability of 0.95, say. All that we can say with absolute certainty is that the true number of balls lies between 17 and 67, and the upper limit is known only because the number of squares is 100 (there is a non-zero probability that every one of the 50 unexamined squares contains a ball).

From experimental observations we can obtain an estimator of a parameter, and a measure of the degree of confidence in this parameter. For example,

$$P(-0.26 < z < 0.26) = 0.95$$

means that 95% of the time in repeated experiments z will have a value within $(-0.26, 0.26)$. The main applications in computing relate to the measurement of software reliability and communication delays.

In most cases probabilities are established by experiments (observations). For example, to determine the frequency of left-handed people in the population, count them in a sample. This is the accepted scientific approach since Galileo. The probabilities obtained in this manner are called *objective* or *empirical.*

Subjectivists or Bayesians want to support estimates also with prior knowledge, and beliefs and values. For example, if you know that at the time of counting there is a convention of left-handed people in your city, this knowledge can be used to adjust the observed probability. This is fine, but sometimes a subjectivist may introduce beliefs in the analysis when this is not justified — in complex situations experiments have been known to give counterintuitive results. Bayesian estimation can be very useful, but care has to be taken when and how it is used.

Let us find an estimate of a parameter θ for the population defined by the function $f(x, \theta)$. To make this more concrete, consider Bernoulli trials, i.e., trials that can result in success with probability p and failure with probability $(1-p)$. The binomial distribution applies, and it gives the probability of x successes in n trials as

$$b(x; n, p) = \binom{n}{x} p^x (1-p)^{n-x}.$$

In terms of this distribution, if we observe x successes in n trials, the θ is

$$\widehat{p} = x/n.$$

Now suppose that we have additional information regarding θ, e.g., that we have a probability distribution $f(\theta)$ for it. This information allows us to improve the objective result.

The $f(\theta)$ is called a *prior distribution*, and the probabilities associated with it are called *subjective* in that experience, prior knowledge, and beliefs define the prior distribution. Bayesian techniques combine $f(\theta)$ with the joint distribution $f(x_1, \ldots, x_n; \theta)$ into a *posterior distribution*

$$f(\theta | x_1, \ldots, x_n) = \frac{f(x_1, \ldots, x_n; \theta)}{g(x_1, \ldots, x_n)},$$

where $g(x_1, \ldots, x_n)$ is the marginal distribution $\sum_\theta f(x_1, \ldots, x_n; \theta)$.

Example. Consider the estimation of defect probability from an observation that 1 item is found defective in a sample of 2. The objective probability of defectiveness is $1/2 = 0.5$. But suppose that the manufacturer tells us that the probability of finding 1 defective in 10 items is 0.6, and that of finding 2 defectives in 10 is 0.4, i.e., that $f(p) = 0.6$ when p $= 0.1$, and $f(p) = 0.4$ when p $= 0.2$. This is a prior distribution, subjective in the sense that it depends on how ready we are to believe the manufacturer. Let x be the number of defectives in the sample. Then the probability distribution for the sample is

$$f(x|p) = b(x; n, p) = \binom{2}{x} p^x (1-p)^{2-x}.$$

From this, the probabilities of a random sample of 2 yielding 1 defective are

$$f(1|0.1) = b(1;2,0.1) = \binom{2}{1}(0.1)(0.9) = 0.18,$$

$$f(1|0.2) = b(1;2,0.2) = \binom{2}{1}(0.2)(0.8) = 0.32.$$

But $f(x,p) = f(x|p)f(p)$, so that $f(1,0.1) = 0.108$ and $f(1,0.2) = 0.128$, and hence $g(1) = 0.108 + 0.128 = 0.236$. The posterior distribution for the proportion of defectives p when $x = 1$ is $f(1,p)/g(1)$, so that $f(0.1|x = 1) = 0.458$ and $f(0.2|x = 1) = 0.542$. Thus the Bayesian estimate of p is $(0.1)(0.458) + (0.2)(0.542) = 0.154$, quite different from the objective estimate of 0.5.

4 Fuzzy sets

If we can tell whether or not an element of a universal set U belongs to a set A, this set can be defined by its *characteristic function*:

$$f_A : U \to \{0,1\},$$

such that $f_A(a) = 1$ if $a \in A$, $f_A(a) = 0$ if $a \notin A$. A set that has a characteristic function is *crisp*. This concept can be generalized by defining a *membership function*:

$$\mu_A : U \to [0,1],$$

where $\mu_A(a)$ expresses a degree of membership of a in A, or the strength of a belief that a belongs to A. Such a set A is *fuzzy*. Fuzziness finds application in contexts where a strict demarcation between attribute values is inappropriate — we cannot select a particular value and say that it is cool when the temperature is below this value, and warm when it is above it; the transition from cool to warm is gradual. This observation led Zadeh to develop the theory of fuzzy sets.

Characteristic functions can be given a probabilistic interpretation. Thus, if we have a set of integers $A = \{n | 1 \leq n \leq 10\}$, then for all $1 \leq i \leq 10$ we have $f_A(i) = 1$, but also, for all $1 \leq i \leq 10$, the probability $P(i \in A)$ is 1. This suggests that the $\mu_A(a)$ can also be interpreted as probabilities, giving rise to a view that probability theory is all that is needed. However, fuzzy sets have been found very useful in practice, so that theoretical arguments relating to their necessity are of academic interest alone. In what follows we shall use an alternative notation, writing $A(x)$ for μ_A, and $A(a)$ for $\mu_A(a)$.

Fig.2 shows plots of membership functions A_1, A_2, A_3, which stand, respectively, for "young", "middle aged", and "old". There are two sets of plots, illustrating one problem we face with fuzzy concepts — they are often given very subjective interpretations. The upper diagram relates to how a 15-year old would define the membership functions; the lower diagram shows the interpretation given by a 70-year old.

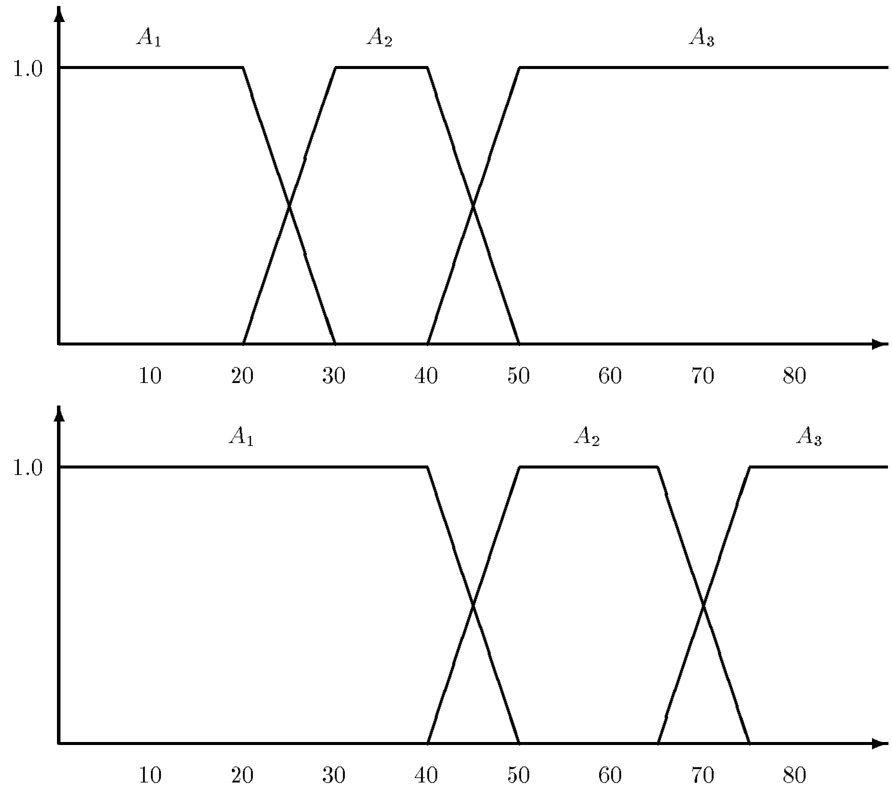

Figure 2: Membership functions relating to age

Some definitions. A fuzzy set A is *normal* if $\exists a : A(a) = 1$. The *height* of A is $\max_a A(a)$. Set A is a *subset* of set B, $(A \subset B)$, if $\forall a \in U : B(a) \geq A(a)$. *Support* of A: $Supp(A) = \{a|A(a) > 0\}$; *core* of A: $Core(A) = \{a|A(a) = 1\}$. To illustrate these terms, define $U = \{a, b, c, d, e\}$, and fuzzy sets based on U:

$$A = \{1/a,\ 0.3/b,\ 0.2/c,\ 0.8/d,\ 0/e\},$$

$$B = \{0.6/a,\ 0.9/b,\ 0.1/c,\ 0.3/d,\ 0.2/e\},$$

$$C = \{1/a,\ 0.4/b,\ 0.2/c,\ 0.9/d,\ 0.2/e\},$$

where the value x in x/y is the degree of membership of y in its fuzzy set. Fuzzy sets A and C are normal (with height 1); the height of B is 0.9; $Supp(B) = Supp(C) = U$; $Supp(A) = \{a, b, c, d\}$. $Core(A) = Core(C) = \{a\}$; $Core(B) = \emptyset$; $A \subset C$.

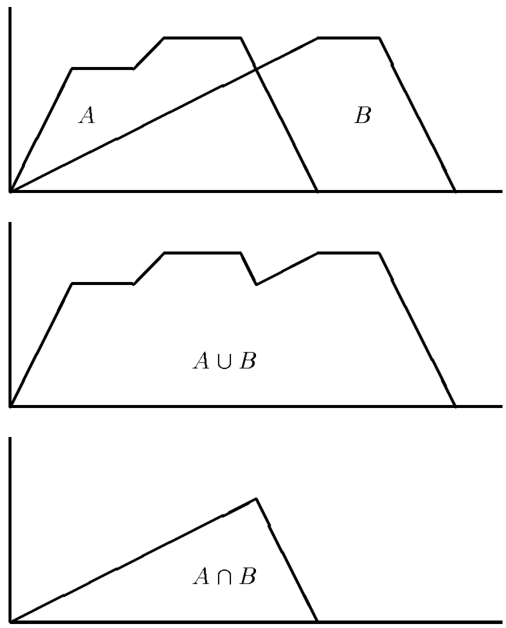

Figure 3: Standard fuzzy set union and intersection

Matters get very complicated when we turn to the fuzzy versions of set operations. There are several versions of intersections (called t-norms) and unions (called t-conorms). Much of the really extensive literature on fuzzy sets is concerned with selecting the appropriate t-norm and t-conorm for a particular application. The three *standard* fuzzy set operations are defined as follows:

$$\overline{A}(a) = 1 - A(a),$$

$$(A \cap B)(a) = \min[A(a), B(a)],$$

$$(A \cup B)(a) = \max[A(a), B(a)].$$

Fig.3 shows fuzzy sets A and B, represented by their membership functions, and their standard union and intersection.

Relations on fuzzy sets allow us to deal with similarity in a natural manner. An equivalence relation splits up a set into equivalence classes, e.g., "x is the same age as y" partitions a population according to age. In order for relation R on set S to be an equivalence relation, the relation has satisfy the following conditions for all $x, y, z \in S$:

$$\text{reflexivity}: \ R(x, x) = \text{true};$$

$$\text{symmetry}: \ R(x, y) = R(y, x);$$

$$\text{transitivity}: \ R(x, y) \wedge R(y, z) \rightarrow R(x, z).$$

A similarity relation groups together similar entities, so we expect it to resemble the equivalence relation. However, transitivity does not hold, as, for example, "x and y have similar salaries" shows — salary x may be similar to salary $x + \delta x$, salary $x + \delta x$ similar to $x + 2\delta x$, $x + 2\delta x$ similar to $x + 3\delta x$, and so forth, but salaries x and, say, $x + 5\delta x$ need no longer be regarded as similar.

So similarity differs from equivalence with respect to transitivity. Transitivity now has to be fuzzy, and by making it fuzzy we obtain the fuzzy equivalence relation, which is commonly called the *similarity relation*. It satisfies, for all $x, y, z \in S$:

$$R(x, x) = 1;$$

$$R(x, y) = R(y, x);$$

$$R(x, z) \geq \max_{y \in S} \min[R(x, y), R(y, z)].$$

The concept of equivalence classes can be preserved by the use of α-cuts. An α-cut of fuzzy set A is a crisp set ${}^{\alpha}A$ consisting of all those elements of A whose membership grades are not below α:

$${}^{\alpha}A = \{x | A(x) \geq \alpha\}.$$

Applying this to the fuzzy relation R, for any α in $(0, 1]$, the α-cut ${}^{\alpha}R$ creates a crisp equivalence relation representing similarity between its elements to degree α. Each such α-equivalence forms a partition of S, and elements x and y of S belong to the same block of this partition if $R(x, y) \geq \alpha$. If $\alpha_2 > \alpha_1$, then the partition induced by α_2 is a refinement of the partition induced by α_1, or is the same as this partition.

Example. Let the membership function for a fuzzy relation R on $\{a, b, \ldots, h\}$ be as shown in Fig.4. Then the partition corresponding to $\alpha = 0.4$ has the blocks $\{a, e, f, g\}$, $\{b, c, h\}$, $\{d\}$; the partition corresponding to $\alpha = 0.7$ has the blocks $\{a, e, f\}$, $\{g\}$, $\{b, h\}$, $\{c\}$, $\{d\}$; the partition corresponding to $\alpha = 1.0$ has the blocks $\{a\}$, $\{e, f\}$, $\{g\}$, $\{b\}$, $\{h\}$, $\{c\}$, $\{d\}$.

Similarity creates uncertainty problems. The most serious is the question of what features determine the similarity of objects. For example, given three objects, characterized by the features of color, weight, cost, and the number of legs, where the characterization of object A is $<$white, 6 Kg, USD 75, 4$>$, that of B is $<$white, 7 Kg, USD 80, 4$>$, and that of C is $<$brown, 15 Kg, USD 400, 0$>$, we would assume A to have greater similarity to B than to C. We would probably think differently on learning that A and C are both chairs, but that B is a dog.

5 Fuzzy logic

Here we shall merely sketch some of the approaches to the fuzzification of some concepts of classical logic, particularly with regard to implication. A fuzzy proposition is defined in terms of a universal set U, and a fuzzy set F based on U that represents some fuzzy predicate, such as "tall", "expensive", "high". Element u

	a	b	c	d	e	f	g	h
a	1.0	0	0	0	0.9	0.9	0.5	0
b	0	1.0	0.4	0	0	0	0	0.8
c	0	0.4	1.0	0	0	0	0	0.4
d	0	0	0	1.0	0	0	0	0
e	0.9	0	0	0	1.0	1.0	0.5	0
f	0.9	0	0	0	1.0	1.0	0.5	0
g	0.5	0	0	0	0.5	0.5	1.0	0
h	0	0.8	0.4	0	0	0	0	1.0

Figure 4: Membership grades for a fuzzy relation

of U belongs to F with a membership grade $F(u)$. For example, with regard to a line of clothing we could have a membership function of "expensive" that is zero for prices below USD 50, rises linearly to 1 for prices between USD 50 and USD 300, and is constant at 1 beyond USD 300. Then the membership grade for an item that costs USD 150 is 0.4, and for an item that costs USD 225 it is 0.7. We take this membership grade to be the degree of the truth of the proposition "item x is expensive". The degree of truth of a proposition is thus a value in [0, 1]. Fuzzy union and intersection, as illustrated by Fig.3, are defined in terms of the max and min functions, and so are the disjunction and conjunction of fuzzy propositions. We get

$$A(p \vee q) = \max(A(p), A(q)); \quad A(p \wedge q) = \min(A(p), A(q)).$$

Definitions of fuzzy existential and universal quantifiers follow from this: the max or min operation extends over the entire fuzzy set to which the quantifier is applied. In addition there are specifically fuzzy quantifiers, which may be *absolute* or *relative*. Absolute quantifiers are semi-quantitative, e.g., about 10, much more than 100, at least about 0.5; relative quantifiers are defined on [0,1] — they are qualitative, e.g., almost all, about half, most.

Implication in the classical sense is a mapping from a pair of truth values to a truth value. Representing T and F by 1 and 0, we have:

$$\Im : \{0, 1\} \times \{0, 1\} \rightarrow \{0, 1\}.$$

Classical implication can be expressed in numerous different equivalent forms, such as

$$\Im(p, q) = \overline{p} \vee q;$$

$$\Im(p, q) = \overline{p} \vee (p \wedge q);$$

$$\Im(p, q) = (\overline{p} \wedge \overline{q}) \vee q.$$

The corresponding fuzzy implication function maps from the fuzzy truth values of propositions p and q to the truth value of "if p then q":

$$\Im : [0, 1] \times [0, 1] \to [0, 1].$$

Letting $\mathcal{U}$, $\mathcal{I}$, and $\mathcal{C}$ stand for fuzzy union, intersection, and complement, respectively, we get three fuzzy expressions corresponding to the three expressions above:

$$\Im(p, q) = \mathcal{U}(\mathcal{C}(p), q);$$

$$\Im(p, q) = \mathcal{U}(\mathcal{C}(p), \mathcal{I}(p, q));$$

$$\Im(p, q) = \mathcal{U}(\mathcal{I}(\mathcal{C}(p), \mathcal{C}(q)), q).$$

The problem is that the fuzzy expressions are no longer equivalent. Moreover, if we use different interpretations of the fuzzy operations, each of our expressions results in turn in several non-equivalent implications. For example, there are four well-known expressions corresponding to $\mathcal{U}(\mathcal{C}(p), q)$ alone.

6 Fuzzy control systems

Fuzzy control systems have proven themselves to be very useful in practice. For example, the Sendai subway control system in Japan has been operational since 1986, and its performance has exceeded that of manual operation. The construction of a fuzzy application has four steps.

• The first step is a determination of whether a fuzzy system is appropriate for the application. It is if system knowledge can be expressed as a set of heuristic qualitative rules. If, in the conventional sense, a system has to be defined by a complicated mathematical model, a fuzzy system may be simpler.

• Inputs and outputs are identified, as well as their ranges. The characteristics of measuring devices determine input ranges; the output range is determined by the control actions that are to be performed.

• Membership functions for all input and output parameters are determined, and a rule base is constructed. The rule base defines the control action to be taken for every combination of the inputs. Design of realistic membership functions and control rules is perhaps the hardest task. Usually the task is carried out iteratively.

• The system is validated: a determination is made for various sample inputs that the outputs are within their required ranges. Since it cannot be established analytically that a fuzzy system is stable, extensive testing is required (for the

Sendai subway system there were 300,000 simulation tests and 3.000 riderless subway runs).

The operation of a fuzzy system consists of measurement of input parameters, their fuzzification, determination of the appropriate control rule or rules, and the selection of a control action by a process of defuzzification. For a steam turbine the appropriate inputs are the current temperature and pressure of the steam. These values lead to the selection of appropriate fuzzy sets for the input. These fuzzy sets and the rule base determine a response in the form of a fuzzy set or sets, and defuzzification derives from this output a control action, which for the steam turbine is a throttle setting.

We have chosen an example in which there are two inputs, but the specific nature of these inputs is left undefined. One input, as shown in the upper left diagram of Fig.6, is represented by a single fuzzy membership function. The domain of these input values is split into four distinct non-overlapping regions, A, B, C, and D, so that this input is not really fuzzy in the strict sense. The other input, as shown in the lower left diagram, is represented by five overlapping membership functions. We thus have 4×5 possible input classes, and Fig.5 shows what control action is to be taken for each such class. This is a fuzzy rule base. The control actions are fuzzy, and are denoted S(trong), H(igh), N(ormal), L(ow), and W(eak). For our example, we assume that the first input falls into range B, and corresponds to a membership grade of 0.6. The other input is assumed to correspond to both Low (membership 0.2) and Normal (membership 0.8). Hence two control actions are relevant, H and N. Their membership functions are as shown in the upper right diagram of Fig.6. Now, when two membership grades are combined, their minimum is taken. Hence, for the B & Low input the combined value is 0.6, and for B & Normal it is 0.2.

These values are used to cut off the top portions of the two membership functions of the fuzzy control variable. The result is the shape shown as the lower right diagram of Fig.6. The horizontal axis represents the values of the control variable, e.g., the setting of a fuel valve, or the amount of pressure that is to be applied to the brakes of a vehicle.

In our graph the values of the control variable extend from 15.5 to 23.0. Defuzzification is the process of selecting a single value in this range. For control applications the most popular is the centroid method. It selects a value such that a line extended upward from this value on the horizontal axis splits the composite truncated membership function graph into two equal areas. The value 18.46 does so in our case. Other approaches are possible, e.g., chosing the midpoint of (15.5, 23.0), namely 19.25. The choice makes little difference as long as the system remains stable.

	Weak	Low	Normal	High	Strong
D	H	N	L	W	W
C	H	H	N	L	W
B	S	H	N	N	L
A	S	S	H	H	N

Figure 5: A fuzzy rule base for a control system

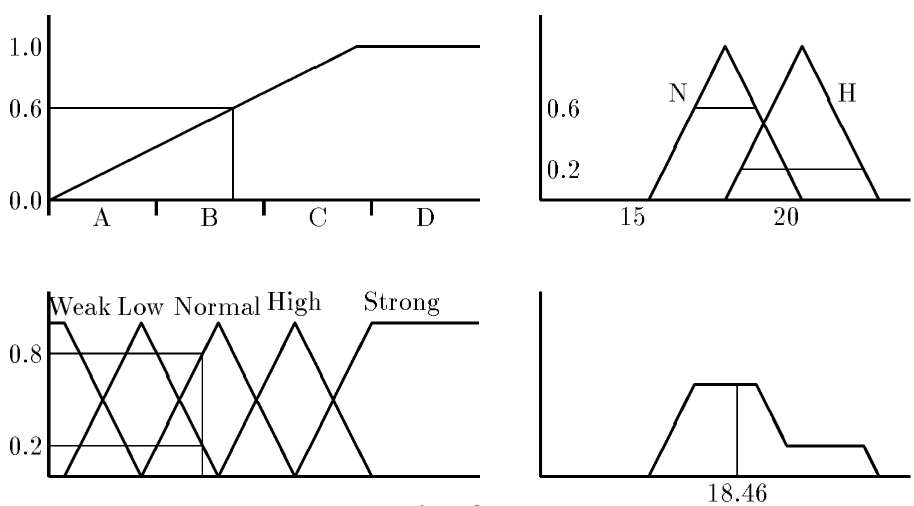

Figure 6: Defuzzification

7 Time Petri nets

For this section we assume some familiarity with basic concepts of Petri nets. To refresh memory, define a Petri net as the tuple $N =< P, T, F, m >$, where

P is a set of places,
T is a set of transitions,
$F \subseteq (P \times T) \cup (T \times P)$ is a flow relation,
m: $P \rightarrow N^+ \cup \{0\}$ is an initial marking that assigns $m(p) \geq 0$ to each place.

A transition t is enabled, i.e., it can fire if every place p_a, such that $< p_a, t >\in F$, contains at least one token. The result of the firing is the loss of a token from each such place p_a and a gain of a token at each place p_b such that $< t, p_b >\in F$. Petri nets can model non-determinacy: suppose place p contains a token, $< p, t_a >$ and $< p, t_b >$ belong to F, and both t_a and t_b are enabled; then either t_a or t_b can fire, and there is nothing in Petri net theory that determines which of the two is to fire or when the firing is to take place.

In time Petri nets the non-determinacy with respect to time is partially removed. A time Petri net is the tuple $N_t =< P, T, F, m, q >$, where P, T, F, m are as before, and $q : T \to (\tau \times \tau)$ defines a time interval within which an enabled transition is to fire. The mapping q provides each transition with a pair of times (u, v). If the transition becomes enabled at time t, then it must fire within the time interval $(t + u, t + v)$, unless it has become disabled before time $t + v$ due to the firing of some other transition.

Figure 7 illustrates the use of time Petri nets to model uncertainty that may arise with real-time processes. Suppose we have a processing task that is to take between α and β time units, i.e., there is some uncertainty as to the length of time the task needs. The period $\delta = \beta - \alpha$ represents the valid time interval during which the task may take place. If it begins too early, Exception I arises; if it takes too long, Exception II goes into effect. The situation is illustrated by the upper diagram of Fig. 7.

In most processes early completion of a task is no problem. An example of where it is a problem arises with package routing. A bar code reader determines the destination of a package, and the package is sent toward a gate that is to open to receive the package. If the package arrives too early, the gate is not yet open; if too late, the gate has closed. In either case it shoots past a closed gate. In setting the times for which the gate is to be open, allowance has to be made for some uncertainty in the time the package needs to reach the gate, but if the gate stays open for too long, packages will enter that are not destined for this gate.

In terms of Fig.7, the task begins when a token arrives in place A, and we take this moment to be time zero. The completion of the task is signaled by the insertion of a token in place C. If this happens before time α, the Exception I transition fires immediately. If the task is taking longer, the transition marked (α, α) fires at time α, and the task moves into a second phase represented by place B. Now, if the task is completed before time β, we get a token in C, and since B already has a token, the lower transition marked (0, 0) fires, and the process continues normally. If the firing of this transition is prevented because the task has not been completed in time, the transition that leads to Exception II fires. If C receives its token after time β, the only enabled transition is the one marked $(\Delta t, \Delta t)$, which merely removes the token from C by "earthing" it.

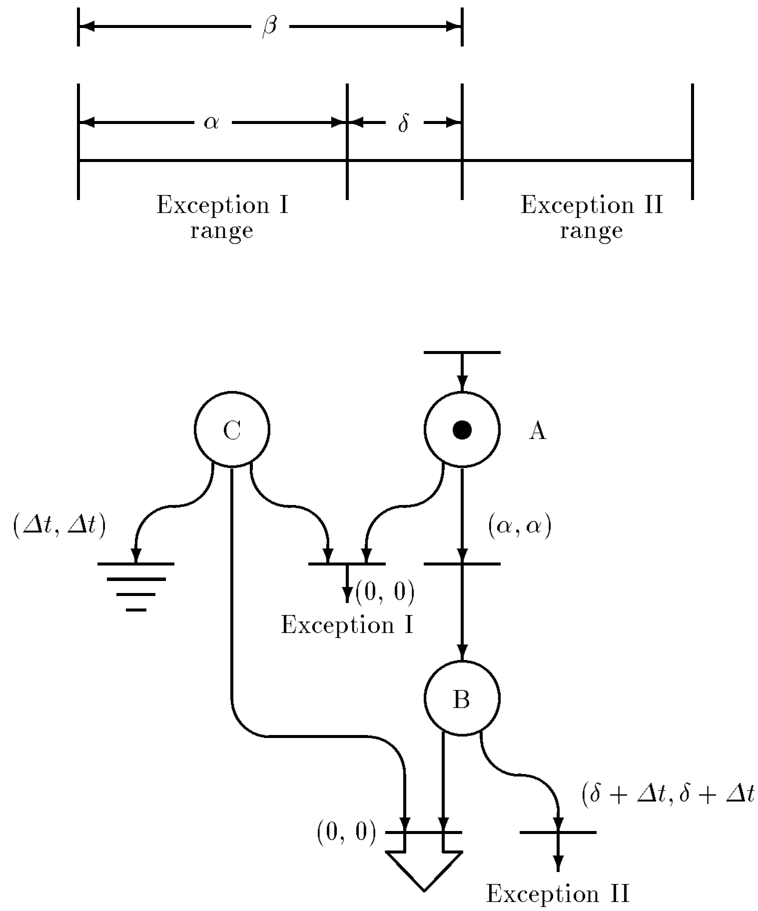

Figure 7: Time Petri net for a real-time task

8 Rough sets

Suppose we are to find X, the set of all immigrants into country C for the year 1985. If the date of immigration for some immigrants is not known, we can construct an approximation, defined by two sets

$$X_* = \{x | ImmigrationDate(x) = 1985\},$$
$$X^* = X_* \cup \{x | ImmigrationDate(x) = ?\}.$$

Then $X_* \subseteq X \subseteq X^*$. The pair (X_*, X^*) is a rough set.

Rough set theory has been developed by Pawlak. The main use of rough sets is in separating large amounts of data into equivalence classes. Suppose that hospital admission records contain the following attribute values for patients:

body temperature:	L(ow), N(ormal), H(igh);
blood pressure:	L(ow), N(ormal), H(igh);
heart rate:	N(ormal), A(bnormal);
weight:	L(ow), N(ormal), H(igh);
insurance:	Y(es), N(o).

The patients can be grouped into $3 \times 3 \times 2 \times 3 \times 2 = 108$ classes. The patients in a particular class are said to be *indiscernible*. Each class is an *elementary set* in rough set terminology. Let us look at the records of eight patients.

	BT	BP	HR	Wt	Ins
p1	N	H	A	H	Y
p2	H	H	A	H	N
p3	H	N	N	N	Y
p4	H	L	N	L	Y
p5	N	L	A	L	Y
p6	H	H	A	H	N
p7	H	H	N	L	N
p8	L	L	A	L	Y

The attributes, separately and in combination, generate various elementary sets. Thus blood pressure generates {p1, p2, p6, p7}, {p3}, and {p4, p5, p8}, which correspond to H, N, and L, respectively; weight generates {p1, p2, p6}, {p3}, and {p4, p5, p7, p8}; and blood pressure and weight in combination generate {p1, p2, p6}, {p3}, {p4, p5, p8}, {p7}, which correspond to HH, NN, LL, and HL, respectively.

Suppose now that analysis of very many patient records shows that some combinations of BP and Wt can determine whether or not a patient has insurance.

BP	H	H	H	N	N	N	L	L	L
Wt	H	N	L	H	N	L	H	N	L
Ins	Y/N	Y	N	Y/N	Y	Y	Y/N	Y	Y

Patients in the HN, NN, NL, LN, and LL groups all have insurance, patients in the HL group do not, and about the rest we cannot tell. Hence, if we were asked which of our eight patients have insurance, and we only had the blood pressure and weight data to go by, we could say that p3, p4, p5, p8 definitely have it, but that we are not sure about p1, p2, and p6 (p7 definitely does not have it). The set of patients with insurance has thus the lower approximation {p3, p4, p5, p8}, and the upper approximation {p1, p2, p3, p4, p5, p6, p8}, which together define a rough set.

To put our discussion on a formal basis, introduce U, the universe of discourse, i.e., the set of all entities of interest for a particular application, A, the set of attributes that apply to elements of U, and $B \subseteq A$, the attributes under particular consideration. Define the indiscernability relation R_B:

$$R_B(x, y) \text{ iff } a(x) = a(y) \text{ for all } a \in B.$$

Let U/R_B (or simply U/B) be the partition induced on U by R_B. The blocks of U/B are B-elementary sets. Denote by $B(x)$ the block of U/B that contains element x. Every subset X of U has its B-lower and B-upper approximations:

$$B_*(X) = \{x \in U | B(x) \subseteq X\},$$

$$B^*(X) = \{x \in U | B(x) \cap X \neq \emptyset\}.$$

Thus, $B_*(X)$ is the union of all B-elementary sets that are contained in X, and $B^*(X)$ is the union of all B-elementary sets that overlap X. Sometimes $B_*(X)$ and $B^*(X)$ are written as $\underline{B}$ and $\overline{B}$, respectively. The accuracy of the approximation of set X with respect to B, written $\alpha_B(X)$, is the ratio $|B_*(X)|/|B^*(X)|$. Clearly $\alpha_B(X)$ lies between 0 and 1. The B-boundary of X is the set

$$BN_B = B^*(X) - B_*(X).$$

Rough sets can be defined in terms of the *rough membership function*:

$$\mu_X^B(x) = \frac{|X \cup B(x)|}{|B(x)|}.$$

Since we do not know $|X|$, this function is of no direct use, but it makes some concepts clearer. We have

$$B_*(X) = \{x \in U | \mu_X^B(x) = 1\},$$

$$B^*(X) = \{x \in U | \mu_X^B(x) > 0\},$$

$$BN_B(X) = \{x \in U | 0 < \mu_X^B(x) \leq 1\}.$$

Also,

$$\mu_{X \cup Y}^B(x) \geq \max(\mu_X^B(x), \mu_Y^B(x)),$$

$$\mu_{X \cap Y}^B(x) \leq \min(\mu_X^B(x), \mu_Y^B(x)).$$

Recall that the corresponding relationships for fuzzy memberships are equalities. This shows that fuzzy sets are a specialization of rough sets, but there is a difference between the two. Under fuzzy set theory we can regard, say, the blood pressure of a patient as both low and normal, but with different membership values. Under rough set theory the blood pressure is either low or normal, but the cardinalities of the sets of patients with low and normal blood pressure are uncertain. Whereas fuzzy sets are well suited for control systems, rough sets work well for classification.

Rough sets appear to be particularly well suited for data mining, which is the detection of significant relationships in data, particularly in data warehouses. This requires the identification of dependencies between attributes. A set of attributes D depends totally on a set of attributes C, denoted $C \Rightarrow D$, if values of attributes of D are uniquely determined by values of attributes of C. We call C the condition attributes, and D the decision attributes. More common is partial dependency, denoted $C \overset{\gamma}{\Rightarrow} D$, The parameter $\gamma(C, D)$, which takes values between 0 and 1, indicates the strength of the dependency:

$$\gamma(C, D) = \sum_{X \in U/D} \frac{|C_*(X)|}{|U|},$$

where U/D is the partition of U induced by D.

To illustrate dependencies we take a population of six people. The attributes are income (INC), with values {Low, Medium, High}, home ownership (HO), ownership of a PC (PC), and ownership of a mobile phone (MOB). We shall consider the first three to be condition attributes, and the last to be a decision attribute.

	INC	HO	PC	MOB
p1	Low	Yes	No	No
p2	Medium	No	Yes	Yes
p3	Medium	Yes	No	Yes
p4	High	Yes	Yes	Yes
p5	High	Yes	No	Yes
p6	Medium	No	Yes	No

The dependence of D = MOB on C = {INC, HO, PC} is not total — the condition attributes are the same for p2 and p6, but p2 does have a mobile while p6 does not. Hence $\gamma(C, D) = (1+1+1+1)/6 = 4/6$. But we also get $\gamma(C, D) = 4/6$ with C = {INC, HO} and with C = {INC, PC}. This shows that either of the attributes HO or PC is redundant, but not both. We have γ(INC, MOB) = 3/6, but γ(HO, MOB) = γ(PC, MOB) = 0.

A set of conditions $C'(C' \subset C)$ is a *D-reduct* of C if C' is a minimal subset of C such that

$$\gamma(C, D) = \gamma(C', D).$$

The intersection of all D-reducts is called a D-core. Here we have D-reducts {INC, PC} and {INC, HO}, so that the D-core is {INC}. A core contains the most important attributes for an application in that removal of a core-attribute reduces the classification power.

9 Belief and evidence

Suppose we assign probability 0.6 to the proposition that the manager of the Internet company Hype.com is competent and trustworthy, and the manager states that in the last business quarter Hype.com lost only USD 20,000,000. Under the belief-function approach a belief of 0.6 is assigned to the manager's statement. This is an indirect approach: a confidence measure for one statement results in a confidence measure for another statement.

Now look at this from another angle: we gave a 0.4 probability that the manager is incompetent or not to be trusted. Should we say that the loss figure is to be disbelieved at a 0.4 level? Not necessarily. We may, for example, set this belief level at 0.2 — in common speech this is expressed as "giving the benefit of doubt" (also, the positive belief can be less than 0.6). A belief cannot be put above the corresponding probability, but it can be lower. This has led to the use of gambling terminology in some discussions of belief and evidence: the 0.6 belief is interpreted as an offer to give 6:10 odds on it, and the 0.2 belief as an offer of 2:10 odds. Using the symbol Bel to denote belief, $Bel(p \wedge \overline{p}) = 1$ for any proposition p. But in our example $Bel(p) = 0.6$ and $Bel(\overline{p}) = 0.2$, so that in general $Bel(p) + Bel(\overline{p}) \leq 1$.

In terms of sets, given a finite universal set U, the belief function Bel is a mapping from the power set of U to the unit interval,

$$Bel : \mathcal{P}(U) \to [0, 1],$$

with properties

$$Bel(\emptyset) = 0, \quad Bel(U) = 1;$$

$$Bel(A) + Bel(\overline{A}) \leq 1.$$

For a set $A \subset U$, $Bel(A)$ is the degree of belief (based on available evidence) that an element of U belongs to A.

Plausibility Pl is the dual of Bel:

$$Pl(A) = 1 - Bel(\overline{A}), \quad Bel(A) = 1 - Pl(\overline{A});$$

$$Pl(\emptyset) = 0, \quad Pl(U) = 1;$$

$$Pl(A) + Pl(\overline{A}) \geq 1.$$

Introduce a basic probability assignment,

$$m : \mathcal{P}(U) \to [0, 1],$$

such that $m(\emptyset) = 0$ and

$$\sum_{A \in \mathcal{P}(U)} m(A) = 1.$$

For each subset A of U, $m(A)$ represents the proportion to which available evidence supports the claim that a particular element of the universe belongs to A. If $|U| = 8$, then $|\mathcal{P}(U)| = 256$, which means that the average value of m is 0.004. However, for most subsets of U we can expect the value to be zero, so that the non-zero values will not be very low. Consider the sets A, B, C, and D shown in Fig.8, where D is also the universal set. Although here the powerset has 16 elements, not all are distinct. For example, AB = A, where we are adopting a convention under which AB stands for A∪B. Then the only sets of interest are A, B, C, D, AC, and BC.

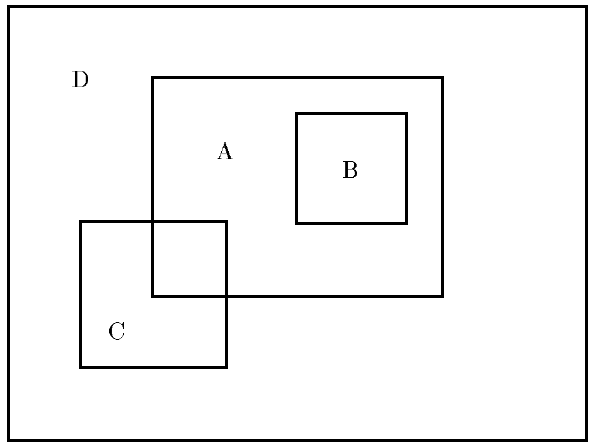

Figure 8: A universal set with subsets

We have

$$Bel(A) = \sum_{B|B\subseteq A} m(B), \qquad Pl(A) = \sum_{B|A\cap B\neq\emptyset} m(B).$$

Explanation: $m(A)$ measures the degree of belief or evidence that an element of U belongs to set A and only to set A; $Bel(A)$ represents the belief that the element belongs to A as well as to subsets of A; $Pl(A)$ represents the belief that it belongs to A, to any of its subsets, and also to subsets that overlap A (hence $Pl(A) \geq Bel(A)$). With respect to Fig.8,

$$Bel(A) = m(A) + m(B);$$

$$Pl(A) = m(A) + m(B) + m(C) + m(D) + m(AC).$$

A set A with $m(A) > 0$ is a *focal element* of m. It is a subset of U on which the available element focuses. The set of pairs $\{< A, m > | m(A) > 0\}$ is a *body of evidence.*

Example 1. Let the universe consist of all possible cost estimates of a project, bounded by say USD 30,000 and 100,000. Define three sets on these costs:

$$30,000 \leq A < 50,000$$

$$50,000 \leq B < 70,000$$

$$70,000 \leq C < 100,000$$

We construct a body of evidence, where the belief in particular cost estimates is based on past experience:

A	B	C	AB	AC	BC	ABC
0.05	0.20	0.15	0.15	0	0.35	0.10

Example 2. We take the universe to consist of three causes, a, b, c. An observed phenomenon (symptom) may be due to a single cause (a disease), or two causes in combination, or all causes. Here we assume the body of evidence to be as follows:

a	b	c	ab	ac	bc	abc
0.30	0.15	0.20	0.10	0.15	0.05	0.05

Let us return to Example 1. Suppose that another expert arrives at a different body of evidence:

A	B	C	AB	AC	BC	ABC
0.05	0.15	0.15	0.20	0	0.30	0.15

Evidence theory provides a method for combining evidence. Given evidence from two (independent) sources, a joint basic probability assignment is given by Dempster's rule:

$$m_{1,2}(A) = \frac{\sum_{B \cap C = A} m_1(B) \times m_2(C)}{1 - \sum_{B \cup C = \emptyset} m_1(B) \times m_2(C)}.$$

Use of the m-values, and their combination by means of Dempster's rule is known as the Dempster-Shafer theory of evidence.
Example 3. If we combine the two sets of probabilities for the software cost estimation problem by means of Dempster's rule, the result is as follows:

	A	B	C	AB	AC	BC	ABC
m_1	0.05	0.20	0.15	0.15	0	0.35	0.10
m_2	0.05	0.15	0.15	0.20	0	0.30	0.15
$m_{1,2}$	0.039	0.440	0.190	0.087	0	0.226	0.018

Let us now look at how some of the values were arrived at. First we compute the denominator in Dempster's rule, which acts as normalizer to ensure that the $m_{1,2}$-values add up to 1. Denote it by K. Then we show how the values for $m_{1,2}(B)$ and $m_{1,2}(AB)$ were arrived at.

$$\begin{aligned}
K &= 1 - [m_1(A)m_2(B) + m_1(A)m_2(C) + m_1(A)m_2(BC) \\
&\quad + m_1(B)m_2(A) + m_1(B)m_2(C) + m_1(B)m_2(AC) \\
&\quad + m_1(C)m_2(A) + m_1(C)m_2(B) + m_1(C)m_2(AB) \\
&\quad + m_1(AB)m_2(C) + m_1(AC)m_2(B) + m_1(BC)m_2(A)] \\
&= 1 - (0.0075 + 0.0075 + \cdots + 0.0175) \\
&= 0.8300; \\
m_{1,2}(B) &= [m_1(B)m_2(B) + m_1(B)m_2(AB) + m_1(B)m_2(BC)
\end{aligned}$$

$$
\begin{aligned}
&+ m_1(B)m_2(ABC) + m_1(AB)m_2(B) + m_1(AB)m_2(BC) \\
&+ m_1(BC)m_2(B) + m_1(BC)m_2(AB) + m_1(ABC)m_2(B)]/K \\
&= (0.0300 + 0.0400 + \cdots + 0.0150)/0.8300 \\
&= 0.440; \\
m_{1,2}(AB) &= [m_1(AB)m_2(AB) + m_1(AB)m_2(ABC) + m_1ABC)m_2(AB)]/K \\
&= (0.0300 + 0.0225 + 0.0200)/0.8300 = 0.087.
\end{aligned}
$$

10 Possibility theory

Zadeh has been concerned with linguistic values, which leads to the concept of "computing with words." The basis for this are possibility distributions. Consider V, a variable defined on a set X. This set can consist of numbers, but also of colors, employees, etc. Instead of allowing V to take just single values, we shall also allow fuzzy subsets of X as its values. Then we can state not only that "*John's age* is 25", but also that "*John's age* is *young*." In the latter case the variable *John's age* has assumed a linguistic value and *young* is represented by a fuzzy subset of X. In terms of a predicate *Age* we have that $Age(John, 25)$ is true or false, but $Age(John, young)$ does not work out so well. We can be fairly sure that John is not over 50, and all values up to 20 are certainly possible, but the middle range is uncertain. To measure such uncertainties, Zadeh introduced the concept of *possibility*.

In probability theory we have probability distribution functions

$$p : X \to [0,1], \text{ with } \sum_{x \in X} p(x) = 1.$$

For a subset A of X, the probability measure for this set is

$$Prob(A) = \sum_{x \in A} p(x).$$

Possibility distribution functions are the counterpart of this. For each $x \in X$, the possibility of x being a value of V (*John's age* in our example) in terms of the linguistic variable A (represented by a fuzzy subset of A), is written $\pi(x)$ — some authors use $r(x)$. Then

$$\pi : X \to [0,1], \text{ with } \max_{x \in X} \pi(x) = 1,$$

such that $\pi(x) = A(x)$, where $A(x)$ is the membership grade of x in A, a fuzzy subset of X. The counterpart of $Prob(A)$ is $Poss(A)$, with the possibility measure for set A given by

$$Poss(A) = \max_{x \in A} \pi(x).$$

Necessity is a measure related to possibility:

$$Nec(A) = 1 - Poss(\overline{A}).$$

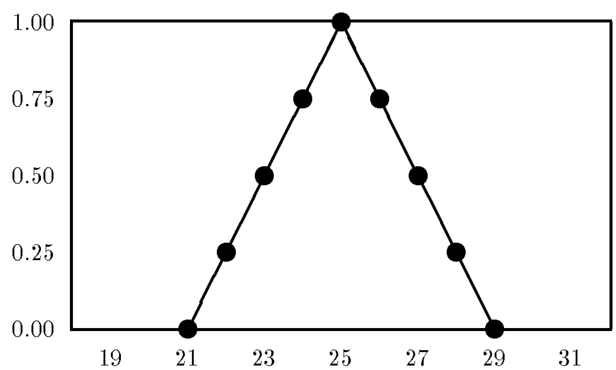

Figure 9: A possibility distribution

We can assume that the possibility distribution for John being young is 1.0 up to age 20, and that it then drops linearly to reach 0 at age 50. For any age up to 100, say, and any given person, if we have no additional knowledge, the possibility that this is the person's age is 1.0. Probability theory in a case like this assumes every age to be equally probable, so, given that the age has to be in [0, 100], all probabilities are 0.01.

In probability theory ignorance is interpreted as complete randomness, which implies a uniform distribution. But there is nothing that says that a uniform distribution is appropriate. In possibility theory one merely states that all values are equally possible, which does not imply that they are equally probable. Possibility statements are weaker than probability statements, and this is how it should be — it allows possibility theory to deal with states of knowledge that range from total ignorance to total information in a natural and simple manner.

Possibilities may be linked to bodies of evidence $< A, m >$. For this we have to use α-cut, which were introduced in Sec. 5. For fuzzy set A,

$$\alpha_1 \leq \alpha_2 \ \rightarrow \ {}^{\alpha_1}A \supseteq {}^{\alpha_2}A.$$

Example. Let V be the set of ages, and consider the statement "V is about 25". The concept "about 25" is represented by fuzzy set T. The possibility distribution π with respect to "about 25" has the same values as the membership function T. We show it as Fig.9.

We have the following significant α-cuts:

$\alpha = 1.00:$ $A_1 = \{25\};$
$\alpha = 0.75:$ $A_2 = \{24, 25, 26\};$
$\alpha = 0.50:$ $A_3 = \{23, 24, 25, 26, 27\};$
$\alpha = 0.25:$ $A_4 = \{22, 23, 24, 25, 26, 27, 28\};$
$\alpha = 0.00:$ $A_5 = \{21, 22, 23, 24, 25, 26, 27, 28, 29\}.$

These α-cuts correspond to focal elements, i.e., they are sets $A \in \mathcal{P}(V)$ such that $m(A) > 0$. It can be shown that for values $\pi_1, \pi_2, \ldots, \pi_n$,

$$\pi_i = \sum_{k=i}^{n} m_k,$$

from which

$$\pi_{i-1} = \sum_{k=i-1}^{n} m_k,$$

giving $m_i = \pi_i - \pi_{i+1}$, with $\pi_{n+1} = 0$ by convention.

Then, for our example, as expected,

$m(A_1$ = 1.00 - 0.75 = 0.25;
$m(A_2$ = 0.75 - 0.50 = 0.25;
$m(A_3$ = 0.50 - 0.25 = 0.25;
$m(A_4$ = 0.25 - 0.25 = 0.25.

11 Bibliographic notes

In the introduction we referred to the book by Klir and Yuan [1] as possibly the best reference on uncertainty. This encyclopaedic survey goes well beyond just fuzziness, and a literature search can start with its 1731 references, which are classified into categories. Dyreson [3] has generated a bibliography of uncertainty management for information systems. A collection of readings edited by Shafer and Pearl [4] deals with nearly all aspects of uncertainty in the context of AI reasoning systems. However, some of the papers may be somewhat advanced for a non-specialist. Measures of uncertainty in expert systems are considered in [5]. A survey by Klir [6] can also be recommended, and so can the survey by Parsons [7] on imperfect information in data and knowledge bases. The Parson's survey has 172 references, and an earlier survey by Ng and Adamson [8] has 75.

The literature on probability is, of course, overwhelming. For software and knowledge engineers the best choice may be a text addressed to engineers and scientists. Such texts have a high degree of rigor without getting into topics of merely theoretical interest. Walpole and Myers [9] is a good choice. The book by Allen [10] is addressed specifically to computer scientists, but its examples are somewhat dated for this Internet age. A collection of papers [11] provides examples of application of the Bayesian approach. The Bayesian approach has been used in a study of empirical software engineering cost models [12].

An outgrowth of the Bayesian method is the study of Bayesian networks [13]. Here the description of a system is separated into a qualitative and a quantitative component. The qualitative component represents relationships between variables or events by means of arcs of a graph; the quantitative component associates conditional probabilities with the arcs. The guest editors' introduction to a special section of the IEEE Transactions on Knowledge and Data Engineering [14] lists 43 references, most of which are recent; the special section contains 4 articles. An earlier survey provides 166 references [15]. Although often our aim in studying relationships is to discover cause-effect patterns, a relationship does not have to be a causal relationship. This aspect of uncertainty is studied in great detail in [16] (the book contains 379 references).

There are very many books on fuzzy sets and their extensions. Five are listed here, where my criterion for listing these books is that I have them. As noted earlier, [1] is the main reference. A book by Cox [17] takes a very practical approach, and the collection edited by Marks [18] contains papers, many of which are experience reports, from various IEEE conferences that took place in 1992 and 1993. Yager and Filev [19] give additional examples. The Terano *et al* book [20] complements the others in that it describes some successfully implemented fuzzy systems in Japan, where the earliest practical applications of fuzziness were made — of particular interest is the data on the Sendai Subway system. Two applications of fuzzy control in automobile design are discussed in [21, 22]. Defuzzification is very important in such work. Runkler [23] presents some of the more important defuzzification methods and investigates their properties. Three examples are used. Zadeh [24] discusses the notion that fuzzy logic is computing with words. Yen [25] brings new insights to fuzzy logic. For a symposium on fuzzy logic see the August 1994 issue of IEEE Expert — the symposium discussion was based on a paper by Elkan [26], which is a formal presentation of some misgivings about fuzzy logic. Mathematicians will find [27] informative.

Similarity, which we considered as a type of fuzzy relationship, has been defined as indistinguishability under an appropriate resolution [28]. Although the title of the paper does not suggest this, similarity is studied very thoroughly in [29]. A particular aspect of similarity arises with identification. Since no two signatures by the same person are the same, the recognition of a signature depends on how similar it is to a reference signature. It is nearly as important to prevent rejection of a true signature as to accept a falsification. Because of the difficulty of achieving this, the use of biometric features has become popular, particularly identification based on unique characteristics of the human iris [30]. A major application of similarity arises in the retrieval of reusable software components — see [31–33] on this. Similarity plays also an important role in case-based reasoning [34].

On Petri nets in general the books by Peterson [35] and Reisig [36] give excellent introduction. Although published a while back, they are still current as regards the needs of practitioners. The topic of time Petri nets is covered thoroughly in [37]. Fuzzy Petri nets have been used for the representation of knowledge [38].

The literature on rough sets is not easy to read in that the most active researchers in this area make use of rather advanced mathematics, but [39] is a fairly accessible introduction. The major source of information on rough sets and their application is [40, 41]; [40] contains an excellent tutorial by Pawlak. For another example of the application of rough sets see [42].

A very important uncertainty problem relates to the quality of data. There are two issues. First, the data in a data base are to be dependable, i.e., they are to model an up-to-date aspect of the world. Second, improper decisions are to be prevented when data in support of real-time decisions come from various sources of variable dependability. The second issue can be expressed as follows: given a set of evidence $\{E_j\}$, which consists of reported facts or measurements, and for each, a measure W_j of the data quality, i.e., a measure of its accuracy or of confidence in its source, decide among an exhaustive set of mutually exclusive hypotheses $\{H_i\}$. Dillard [43] considers several decision methods for selecting the appropriate hypothesis, including Dempster's rule, but in a simplified form. For a later investigation of the effect of data quality on decision making see [44]. In [45] the Dempster-Shafer theory is combined with fuzzy inferencing. The Dempster-Shafer theory is the basis for Yager's work on multicriteria decisions [46].

Management of inconsistency in software development is a major concern of software engineers. Because this topic has been recently surveyed by Spanoudakis and Zisman [47] we have not discussed it here. We merely draw attention to the distinction made by Robinson and Pawlowski [48] between inconsistencies, which are technical, and conflicts, which have a social basis. Awareness of this distinction can be useful in the management of uncertainties in other areas as well.

The management of uncertainty has been extensively studied in the context of information systems. The main topics have been uncertainty in general [49, 50], null values in data bases [51–53], avoidance of null responses to queries [54–56], vague and fuzzy queries [57–60], and the association of probabilities with attributes [61, 62]. The plausibility concept has been used in data base design [63].

References

1. G.J. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic* (Prentice Hall PTR, 1995).
2. A.T. Berztiss, "Software methodologies for decision support", *Information & Management* **18** (1990) 221–229.
3. C.E. Dyreson, "A bibliography on uncertainty management in information systems", A. Motro and P. Smets (eds.), *Uncertainty Management in Information Systems* (Kluwer, 1997) 413–458.
4. G. Shafer and J. Pearl (eds.), *Readings in Uncertain Reasoning* (Morgan Kaufmann, 1990).
5. P. Walley, "Measures of uncertainty in expert systems", *Artificial Intelligence* **83** (1996) 1–58.
6. G.J. Klir, "Developments in uncertainty-based information", *Advances in Computers* **36** (1993) 255–332.
7. S. Parsons, "Current approaches to handling imperfect information in data and knowledge bases", *IEEE Transactions on Knowledge and Data Engineering* **8** (1996) 353–372.
8. K.-C. Ng and B. Abramson, "Uncertainty management in expert systems", *IEEE Expert* **5**, no. 2 (April 1990) 29–48.
9. R.E. Walpole and R.H. Myers, *Probability and Statistics for Engineers and Scientists, 4th ed.* (Macmillan, 1989).

10. A.O. Allen, *Probability, Statistics, and Queueing Theory with Computer Science Applications* (Academic Press, 1978).
11. S. French and J.Q. Smith (eds.), *The Practice of Bayesian Analysis* (Arnold with Wiley, 1997).
12. S. Chulani, B. Boehm and B. Steece, "Bayesian analysis of empirical software engineering cost models", *IEEE Transactions on Software Engineering* **25** (1999) 573–583.
13. J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference* (Morgan Kaufmann, 1988).
14. M.J. Druzdzel and L.C. v. d. Gaag, "Building probabilistic networks: 'Where do the numbers come from?'", *IEEE Transactions on Knowledge and Data Engineering* **12** (2000) 481–486.
15. W. Buntine, "A guide to the literature on learning probabilistic networks from data", *IEEE Transactions on Knowledge and Data Engineering* **8** (1996) 195–210.
16. J. Pearl, *Causality: Models, Reasoning, and Inference* (Cambridge University Press, 2000).
17. E. Cox, *The Fuzzy Systems Handbook* (Academic Press, 1994).
18. R.J. Marks (ed.), *Fuzzy Logic Technology and Applications* (IEEE, 1994).
19. R.R. Yager and D.P. Filev, *Essential Fuzzy Modeling and Control* (Wiley, 1994).
20. T. Terano, K. Asai and M. Sugeno, *Applied Fuzzy Systems* (Academic Press, 1994) [Japanese edition in 1989].
21. R. Müller and G. Nöcker, "Intelligent cruise control with fuzzy logic", in [17], pp. 74–80.
22. G.F. Maurer, "A fuzzy logic controller for an ABS braking system", *IEEE Transactions on Fuzzy Systems* **3** (1995) 381–388.
23. T.A. Runkler, "Selection of appropriate defuzzification methods using application specific properties", *IEEE Transactions on Fuzzy Systems* **5** (1997) 72–79.
24. L.A. Zadeh, "Fuzzy logic = computing with words", *IEEE Transactions on Fuzzy Systems* **4** (1996) 103–111.
25. J. Yen, "Fuzzy logic — A modern perspective", *IEEE Transactions on Knowledge and Data Engineering* **11** (1999) 153-165.
26. C. Elkan, "The paradoxical success of fuzzy logic", *IEEE Expert* **9**, no. 4 (August 1994) 3–8.
27. E. Turunen, *Mathematics Behind Fuzzy Logic* (Physica-Verlag, 1999).
28. K. Georgatos, "Resolution spaces: A topological approach to similarity", *Proceedings of the 11th International Workshop on Database and Expert Systems Applications* (IEEE CS Press, 2000) 553–557.
29. S. Castano, G. Martella and P. Samarati, "Analysis, comparison, and design of role-based security specifications", *Data & Knowledge Engineering* **21** (1997) 31–55.
30. M. Negin et al, "An iris biometric system for public and personal use", *Computer* **33**, no. 2 (February 2000) 70–75.
31. E. Ostertag, J. Hendler, R.P. Diaz and C. Braun, "Computing similarity in a reuse library system: An AI-based approach", *ACM Transactions on Software Engineering Methodology* **1** (1992) 205–228.
32. M.G. Fugini and S. Faustle, "Retrieval of reusable components in a development information system", *Proceedings of the 3nd International Conference on Software Reusability* (1994) 89–98.
33. D. Merkl, A.M. Tjoa and G. Kappel, "Learning the semantic similarity of reusable software components", *Proceedings of the 3nd International Conference on Software Reusability* (1994) 33–41.

34. K.-D. Althoff, "Case based reasoning", this volume.
35. J.L. Peterson, *Petri Net Theory and the Modeling of Systems* (Prentice-Hall, 1981).
36. W. Reisig, *Petri Nets — An Introduction* (Springer-Verlag, 1985).
37. D. Berthomieu and M. Diaz, "Modeling and verification of time dependent systems using time Petri nets", *IEEE Transactions on Software Engineering* **17** (1991) 259–273.
38. S.-M. Chen, J.-S. Ke and J.-F. Chang, "Knowledge representation using fuzzy Petri nets", *IEEE Transactions on Knowledge and Data Engineering* **2** (1990) 311–319 [correspondence on this paper: *ibid* **7** (1995) 190–192].
39. Z. Pawlak, J. Grzymala-Busse, R. Slowinski and W. Ziarko, "Rough sets", *Communications of the ACM* **38**, no. 11, November 1995, 88–95.
40. L. Polkowski and A. Skowron (eds.), *Rough Sets in Knowledge Discovery 1: Methodology and Applications* (Physica-Verlag, 1998).
41. L. Polkowski and A. Skowron (eds.), *Rough Sets in Knowledge Discovery 2: Applications, Case Studies and Software Systems* (Physica-Verlag, 1998).
42. C.-C. Chan, "Incremental learning of production rules from examples under uncertainty: A rough set approach", *International Journal of Software Engineering and Knowledge Engineering* **1** (1991) 439–461.
43. R.A. Dillard, "Using data quality measures in decision-making algorithms", *IEEE Expert* **7**, no. 6 (December 1992) 63–72.
44. I.N. Chengalur-Smith, D.P. Ballou and H.L. Pazer, "The impact of data quality information on decision making: An exploratory analysis", *IEEE Transactions on Knowledge and Data Engineering* **11** (1999) 853–864.
45. C. Romer and A. Kandel, "Applicability analysis of fuzzy inference by means of generalized Dempster-Shafer theory", *IEEE Transactions on Fuzzy Systems* **3** (1995) 448–453.
46. R.R. Yager, "Hierarchical aggregation functions generated from belief structures", *IEEE Transactions on Fuzzy Systems* **8** (2000) 481–490.
47. G. Spanoudakis and A. Zisman, "Inconsistency management in software engineering: Survey and open research issues", this handbook, volume 1.
48. W. Robinson and S. Pawlowski, "Managing requirements inconsistency with development goal monitors", *IEEE Transactions on Software Engineering* **25** (1999) 816–835.
49. J.M. Morrissey, "Imprecise information and uncertainty in information systems", *ACM Transactions on Information Systems* **8** (1990) 159–180.
50. A. Yazici, B.P. Buckles and F.E. Petry, "Handling complex and uncertain information in the ExIFO and NF^2 data models", *IEEE Transactions on Fuzzy Systems* **7** (1999) 659–676.
51. K.-C. Liu and R. Sunderraman, "Indefinite and maybe information in relational databases", *ACM Transactions on Database Systems* **15** (1990) 1–39.
52. K.-C. Liu and R. Sunderraman, "A generalized relational model for indefinite and maybe information", *IEEE Transactions on Knowledge and Data Engineering* **3** (1991) 65–77.
53. Q. Kong and G. Chen, "On deductive databases with incomplete information", *ACM Transactions on Information Systems* **13** (1995) 354–369.
54. T. Ichikawa and M. Hirakawa, "ARES: A relational database with the capability of performing flexible interpretation of queries", *IEEE Transactions on Software Engineering* **SE-12** (1986) 624–634.
55. M. Kao, N. Cercone and W.-S. Luk, "Providing quality responses with natural language interfaces: The null value problem", *IEEE Transactions on Software Engineering* **14** (1988) 959–984.

56. A. Motro, "FLEX: A tolerant and cooperative user interface to databases", *IEEE Transactions on Knowledge and Data Engineerine* **2** (1990) 231–246.
57. A.T. Berztiss, "Imprecise queries and the quality of conceptual models", *Information Modelling and Knowledge Bases V* (IOS Press, 1994) 174–185.
58. P. Bosc and O. Pivert, "SQLf: A relational database language for fuzzy querying", *IEEE Transactions on Fuzzy Systems* **3** (1995) 1–17.
59. A.T. Berztiss, "Domain analysis and queries in context", *Proceedings of the 5th International Conference on Application of Natural Languages in Information Systems* (2001, to be published in the Springer LNCS series).
60. G. S. Loo and K.-H. Lee, "An interface to databases for flexible query answering: A fuzzy-set approach", *Proceedings of the 10th International Conference on Database and Expert Systems Applications (Lecture Notes in Computer Science 1873)* (Springer, 2000) 654–663.
61. D. Barbara, H. Garcia-Molina and D. Porter, "The management of probabilistic data", *IEEE Transactions on Knowledge and Data Engineering* **4** (1992) 487–502.
62. L. Lakshmanan, N. Leone, R. Ross and V. Subrahmanian, "ProbView: A flexible probabilistic database system", *ACM Transactions on Database Systems* **22** (1997) 419–469.
63. M. Altus, "Meta-modeling in conceptual data base design using a plausibility function", *Information Modelling and Knowledge Bases VIII* (IOS Press, 1997) 113–128.