journal of statistical planning and inference

Journal of Statistical Planning and Inference 105 (2002) 233–264

www.elsevier.com/locate/jspi

Toward a perception-based theory of probabilistic reasoning with imprecise probabilities ☆

Lotfi A. Zadeh *

Computer Science Division and the Electronics Research Laboratory, Department of EECS, University of California, Berkeley, CA 94720-1776, USA

Received 5 November 1999; received in revised form 29 February 2000; accepted 25 January 2001

To Herbert Robbins, Hung T. Nguyen and the memory of Professor Kampe de Feriet

Abstract

The perception-based theory of probabilistic reasoning which is outlined in this paper is not in the traditional spirit. Its principal aim is to lay the groundwork for a radical enlargement of the role of natural languages in probability theory and its applications, especially in the realm of decision analysis. To this end, probability theory is generalized by adding to the theory the capability to operate on perception-based information, e.g., "Usually Robert returns from work at about 6 p.m." or "It is very unlikely that there will be a significant increase in the price of oil in the near future". A key idea on which perception-based theory is based is that the meaning of a proposition, p, which describes a perception, may be expressed as a generalized constraint of the form X is*r* R, where X is the constrained variable, R is the constraining relation and is*r* is a copula in which r is a discrete variable whose value defines the way in which R constrains X. In the theory, generalized constraints serve to define imprecise probabilities, utilities and other constructs, and generalized constraint propagation is employed as a mechanism for reasoning with imprecise probabilities as well as for computation with perception-based information.

MSC: primary 62A01; secondary 03B52

Keywords: Perception-based information; Fuzzy set theory; Fuzzy logic; Generalized constraints; Constraint languages

1. Introduction

Interest in probability theory has grown markedly during the past decade. Underlying this growth is the ballistic ascent in the importance of information technology. A related cause is the concerted drive toward automation of decision-making in a wide variety of fields ranging from assessment of creditworthiness, biometric authentication,

☆ Program of UC Berkeley.

* Tel.: +1-510-642-4959; fax: +1-510-642-1712.

E-mail address: zadeh@cs.berkeley.edu (L.A. Zadeh).

0378-3758/02/$ - see front matter
PII: S0378-3758(01)00212-9

and fraud detection to stock market forecasting, and management of uncertainty in knowledge-based systems. Probabilistic reasoning plays a key role in these and related applications.

A side effect of the growth of interest in probability theory is the widening realization that most real-world probabilities are far from being precisely known or measurable numbers. Actually, reasoning with imprecise probabilities has a long history (Walley, 1991) but the issue is of much greater importance today than it was in the past, largely because the vast increase in the computational power of information processing systems makes it practicable to compute with imprecise probabilities—to perform computations which are far more complex and less amenable to precise analysis than computations involving precise probabilities.

Transition from precise probabilities to imprecise probabilities in probability theory is a form of generalization and as such it enhances the ability of probability theory to deal with real-world problems. The question is: Is this mode of generalization sufficient? Is there a need for additional modes of generalization? In what follows, I argue that the answers to these questions are, respectively, No and Yes. In essence, my thesis is that what is needed is a move from imprecise probabilities to perception-based probability theory—a theory in which perceptions and their descriptions in a natural language play a pivotal role.

The perception-based theory of probabilistic reasoning which is outlined in the following is not in the traditional spirit. Its principal aim is to lay the groundwork for a radical enlargement in the role of natural languages in probability theory and its applications, especially in the realm of decision analysis.

For convenience, let PT denote standard probability theory of the kind found in textbooks and taught in courses. What is not in dispute is that standard probability theory provides a vast array of concepts and techniques which are highly effective in dealing with a wide variety of problems in which the available information is lacking in certainty. But alongside such problems we see many very simple problems for which PT offers no solutions. Here are a few typical examples:

1. What is the probability that my tax return will be audited?
2. What is the probability that my car may be stolen?
3. How long does it take to get from the hotel to the airport by taxi?
4. Usually Robert returns from work at about 6 p.m. What is the probability that he is home at 6:30 p.m.?
5. A box contains about 20 balls of various sizes. A few are small and several are large. What is the probability that a ball drawn at random is neither large nor small?

Another class of simple problems which PT cannot handle relates to commonsense reasoning (Kuipers, 1994; Fikes and Nilsson, 1971; Smithson, 1989; Shen and Leitch, 1992; Novak et al., 1992; Krause and Clark, 1993) exemplified by

6. Most young men are healthy; Robert is young. What can be said about Robert's health?

7. Most young men are healthy; it is likely that Robert is young. What can be said about Robert's health?
8. Slimness is attractive; Cindy is slim. What can be said about Cindy's attractiveness?

Questions of this kind are routinely faced and answered by humans. The answers, however, are not numbers; they are linguistic descriptions of fuzzy perceptions of probabilities, e.g., not very high, quite unlikely, about 0.8, etc. Such answers cannot be arrived at through the use of standard probability theory. This assertion may appear to be in contradiction with the existence of a voluminous literature on imprecise probabilities (Walley, 1991). In may view, this is not the case.

What are the sources of difficulty in using PT? In Problems 1 and 2, the difficulty is rooted in the basic property of conditional probabilities, namely, given $P(X)$, all that can be said about $P(X|Y)$ is that its value is between 0 and 1, assuming that Y is not contained in X or its complement. Thus, if I start with the knowledge that 1% of tax returns are audited, it tells me nothing about the probability that my tax return will be audited. The same holds true when I add more detailed information about myself, e.g., my profession, income, age, place of residence, etc. The Internal Revenue Service may be able to tell me what fraction of returns in a particular category are audited, but all that can be said about the probability that my return will be audited is that it is between 0 and 1. The tax-return-audit example raises some non-trivial issues which are analyzed in depth in a paper by Nguyen et al. (1999).

A closely related problem which does not involve probabilities is the following.

Consider a function, $y = f(x)$, defined on an interval, say $[0, 10]$, which takes values in the interval $[0, 1]$. Suppose that I am given the average value, a, of f over $[0, 10]$, and am asked: What is the value of f at $x = 3$? Clearly, all I can say is that the value is between 0 and 1.

Next, assume that I am given the average value of f over the interval $[2, 4]$, and am asked the same question. Again, all I can say is that the value is between 0 and 1. As the length of the interval decreases, the answer remains the same so long as the interval contains the point $x = 3$ and its length is not zero. As in the previous example, additional information does not improve my ability to estimate $f(3)$.

The reason why this conclusion appears to be somewhat counterintuitive is that usually there is a tacit assumption that f is a smooth function. In this case, in the limit the average value will converge to $f(3)$. Note that the answer depends on the way in which smoothness is defined.

In Problem 3, the difficulty is that we are dealing with a time series drawn from a nonstationary process. When I pose the question to a hotel clerk, he/she may tell me that it would take approximately 20–25 min. In giving this answer, the clerk may take into consideration that it is raining lightly and that as a result it would take a little longer than usual to get to the airport. PT does not have the capability to operate on the perception-based information that "it is raining lightly" and factor-in its effect on the time of travel to the airport.

In problems 4–8, the difficulty is more fundamental. Specifically, the problem is that PT—as stated above—has no capability to operate on perceptions described in a natural language, e.g., “usually Robert returns from work at about 6 p.m.”, or “the box contains several large balls” or “most young men are healthy”. This is a basic shortcoming that will be discussed in greater detail at a later point.

What we see is that standard probability theory has many strengths and many limitations. The limitations of standard probability theory fall into several categories. To see them in a broad perspective, what has to be considered is that a basic concept which is immanent in human cognition is that of partiality. Thus, we accept the reality of partial certainty, partial truth, partial precision, partial possibility, partial knowledge, partial understanding, partial belief, partial solution and partial capability, whatever it may be. Viewed through the prism of partiality, probability theory is, in essence, a theory of partial certainty and random behavior. What it does not address—at least not explicitly—is partial truth, partial precision and partial possibility—facets which are distinct from partial certainty and fall within the province of fuzzy logic (FL) (Zadeh, 1978; Dubois and Prade, 1988; Novak, 1991; Klir and Folger, 1988; Reghis and Roventa, 1998; Klir and Yuan, 1995; Grabisch et al., 1995). This observation explains why PT and FL are, for the most part, complementary rather than competitive (Zadeh, 1995; Krause and Clark, 1993; Thomas, 1995).

A simple example will illustrate the point. Suppose that Robert is three-quarters German and one-quarter French. If he were characterized as German, the characterization would be imprecise but not uncertain. Equivalently, if Robert stated that he is German, his statement would be partially true; more specifically, its truth value would be 0.75. Again, 0.75 has no relation to probability.

Within probability theory, the basic concepts on which PT rests do not reflect the reality of partiality because probability theory is based on two-valued Aristotelian logic. Thus, in PT, a process is random or not random; a time series is stationary or not stationary; an event happens or does not happen; events A and B are either independent or not independent; and so on. The denial of partiality of truth and possibility has the effect of seriously restricting the ability of probability theory to deal with those problems in which truth and possibility are matters of degree.

A case in point is the concept of an event. A recent Associated Press article carried the headline, “Balding on Top Tied to Heart Problems; Risk of disease is 36 percent higher, a study finds”. Now it is evident that both “balding on top”, and “heart problems”, are matters of degree or, more concretely, are fuzzy events, as defined in Zadeh (1968), Kruse and Meyer (1987) and Wang and Klir (1992). Such events are the norm rather than exception in real-world settings. And yet, in PT the basic concept of conditional probability of an event B given an event A is not defined when A and B are fuzzy events.

Another basic, and perhaps more serious, limitation is rooted in the fact that, in general, our assessment of probabilities is based on information which is a mixture of

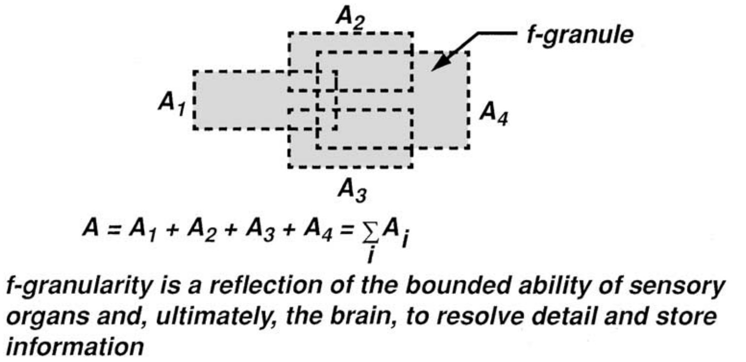

Fig. 1. f-Granularity (fuzzy granularity).

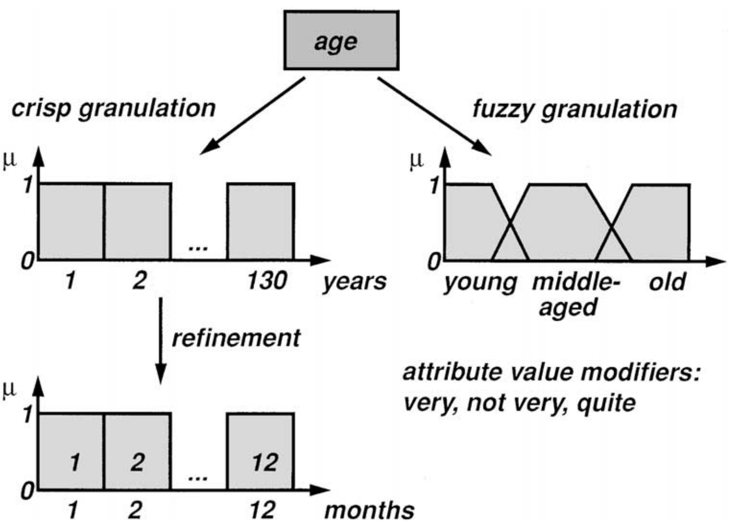

Fig. 2. Crisp and fuzzy granulation of *Age*.

measurements and perceptions (Vallee, 1995; Barsalou, 1999). Reflecting the bounded human ability to resolve detail and store information, perceptions are intrinsically imprecise. More specifically, perceptions are f-granular (Zadeh, 1979, 1997), that is: (a) perceptions are fuzzy in the sense that perceived values of variables are not sharply defined and (b) perceptions are granular in the sense that perceived values of variables are grouped into granules, with a granule being a clump of points drawn together by indistinguishability, similarity, proximity or functionality (Fig. 1). For example, the fuzzy granules of the variable Age might be young, middle-aged and old (Fig. 2). Similarly, the fuzzy granules of the variable Probability might be likely, not likely, very unlikely, very likely, etc.

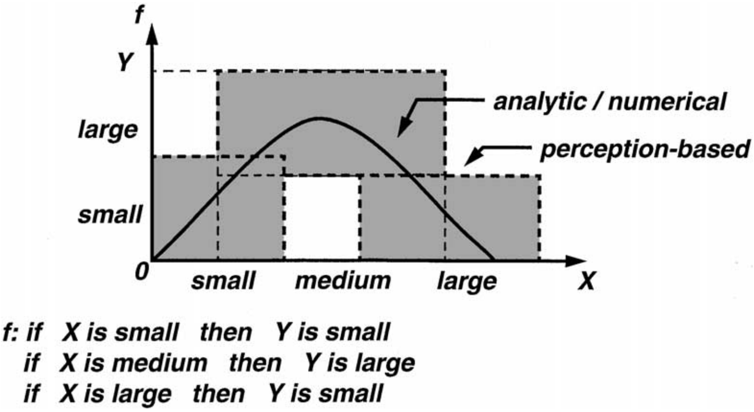

Fig. 3. Coarse description of a function by a collection of linguistic rules. Linguistic representation is perception-based.

Perceptions are described by propositions expressed in a natural language. For example

- Dana is young,
- it is a warm day,
- it is likely to rain in the evening,
- the economy is improving,
- a box contains several large balls, most of which are black.

An important class of perceptions relates to mathematical constructs such as functions, relations and counts. For example, a function such as shown in Fig. 3 may be described in words by a collection of linguistic rules (Zadeh, 1973, 1975, 1996). In particular, a probability distribution, e.g., discrete-valued probability distribution of Carol's age, P^*, may be described in words as

Prob{Carol is *young*} is *low*,
Prob{Carol is *middle-aged*} is *high*,
Prob{Carol is *old*} is *low*

or as a linguistic rule-set

if *Age* is *young* then P^* is *low*,
if *Age* is *middle-aged* then P^* is *high*,
if *Age* is *old* then P^* is *low*.

For the latter representation, using the concept of a fuzzy graph (Zadeh, 1996, 1997), which will be discussed later, the probability distribution of Carol's age may be represented as a fuzzy graph and written as

$$P^* = young \times low + middle\text{-}aged \times high + old \times low$$

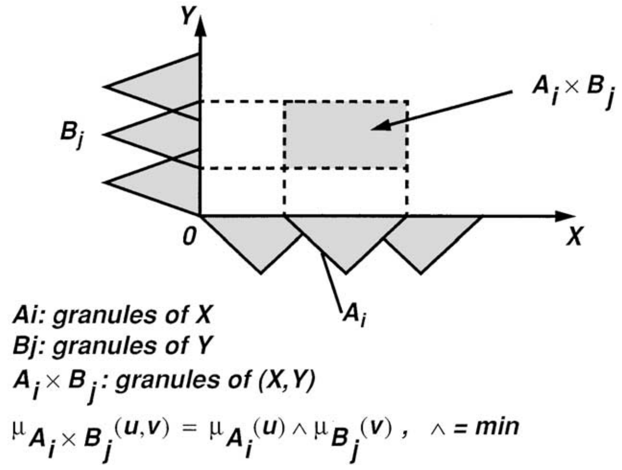

Fig. 4. Cartesian granulation. Granulation of X and Y induces granulation of (X, Y).

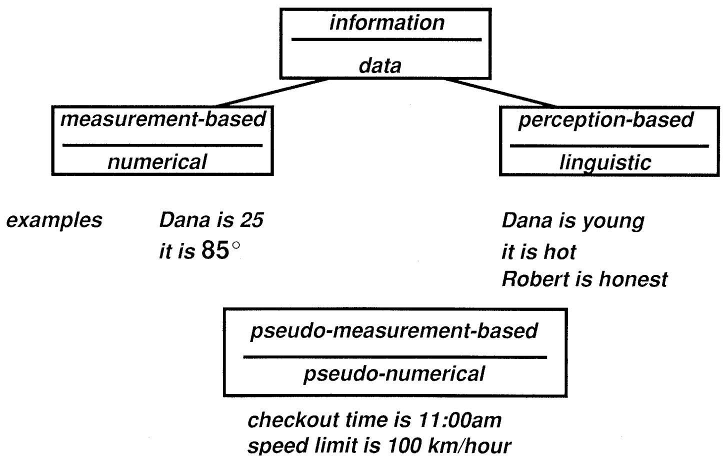

Fig. 5. Structure of information: measurement-based, perception-based and pseudo-measurement-based information.

which, as shown in Fig. 4, should be interpreted as a disjunction of cartesian products of linguistic values of *Age* and *Probability* (Zadeh, 1997; Pedrycz and Gomide, 1998).

An important observation is in order. If I were asked to estimate Carol's age, it would be unrealistic to expect that I would come up with a numerical probability distribution. But I would be able to describe my perception of the probability distribution of Carol's age in a natural language in which *Age* and *Probability* are represented—as described above—as linguistic, that is, granular variables (Zadeh, 1973, 1975, 1996, 1997).

Information which is conveyed by propositions drawn from a natural language will be said to be perception-based (Fig. 5). In my view, the most important

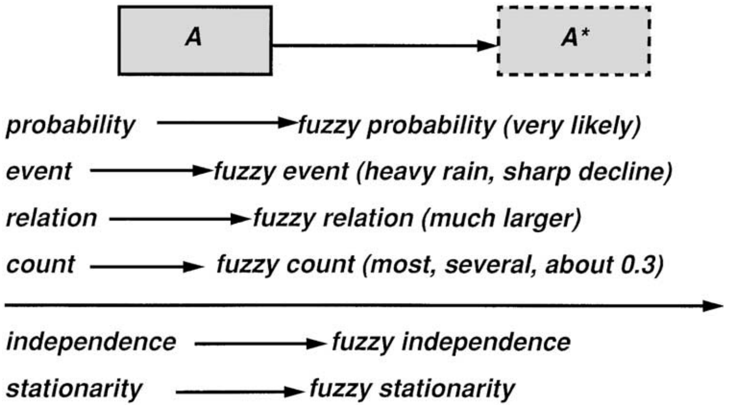

Fig. 6. f-Generalization (fuzzification). Fuzzification is a mode of generalization from crisp concepts to fuzzy concepts.

shortcoming of standard probability theory is that it does not have the capability to process perception-based information. It does not have this capability principally because there is no mechanism in PT for (a) representing the meaning of perceptions and (b) computing and reasoning with representations of meaning.

To add this capability to standard probability theory, three stages of generalization are required.

The first stage is referred to as f-generalization (Zadeh, 1997). In this mode of generalization, a point or a set is replaced by a fuzzy set. f-generalization of standard probability theory, PT, leads to a generalized probability theory which will be denoted as PT+. In relation to PT, PT+ has the capability to deal with

1. fuzzy numbers, quantifiers and probabilities, e.g., about 0.7, most, not very likely,
2. fuzzy events, e.g., warm day,
3. fuzzy relations, e.g., much larger than,
4. fuzzy truths and fuzzy possibilities, e.g., very true, quite possible.

In addition, PT+ has the potential—as yet largely unrealized—to fuzzify such basic concepts as independence, stationarity and causality. A move in this direction would be a significant paradigm shift in probability theory.

The second stage is referred to as f.g-generalization (fuzzy granulation) (Zadeh, 1997). In this mode of generalization, a point or a set is replaced by a granulated fuzzy set (Fig. 6). For example, a function, f, is replaced by its fuzzy graph, f^* (Fig. 7). f.g-generalization of PT leads to a generalized probability theory denoted as PT++.

PT++ adds to PT+ further capabilities which derive from the use of granulation. They are, mainly

1. linguistic (granular) variables,
2. linguistic (granular) functions and relations,

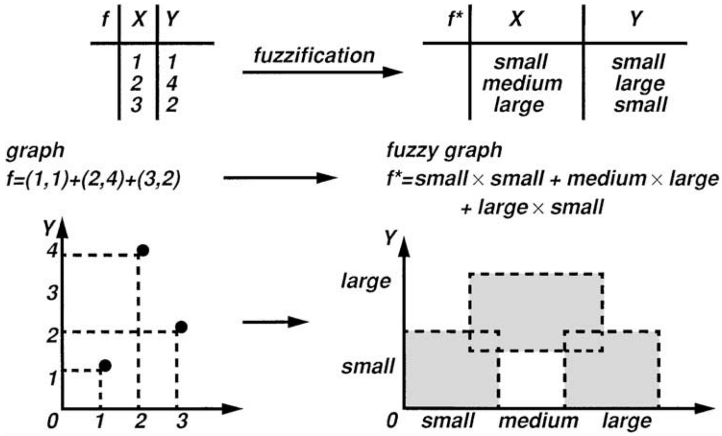

Fig. 7. Fuzzy graph of a function. A fuzzy graph is a generalization of the concept of a graph of a function.

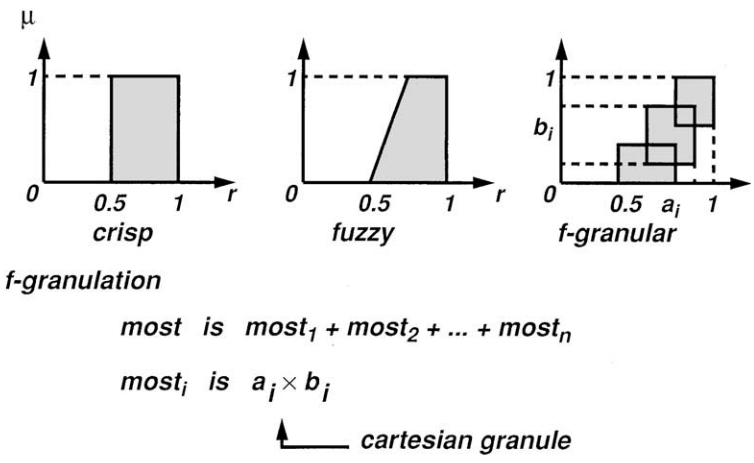

Fig. 8. Representation of *most*. Crisp, fuzzy and f-granular.

3. fuzzy rule-sets and fuzzy graphs,
4. granular goals and constraints,
5. granular probability distributions.

As a simple example, representation of the membership function of the fuzzy quantifier *most* (Zadeh, 1983) in PT, PT+ and PT++ is shown in Fig. 8.

The third stage is referred to a p-generalization (perceptualization). In this mode of generalization, what is added to PT++ is the capability to process perception-based information through the use of the computational theory of perceptions (CTP) (Zadeh, 1999, 2000). p-generalization of PT leads to what will be referred to as perception-based probability theory (PT_P).

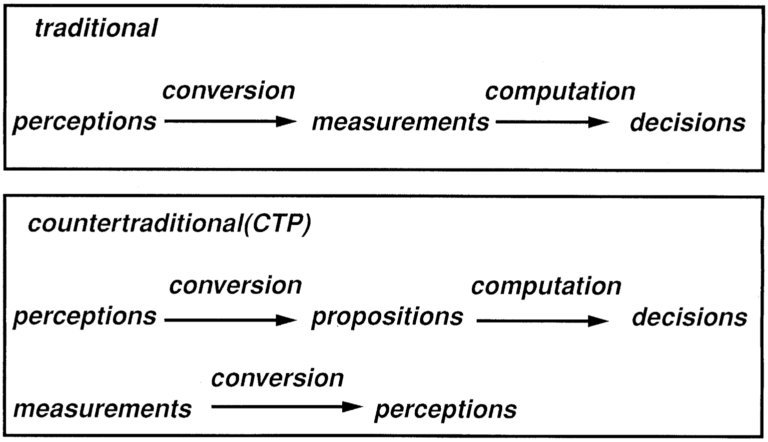

Fig. 9. Countertraditional conversion of measurements into perceptions. Traditionally, perceptions are converted into measurements.

The capability of PT_P to process perception-based information has an important implication. Specifically, it opens the door to a major enlargement of the role of natural languages in probability theory. As a simple illustration, instead of describing a probability distribution, P, analytically or numerically, as we normally do, P could be interpreted as a perception and described as a collection of propositions expressed in a natural language. A special case of such description is the widely used technique of describing a function via a collection of linguistic if–then rules (Zadeh, 1996). For example, the function shown in Fig. 7 may be described coarsely by the rule-set

f: if X is *small* then Y is *small*,
if X is *medium* then Y is *large*,
if X is *large* then Y is *small*,

with the understanding that the coarseness of granulation is a matter of choice.

In probability theory, as in other fields of science, it is a long-standing tradition to deal with perceptions by converting them into measurements. PT_p does not put this tradition aside. Rather, it adds to PT a countertraditional capability to convert measurements into perceptions, or to deal with perceptions directly, when conversion of perceptions into measurements is infeasible, unrealistic or counterproductive (Fig. 9).

There are three important points that are in need of clarification. First, when we allude to an enlarged role for natural languages in probability theory, what we have in mind is not a commonly used natural language but a subset which will be referred to as a precisiated natural language (PNL). In essence, PNL is a descriptive language which is intended to serve as a basis for representing the meaning of perceptions in a way that lends itself to computation. As will be seen later, PNL is a subset of a natural language which is equipped with constraint-centered semantics and is translatable into what is referred to as the generalized constraint language (GCL). At this point, it will suffice to observe that the descriptive power of PNL is much higher than that of the subset of a natural language which is translatable into predicate logic.

The second point is that in moving from measurements to perceptions, we move in the direction of lesser precision. The underlying rationale for this move is that precision carries a cost and that, in general, in any given situation there is a tolerance for imprecision that can be exploited to achieve tractability, robustness, lower cost and better rapport with reality.

The third point is that perceptions are more general than measurements and PT_p is more general that PT. Reflecting its greater generality, PT_p has a more complex mathematical structure than PT and is computationally more intensive. Thus, to exploit the capabilities of PT, it is necessary to have the capability to perform large volumes of computation at a low level of precision.

Perception-based probability theory goes far beyond standard probability theory both in spirit and in content. Full development of PT_p will be a long and tortuous process. In this perspective, my paper should be viewed as a sign pointing in a direction that departs from the deep-seated tradition of according more respect to numbers than to words.

Basically, perception-based probability theory may be regarded as the sum of standard probability theory and the computational theory of perceptions. The principal components of the computational theory of perceptions are (a) meaning representation and (b) reasoning. These components of CTP are discussed in the following sections.

2. The basics of perception-based probability theory; the concept of a generalized constraint

As was stated already, perception-based probability theory may be viewed as a p-generalization of standard probability theory. In the main, this generalization adds to PT the capability to operate on perception-based information through the use of the computational theory of perceptions. What follows is an informal precis of some of the basic concepts which underlie this theory.

To be able to compute and reason with perceptions, it is necessary to have a means of representing their meaning in a form that lends itself to computation. In CTP, this is done through the use of what is called constraint-centered semantics of natural languages (CSNL) (Zadeh, 1999).

A concept which plays a key role in CSNL is that of a generalized constraint (Zadeh, 1986). Introduction of this concept is motivated by the fact that conventional crisp constraints of the form $X \in C$, where X is a variable and C is a set, are insufficient to represent the meaning of perceptions.

A generalized constraint is, in effect, a family of constraints. An unconditional constraint on a variable X is represented as

$$X \text{ is} r\ R, \tag{2.1}$$

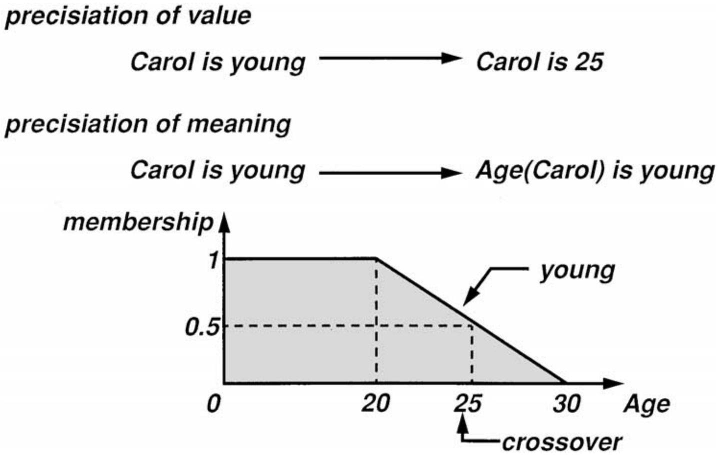

Fig. 10. Membership function of *young* (context-dependent). Two modes of precisiation.

where R is the constraining relation and is$r$, pronounced as ezar, is a variable copula in which the discrete-valued variable r defines the way in which R constrains X.

The principal constraints are the following:

$r :=$	equality constraint; $X = R$
r : blank	possibilistic constraint; X is R; R is the possibility distribution of X (Zadeh, 1978; Dubois and Prade, 1988)
$r : v$	veristic constraint; X is$v$ R; R is the verity distribution of X (Zadeh, 1999)
$r : p$	probabilistic constraint; X is$p$ R; R is the probability distribution of X
$r : pv$	probability-value constraint; X ispv R; X is the probability of a fuzzy event (Zadeh, 1968) and R is its value
$r : rs$	random set constraint; X isrs R; R is the fuzzy-set-valued probability distribution of X
$r : fg$	fuzzy graph constraint; X isfg R; X is a function and R is its fuzzy graph
$r : u$	usuality constraint; X is$u$ R; means: usually (X is R).

As an illustration, the constraint

Carol is *young*

in which *young* is a fuzzy set with a membership function such as shown in Fig. 10, is a possibilistic constraint on the variable X: $Age(Carol)$. This constraint defines the possibility distribution of X through the relation

$$\text{Poss}\{X = u\} = \mu_{young}(u),$$

where u is a numerical value of Age; μ_{young} is the membership function of *young*; and $\text{Poss}\{X = u\}$ is the possibility that Carol's age is u.

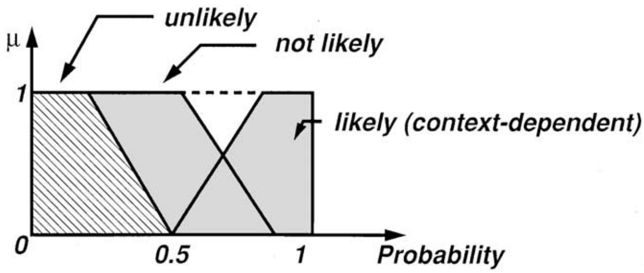

Fig. 11. Membership function of *likely* (context-dependent).

The veristic constraint

$$X \text{ is}v\ R \tag{2.2}$$

means that the verity (truth value) of the proposition $\{X = u\}$ is equal to the value of the verity distribution R at u. For example, in the proposition "Alan is half German, quarter French and quarter Italian", the verity of the proposition "Alan is German" is 0.5. It should be noted that the numbers 0.5 and 0.25 are not probabilities.

The probabilistic constraint

$$X \text{ is}p\ N(m, \sigma^2) \tag{2.3}$$

means that X is a normally distributed random variable with mean m and variance σ^2.

The proposition

$$p: \quad \text{it is likely that Carol is young} \tag{2.4}$$

may be expressed as the probability-value constraint

$$\text{Prob}\{Age(Carol) \text{ is } young\} \text{ is } likely. \tag{2.5}$$

In this expression, the constrained variable is X: Prob$\{Age(Carol)$ is $young\}$ and the constraint

$$X \text{ is } likely \tag{2.6}$$

is a possibilistic constraint in which *likely* is a fuzzy probability whose membership function is shown in Fig. 11.

In the random-set constraint, X is a fuzzy-set-valued random variable. Assuming that the values of X are fuzzy sets $\{A_i, i = 1, \dots, n\}$ with respective probabilities $p_1, \dots, p_n$, the random-set constraint on X is expressed symbolically as

$$X \text{ is}rs\ (p_1 \backslash A_1 + \cdots + p_n \backslash A_n). \tag{2.7}$$

It should be noted that a random-set constraint may be viewed as a combination of (a) a probabilistic constraint, expressed as

$$X \text{ is}p\ (p_1 \backslash u_1 + \cdots + p_n \backslash u_n), \quad u_i \in U \tag{2.8}$$

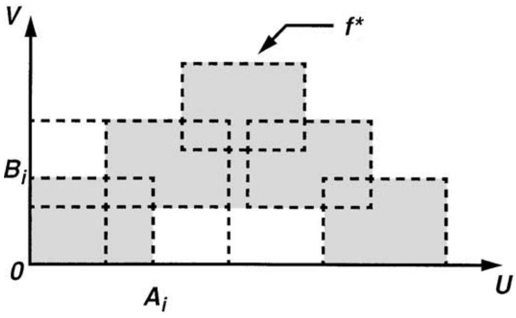

Fig. 12. Fuzzy-graph constraint. f^* is a fuzzy graph which is an approximate representation of f.

and a possibilistic constraint expressed as

$$(X, Y) \text{ is } R, \tag{2.9}$$

where R is a fuzzy relation defined on $U \times V$, with membership function $\mu_R : U \times V \rightarrow [0,1]$.

If A_i is a section of R, defined as in Zadeh (1997) by

$$\mu_{A_i}(v) = \mu_R(u_i, v), \tag{2.10}$$

then the constraint on Y is a random-set constraint expressed as

$$Y \text{ is} rs \ (p_1 \backslash A_1 + \cdots + p_n \backslash A_n). \tag{2.11}$$

Another point that should be noted is that the concept of a random-set constraint is closely related to the Dempster–Shafer theory of evidence (Dempster, 1967; Shafer, 1976) in which the focal sets are allowed to be fuzzy sets (Zadeh, 1979).

In the fuzzy-graph constraint

$$X \text{ is} fg \ R, \tag{2.12}$$

the constrained variable, X, is a function, f, and R is a fuzzy graph (Zadeh, 1997) which plays the role of a possibility distribution of X. More specifically, if $f : U \times V \rightarrow [0,1]$ and A_i, $i = 1, \ldots, m$ and B_j, $j = 1, \ldots, n$, are, respectively, fuzzy granules in U and V (Fig. 12), then the fuzzy graph of f is the disjunction of cartesian products (granules) $U_i \times V_j$, expressed as

$$f^* = \sum_{i=1, j=1}^{m,n} U_i \times V_j, \tag{2.13}$$

with the understanding that the symbol $\sum$ should be interpreted as the union rather than as an arithmetic sum, and U_i and V_j take values in the sets $\{A_1, \ldots, A_m\}$ and $\{B_1, \ldots, B_n\}$, respectively.

A fuzzy graph of f may be viewed as an approximate representation of f. Usually, the granules A_i and B_j play the role of values of linguistic variables. Thus, in the case of the function shown in Fig. 7, its fuzzy graph may be expressed as

$$f^* = small \times small + medium \times large + large \times small. \tag{2.14}$$

Equivalently, if f is written as $Y = f(X)$, then f^* may be expressed as the rule-set

$$f^*: \quad \text{if } X \text{ is } small \text{ then } Y \text{ is } small, \\ \text{if } X \text{ is } medium \text{ then } Y \text{ is } large, \\ \text{if } X \text{ is } large \text{ then } Y \text{ is } small. \tag{2.15}$$

This rule-set may be interpreted as a description—in a natural language—of a perception of f.

The usuality constraint is a special case of the probability-value constraint. Thus,

$$X \text{ is} u\ A \tag{2.16}$$

should be interpreted as an abbreviation of

$$\text{usually } (X \text{ is } A), \tag{2.17}$$

which in turn may be interpreted as

$$\text{Prob}\{X \ is\ A\} \text{ is } usually, \tag{2.18}$$

with *usually* playing the role of a fuzzy probability which is close to 1. In this sense, A is a *usual* value of X. More generally, A is a *usual* value of X if the fuzzy probability of the fuzzy event $\{X \text{ is } A\}$ is close to one and A has high specificity, that is, has a tight possibility distribution, with tightness being a context-dependent characteristic of a fuzzy set. It is important to note that, unlike the concept of the expected value, the usual value of a random variable is not uniquely determined by its probability distribution. What this means is that the usual value depends on the calibration of the context-dependent natural language predicates "close to one" and "high specificity".

The difference between the concepts of the expected and usual values goes to the heart of the difference between precise and imprecise probability theories. The expected value is precisely defined and unique. The usual value is context-dependent and hence is not unique. However, its definition is precise if the natural language predicates which occur in its definition are defined precisely by their membership functions. In this sense, the concept of the usual value has a flexibility that the expected value does not have. Furthermore, it may be argued that the concept of the usual value is closer to our intuitive perception of "expected value" than the concept of the expected value as it is defined in PT.

In the foregoing discussion, we have focused our attention on unconditional generalized constraints. More generally, a generalized constraint may be conditional, in which case it is expressed in a generic form as an if–then rule

$$\text{if } X \text{ is} r\ R \text{ then } Y \text{ is} s\ S \tag{2.19}$$

or, equivalently, as

$$Y \text{ is} s\ S \text{ if } X \text{ is} r\ R. \tag{2.20}$$

Furthermore, a generalized constraint may be exception-qualified, in which case it is expressed as

$$X \text{ is} r\ R \text{ unless } Y \text{ is} s\ S. \tag{2.21}$$

A generalized rule-set is a collection of generalized if–then rules which collectively serve as an approximate representation of a function or a relation. Equivalently, a generalized rule-set may be viewed as a description of a perception of a function or a relation.

As an illustration, consider a function, $f:(U\times V)\to[0,1]$, expressed as $Y=f(X)$, where U and V are the domains of X and Y, respectively. Assume that U and V are granulated, with the granules of U and V denoted, respectively, as A_i, $i=1,\dots,m$, and B_j, $j=1,\dots,n$. Then, a generic form of a generalized rule set may be expressed as

$$f^*:\quad \{\text{if } X \text{ is} r\ U_i \text{ then } Y \text{ is} s\ V_j\} \quad i=1,\dots,m,\ j=1,\dots,n, \tag{2.22}$$

where U_i and V_j take values in the sets $\{A_1,\dots,A_m\}$ and $\{B_1,\dots,B_n\}$, respectively. In this expression, f^* represents a fuzzy graph of f.

A concept which plays a key role in the computational theory of perceptions is that of the Generalized Constraint Language, GCL (Zadeh, 1999). Informally, GCL is a meaning-representation language in which the principal semantic elements are generalized constraints. The use of generalized constraints as its semantic elements makes a GCL a far more expressive language than conventional meaning-representation languages based on predicate logic.

3. Meaning-representation: constraint-centered semantics of natural languages

In perception-based probability theory, perceptions—and, in particular, perceptions of likelihood, dependency, count and variations in time and space—are described by propositions drawn from a natural language. To mechanize reasoning with perceptions, it is necessary to have a method of representing the meaning of propositions in a way that lends itself to computation. In the computational theory of perceptions, a system that is used for this purpose is called the constraint-centered semantics of natural language (CSNL) (Zadeh, 1999).

Meaning-representation is a central part of every logical system. Why, then, is it necessary to introduce a system that is significantly different from the many meaning-representation methods that are in use? The reason has to do with the intrinsic imprecision of perceptions and, more particularly, with their f-granularity. It is this characteristic of perceptions that puts them well beyond the expressive power of conventional meaning-representation methods, most of which are based on predicate logic.

To illustrate, consider the following simple perceptions:

- Ann is much younger than Mary.
- A box contains black and white balls of various sizes. Most are large. Most of the large balls are black.
- Usually it is rather cold in San Francisco during the summer.
- It is very unlikely that there will be a significant increase in the price of oil in the near future.

Conventional meaning-representation methods do not have the capability to represent the meaning of such perceptions in a form that lends itself to computation.

A key idea which differentiates CSNL from conventional methods is that the meaning of a proposition, p, drawn from a natural language, is represented as a generalized constraint, with the understanding that the constrained variable and the constraining relation are, in general, implicit rather than explicit in p. For example, in the proposition

p: it is likely that Kate is young,

the constraint is possibilistic; the constrained variable is the probability that Kate is young; and the constraining relation is *likely*.

The principal ideas and assumptions which underlie CSNL may be summarized as follows:

1. Perceptions are described by propositions drawn from a natural language.
2. A proposition, p, may be viewed as an answer to a question.
   In general, the question is implicit and not unique. For example, the proposition "Carol is young" may be viewed as an answer to the question: "How old is Carol", or as the answer to "Who is young?"
3. A proposition is a carrier of information.
4. The meaning of a proposition, p, is represented as a generalized constraint which defines the information conveyed by p.
5. Meaning-representation is viewed as translation from a language into the GCL.

In CSNL, translation of a proposition, p, into GCL is equated to explicitation of the generalized constraint which represents the meaning of p. In symbols

$$p \ \frac{\text{translation}}{\text{explicitation}} \ X \text{ is} r \ R. \tag{3.1}$$

The right-hand member of this relation is referred to as a canonical form of p, written as $CF(p)$. Thus, the canonical form of p places in evidence (a) the constrained variable which, in general, is implicit in p; (b) the constraining relation, R; and (c) the copula variable r which defines the way in which R constrains X.

The canonical form of a question, q, may be expressed as

$$CF(q): \quad X \text{ is}r \ ?R \tag{3.2}$$

and read as "What is the generalized value of X?"

Similarly, the canonical form of p, viewed as an answer to q, is expressed as

$$CF(p): \quad X \text{ is}r \ R \tag{3.3}$$

and reads "The generalized value of X is$r$ R".

As a simple illustration, if the question is "How old is Carol?", its canonical form is

$$CF(q): \quad Age(Carol) \text{ is } ?R. \tag{3.4}$$

Correspondingly, the canonical form of

$$p: \quad \text{Carol is young} \tag{3.5}$$

is

$$CF(p): \quad Age(Carol) \text{ is } young. \tag{3.6}$$

If the answer to the question is

$$p: \quad \text{it is likely that Carol is young} \tag{3.7}$$

then

$$CF(p): \quad \text{Prob}\{Age(Carol) \text{ is } young\} \text{ is } likely. \tag{3.8}$$

More explicitly, if $Age(Carol)$ is a random variable with probability density g, then the probability measure (Zadeh, 1968) of the fuzzy event "Carol is young" may be expressed as

$$\int_0^{120} \mu_{young}(u) g(u)\, \mathrm{d}u, \tag{3.9}$$

where μ_{young} is the membership function of *young*. Thus, in this interpretation the constrained variable is the probability density g, and, as will be seen later, the membership function of the constraining relation is given by

$$\mu_R(g) = \mu_{likely}\left(\int_0^{120} \mu_{young}(u) g(u)\, \mathrm{d}u\right). \tag{3.10}$$

A concept which plays an important role in CSNL is that of cardinality, that is, the count of elements in a fuzzy set (Zadeh, 1983; Ralescu, 1995; Hajek, 1998). Basically, there are two ways in which cardinality can be defined: (a) crisp cardinality and (b) fuzzy cardinality (Zadeh, 1983; Ralescu et al., 1995; Ralescu, 1995). In the case of (a), the count of elements in a fuzzy set is a crisp number; in the case of (b) it is a fuzzy number. For our purposes, it will suffice to restrict our attention to the case where a fuzzy set is defined on a finite set and is associated with a crisp count of its elements.

More specifically, consider a fuzzy set A defined on a finite set $U = \{u_1, \ldots, u_n\}$ through its membership function $\mu_A : U \to [0,1]$. The sigma-count of A is defined as

$$\sum Count(A) = \sum_{i=1}^{n} \mu_A(u_i). \tag{3.11}$$

If A and B are fuzzy sets defined on U, then the relative sigma-count, $\sum Count(A/B)$, is defined as

$$\sum Count(A/B) = \frac{\sum_{i=1}^{n} \mu_A(u_i) \wedge \mu_B(u_i)}{\sum_{i=1}^{n} \mu_B(u_i)}, \tag{3.12}$$

where $\wedge = \min$, and summations are arithmetic.

As a simple illustration, consider the perception

p: most Swedes are tall.

In this case, the canonical form of p may be expressed as

$$CF(p): \quad \sum Count(tall\,.\,Swedes/Swedes) \text{ is } \frac{1}{n}\sum_{i=1}^{n} \mu_{tall\,.\,Swede}(u_i), \tag{3.13}$$

where u_i is the height of the ith Swede and $\mu_{tall\,.\,Swede}(u_i)$ is the grade of membership of the ith Swede in the fuzzy set of tall Swedes.

In a general setting, how can a given proposition, p, be expressed in its canonical form? A framework for translation of propositions drawn from a natural language into GCL is partially provided by the conceptual structure of test-score semantics (Zadeh, 1981). In this semantics, X and R are defined by procedures which act on an explanatory database, ED, with ED playing the role of a collection of possible worlds in possible world semantics (Cresswell, 1973). As a very simple illustration, consider the proposition (Zadeh, 1999)

p: Carol lives in a small city near San Francisco

and assume that the explanatory database consists of three relations:

$$\begin{aligned} ED = {} & POPULATION[Name; Residence] \\ & + SMALL[City; \mu] \\ & + NEAR[City1; City2; \mu]. \end{aligned} \tag{3.14}$$

In this case,

$$X = Residence(Carol) =_{Residence} POPULATION[Name = Carol], \tag{3.15}$$

$$R = SMALL[City; \mu] \cap_{City1} NEAR[City2 = San\_Francisco]. \tag{3.16}$$

In R, the first constituent is the fuzzy set of small cities; the second constituent is the fuzzy set of cities which are near San Francisco; and $\cap$ denotes the intersection of these sets. Left subscripts denote projections, as defined in Zadeh (1981).

There are many issues relating to meaning-representation of perception-based information which go beyond the scope of the present paper. The brief outline presented in this section is sufficient for our purposes. In the following section, our attention will be focused on the basic problem of reasoning based on generalized constraint propagation. The method which will be outlined contains as a special case a basic idea suggested in an early paper of Good (1962). A related idea was employed in Zadeh (1955).

4. Reasoning based on propagation of generalized constraints

One of the basic problems in probability theory is that of computation of the probability of a given event from a body of knowledge which consists of information about the relevant functions, relations, counts, dependencies and probabilities of related events.

As was alluded to earlier, in many cases the available information is a mixture of measurements and perceptions. Standard probability theory provides a vast array of tools for dealing with measurement-based information. But what is not provided is a machinery for dealing with information which is perception-based. This limitation of PT is exemplified by the following elementary problems—problems in which information is perception-based.

1. X is a normally distributed random variable with small mean and small variance.
   Y is much larger than X.
   What is the probability that Y is neither small nor large?
2. Most Swedes are tall.
   Most Swedes are blond.
   What is the probability that a Swede picked at random is tall and blond?
3. Consider a perception-valued times series

   $$T=\{t_1,t_2,t_3,\dots\},$$

   in which the t_i's are perceptions of, say temperature, e.g., warm, very warm, cold,… . For simplicity, assume that the t_i's are independent and identically distributed. Furthermore, assume that the t_i's range over a finite set of linguistic values, $A_1,A_2,\dots,A_n$, with respective probabilities $P_1,\dots,P_n$. What is the average value of T?

To be able to compute with perceptions, it is necessary, as was stressed already, to have a mechanism for representing their meaning in a form that lends itself to computation. In the computational theory of perceptions, this purpose is served by the constraint-centered semantics of natural languages. Through the use of CSNL, propositions drawn from a natural language are translated into the GCL.

The second stage of computation involves generalized constraint propagation from premises to conclusions. Restricted versions of constraint propagation are considered in Zadeh (1979), Bowen et al. (1992), Dubois et al. (1993), Katai et al. (1992) and Yager (1989). The main steps in generalized constraint propagation are summarized in the following. As a preliminary, a simple example is analyzed.

Assume that the premises consist of two perceptions:

p_1: most Swedes are tall,

p_2: most Swedes are blond.

and the question, q, is: What fraction of Swedes are tall and blond? This fraction, then, will be the linguistic value of the probability that a Swede picked at random is tall and blond.

To answer the question, we first convert p_1, p_2 and q into their canonical forms:

$$CF(p_1):\quad \sum Count(tall\,.\,Swedes/Swedes) \text{ is } most, \tag{4.1}$$

$$CF(p_2):\quad \sum Count(blond\,.\,Swedes/Swedes) \text{ is } most, \tag{4.2}$$

$$CF(q):\quad \sum Count(tall\cap blond\,.\,Swedes/Swedes) \text{ is } ?Q, \tag{4.3}$$

where Q is the desired fraction.

Next, we employ the identity (Zadeh, 1983)

$$\sum Count(A \cap B) + \sum Count(A \cup B) = \sum Count(A) + \sum Count(B), \tag{4.4}$$

in which A and B are arbitrary fuzzy sets. From this identity, we can readily deduce that

$$\sum Count(A) + \sum Count(B) - 1 \leqslant \sum Count(A \cap B) \\ \leqslant \min\left(\sum Count(A), \sum Count(B)\right), \tag{4.5}$$

with the understanding that the lower bound is constrained to lie in the interval $[0, 1]$. It should be noted that the identity in question is a generalization of the basic identity for probability measures

$$P(A \cap B) + P(A \cup B) = P(A) + P(B). \tag{4.6}$$

Using the information conveyed by canonical forms, we obtain the bounds

$$2most - 1 \leqslant \sum Count(tall \cap blond\,.\,Swedes/Swedes) \leqslant most, \tag{4.7}$$

which may be expressed equivalently as

$$\sum Count(tall \cap blond\,.\,Swedes/Swedes) \text{ is } \leqslant most \cap \geqslant (2most - 1). \tag{4.8}$$

Now

$$\leqslant most = [0, 1] \tag{4.9}$$

and

$$\geqslant (2most - 1) = 2most - 1, \tag{4.10}$$

in virtue of monotonicity of *most* (Zadeh, 1999).

Consequently,

$$\sum Count(tall \cap blond\,.\,Swedes/Swedes) \text{ is } 2most - 1 \tag{4.11}$$

and hence the answer to the question is

$$a\!: \quad (2most - 1)\ Swedes \text{ are } tall\ and\ blond. \tag{4.12}$$

In a more general setting, the principal elements of the reasoning process are the following.

1. Question (query), q. The canonical form of q is assumed to be

$$X \text{ is} r\ ?Q. \tag{4.13}$$

2. Premises. The collection of premises expressed in a natural language constitutes the initial data set (IDS).
3. Additional premises which are needed to arrive at an answer to q. These premises constitute the external data set (EDS). Addition of EDS to IDS results in what is referred to as the augmented data set (IDS+).

Example. Assume that the initial data set consists of the propositions

p_1: Carol lives near Berkeley,

p_2: Pat lives near Palo Alto.

Suppose that the question is: How far is Carol from Pat? The external data set in this case consists of the proposition

distance between Berkeley and Palo Alto is approximately 45 miles. (4.14)

4. Through the use of CSNL, propositions in IDS+ are translated into the GCL. The resulting collection of generalized constraints is referred to as the augmented initial constraint set ICS+.
5. With the generalized constraints in ICS+ serving as antecedent constraints, the rules which govern generalized constraint propagation in CTP are applied to ICS+, with the goal of deducing a set of generalized constraints, referred to as the terminal constraint set, which collectively provide the information which is needed to compute q.

The rules governing generalized constraint propagation in the computational theory of perceptions coincide with the rules of inference in fuzzy logic (Zadeh, 1999, 2000). In general, the chains of inference in CTP are short because of the intrinsic imprecision of perceptions. The shortness of chains of inference greatly simplifies what would otherwise be a complex problem, namely, the problem of selection of rules which should be applied in succession to arrive at the terminal constraint set. This basic problem plays a central role in theorem proving in the context of standard logical systems (Fikes and Nilsson, 1971).

6. The generalized constraints in the terminal constraint set are re-translated into a natural language, leading to the terminal data set. This set serves as the answer to the posed question. The process of re-translation is referred to as linguistic approximation (Pedrycz and Gomide, 1998). Re-translation will not be addressed in this paper.

The basic rules which govern generalized constraint propagation are of the general form

$$\begin{array}{l} p_1 \\ p_2 \\ \cdot \\ \cdot \\ \cdot \\ p_k \\ \hline p_{k+1}, \end{array} \tag{4.15}$$

where $p_1, \ldots, p_k$ are the premises and p_{k+1} is the conclusion. Generally, $k = 1$ or 2.

In a generic form, the basic constraint-propagation rules in CTP are expressed as follows (Zadeh, 1999):

1. *Conjunctive rule* 1:

$$\frac{\begin{array}{lll} X & \text{is}r & R \\ X & \text{is}s & S \end{array}}{\begin{array}{lll} X & \text{is}t & T. \end{array}} \tag{4.16}$$

The different symbols r, s, t in constraint copulas signify that the constraints need not be of the same type.

2. *Conjunctive rule* 2:

$$\frac{\begin{array}{ccc} X & \text{is}r & R \\ Y & \text{is}s & S \end{array}}{(X, Y) \quad \text{is}t \quad T.} \tag{4.17}$$

3. *Disjunctive rule* 1:

$$\text{or} \ \frac{\begin{array}{lll} X & \text{is}r & R \\ X & \text{is}s & S \end{array}}{\begin{array}{lll} X & \text{is}t & T. \end{array}} \tag{4.18}$$

4. *Disjunctive rule* 2:

$$\text{or} \ \frac{\begin{array}{ccc} X & \text{is}r & R \\ Y & \text{is}s & S \end{array}}{(X, Y) \quad \text{is}t \quad T.} \tag{4.19}$$

5. *Projective rule*:

$$\frac{(X, Y) \quad \text{is}r \quad R}{Y \quad \text{is}s \quad S.} \tag{4.20}$$

6. *Surjective rule*:

$$\frac{X \quad \text{is}r \quad R}{(X, Y) \quad \text{is}s \quad S.} \tag{4.21}$$

7. *Inversive rule*:

$$\frac{f(X) \quad \text{is}r \quad R}{X \quad \text{is}s \quad S,} \tag{4.22}$$

where $f(X)$ is a function of X.

From these basic rules the following frequently used rules may be derived:

8. *Compositional rule*:

$$\frac{\begin{array}{ccc} X & \text{is}r & R \\ (X, Y) & \text{is}s & S \end{array}}{Y \quad \text{is}t \quad T.} \tag{4.23}$$

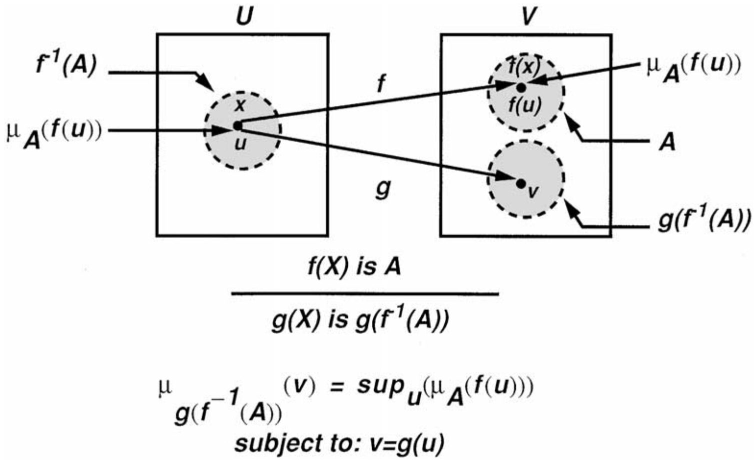

Fig. 13. Generalized extension principle. Constraint on $f(X)$ induces a constraint on $g(X)$.

9. *Generalized extension principle*:

$$\frac{f(X) \quad \text{is}r \quad R}{g(X) \quad \text{is}s \quad S,} \tag{4.24}$$

where f and g are given functions. The generalized extension principle is the principal rule of inference in fuzzy logic.

The generic rules lead to specialized rules for various types of constraints. In particular, for possibilistic constraints we have, for example (Pedrycz and Gomide, 1998)

Conjunctive rule 1:

$$\begin{array}{lll} X & \text{is} & R \\ X & \text{is} & S \\ \hline X & \text{is} & R \cap S, \end{array} \tag{4.25}$$

where R and S are fuzzy sets and $R \cap S$ is their intersection.

Compositional rule:

$$\begin{array}{lll} X & \text{is} & R \\ (X, Y) & \text{is} & S \\ \hline Y & \text{is} & R \bullet S, \end{array} \tag{4.26}$$

where $R \bullet S$ is the composition of R and S. If conjunction and disjunction are identified with min and max, respectively, then

$$\mu_{R \bullet S}(v) = \max_u(\min(\mu_R(u), \mu_S(u, v))), \tag{4.27}$$

where μ_R and μ_S are the membership functions of R and S.

Generalized extension principle (Fig. 13):

$$\frac{f(X) \quad \text{is} \quad R}{g(X) \text{ is } g(f^{-1}(R)),} \tag{4.28}$$

where

$$\mu_{g(f^{-1}(R))}(v) = \max_{u|v=g(u)} \mu_R(f(u)). \tag{4.29}$$

Compositional rule for probabilistic constraints (*Bayes' rule*):

$$\frac{\begin{array}{c} X \quad \text{is}p \quad R \\ Y|X \quad \text{is}p \quad S \end{array}}{Y \quad \text{is}p \quad R \bullet S,} \tag{4.30}$$

where $Y|X$ denotes Y conditioned on X, and $R \bullet S$ is the composition of the probability distributions R and S.

Compositional rule for probabilistic and possibilistic constraints (*random-set constraint*):

$$\frac{\begin{array}{c} X \quad \text{is}p \quad R \\ (X,Y) \quad \text{is} \quad S \end{array}}{Y \quad \text{is}rs \quad T,} \tag{4.31}$$

where T is a random set. As was stated at an earlier point, if X takes values in a finite set $\{u_1,\ldots,u_n\}$ with respective probabilities $p_1,\ldots,p_n$, then the constraint X is$p$ R may be expressed compactly as

$$X \text{ is}p \left(\sum_{i=1}^{n} p_i \backslash u_i\right). \tag{4.32}$$

When X takes a value u_i, the possibilistic constraint (X,Y) is S induces a constraint on Y which is given by

$$Y \text{ is } S_i, \tag{4.33}$$

where S_i is a fuzzy set defined by

$$S_i = S(u_i, Y). \tag{4.34}$$

From this it follows that when X takes the values $u_1,\ldots,u_n$ with respective probabilities $p_1,\ldots,p_n$, the fuzzy-set-valued probability distribution of Y may be expressed as

$$Y \text{ is}p \left(\sum_{i=1}^{n} p_i \backslash S_i\right). \tag{4.35}$$

This fuzzy-set-valued probability distribution defines the random set T in the random-set constraint

$$Y \text{ is}rs\ T. \tag{4.36}$$

Conjunctive rule for random set constraints: For the special case in which R and S in the generic conjunctive rule are random fuzzy sets as defined above, the rule assumes a more specific form:

$$\frac{\begin{array}{lll} X & \text{is}rs & \sum_{i=1}^{m} p_i \backslash R_i \\ X & \text{is}rs & \sum_{j=1}^{n} q_j \backslash S_j \end{array}}{\begin{array}{lll} X & \text{is}rs & \sum_{i=1,j=1}^{m,n} p_i q_j \backslash (R_i \cap S_j). \end{array}} \tag{4.37}$$

In this rule, R_i and S_i are assumed to be fuzzy sets. When R_i and S_i are crisp sets, the rule reduces to the Dempster rule of combination of evidence (Dempster, 1967; Shafer, 1976). An extension of Dempster's rule to fuzzy sets was described in a paper dealing with fuzzy information granularity (Zadeh, 1979). It should be noted that in (4.37) the right-hand member is not normalized, as it is in the Dempster–Shafer theory (Strat, 1992).

The few simple examples discussed above demonstrate that there are many ways in which generic rules can be specialized, with each specialization leading to a distinct theory in its own right. For example, possibilistic constraints lead to possibility theory (Zadeh, 1978; Dubois and Prade, 1988); probabilistic constraints lead to probability theory; and random-set constraints lead to the Dempster–Shafer theory of evidence. In combination, these and other specialized rules of generalized constraint propagation provide the machinery that is needed for a mechanization of reasoning processes in the logic of perceptions and, more particularly, in a perception-based theory of probabilistic reasoning with imprecise probabilities.

As an illustration, let us consider a simple problem that was stated earlier—a typical problem which arises in situations in which the decision-relevant information is perception-based. Given the perception: Usually Robert returns from work at about 6 p.m.; the question is: What is the probability that he is home at 6:30 p.m.?

An applicable constraint-propagation rule in this case is the generalized extension principle. More specifically, let g denote the probability density of the time at which Robert returns from work. The initial data set is the proposition

p: usually Robert returns from work at about 6 p.m.

This proposition may be expressed as the usuality constraint

$$X \text{ is} u\ 6^*, \tag{4.38}$$

where 6^* is an abbreviation for "about 6 p.m.", and X is the time at which Robert returns from work. Equivalently, the constraint in question may be expressed as

$$p: \quad \text{Prob}\{X \text{ is } 6^*\} \text{ is } \textit{usually}. \tag{4.39}$$

Using the definition of the probability measure of a fuzzy event (Zadeh, 1968), the constraint on g may be expressed as

$$\int_0^{12} g(u)\mu_{6^*}(u)\,\mathrm{d}u \text{ is } \textit{usually}, \tag{4.40}$$

where $\mu_{6^*}(u)$ is the membership function of 6^* (Fig. 14).

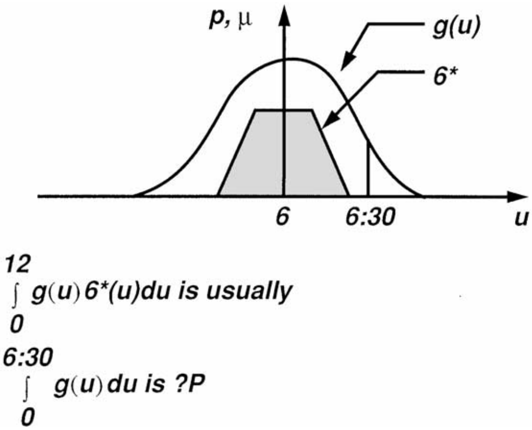

Fig. 14. Application of the generalized extension principle. P is the probability that Robert is at home at 6:30 p.m.

Let $P(g)$ denote the probability that Robert is at home at 6:30 p.m. This probability would be a number if g were known. In our case, information about g is conveyed by the given usuality constraint. This constraint defines the possibility distribution of g as a functional:

$$\mu(g) = \mu_{usually}\left(\int_0^{12} g(u)\mu_{6^*}(u)\,\mathrm{d}u\right). \tag{4.41}$$

In terms of g, the probability that Robert is home at 6:30 p.m. may be written as a functional:

$$P(g) = \int_0^{6:30} g(u)\,\mathrm{d}u. \tag{4.42}$$

The generalized extension principle reduces computation of the possibility distribution of P to the solution of the variational problem

$$\mu_P(v) = \max_g\left(\mu_{usually}\left(\int_0^{12} g(u)\mu_{6^*}(u)\,\mathrm{d}u\right)\right) \tag{4.43}$$

subject to

$$v = \int_0^{6:30} g(u)\,\mathrm{d}u.$$

The reduction of inference to solution of constrained variational problems is a basic feature of fuzzy logic (Zadeh, 1979).

Solution of variational problems of form (4.43) may be simplified by a discretization of g. Thus, if u is assumed to take values in a finite set $U = \{u_1, \ldots, u_n\}$, and the respective probabilities are $p_1, \ldots, p_n$, then the variational problem (4.43) reduces to the nonlinear program

$$\mu_P(v) = \max_P\left(\mu_{usually}\left(\sum_{i=1}^{n} p_i\mu_{6^*}(u_i)\right)\right) \tag{4.44}$$

subject to

$$v=\sum_{j=1}^{m} P_j,$$

$$0 \leqslant P_j \leqslant 1,$$

$$\sum_{i=1}^{n} P_i = 1,$$

where $p=(p_1,\ldots,p_n)$, and m is such that $u_m=6{:}30$.

In general, probabilities serve as a basis for making a rational decision. As an illustration, assume that I want to call Robert at home at 6:30 p.m. and have to decide on whether I should call him person-to-person or station-to-station. Assume that we have solved the variational problem (4.43) and have in hand the value of P defined by its membership function $\mu_P(v)$. Furthermore, assume that the costs of person-to-person and station-to-station calls are a and b, respectively.

Then the expected cost of a person-to-person call is

$$A=aP,$$

while that of a station-to-station call is

$$B=b,$$

where A is a fuzzy number defined by (Kaufmann and Gupta, 1985)

$$\mu_A(v)=a\mu_P(v).$$

More generally, if X is a random variable taking values in the set of numbers $U=\{a_1,\ldots,a_n\}$ with respective imprecise (fuzzy) probabilities $P_1,\ldots,P_n$, then the expected value of X is the fuzzy number (Zadeh, 1975; Kruse and Meyer, 1987)

$$E(X)=\sum_{i=1}^{n} a_i P_i. \tag{4.45}$$

The membership function of $E(X)$ may be computed through the use of fuzzy arithmetic (Kaufmann and Gupta, 1985; Mares, 1994). More specifically, if the membership functions of P_i are μ_i, then the membership function of $E(X)$ is given by the solution of the variational problem

$$\mu_{E(X)}(v)=\max_{u_1,\ldots,u_n}(\mu_{P_1}(u_1)\wedge\cdots\wedge\mu_{P_n}(u_n)) \tag{4.46}$$

subject to the constraints

$$0 \leqslant u_i \leqslant 1,$$

$$\sum_{i=1}^{n} u_i = 1,$$

$$v=\sum_{i=1}^{n} a_i u_i.$$

Returning to our discussion of the Robert example, if we employ a generalized version of the principle of maximization of expected utility to decide on how to place the call, then the problem reduces to that of ranking the fuzzy numbers A and B. The problem of ranking of fuzzy numbers has received considerable attention in the literature (see Pedrycz and Gomide, 1998), and a number of ranking algorithms have been described.

Our discussion of the Robert example is aimed at highlighting some of the principal facets of the perception-based approach to reasoning with imprecise probabilities. The key point is that reasoning with perception-based information may be reduced to solution of variational problems. In general, the problems are computationally intensive, even for simple examples, but well within the capabilities of desktop computers. Eventually, novel methods of computation involving neural computing, evolutionary computing, molecular computing or quantum computing may turn out to be effective in computing with imprecise probabilities in the context of perception-based information.

As a further illustration of reasoning with perception-based information, it is instructive to consider a perception-based version of a basic problem in probability theory.

Let X and Y be random variables in U and V, respectively. Let f be a mapping from U to V. The basic problem is: Given the probability distribution of $X, P(X)$, what is the probability distribution of Y?

In the perception-based version of this problem it is assumed that what we know are perceptions of f and $P(X)$, denoted as f^* and $P^*(X)$, respectively. More specifically, we assume that X and f are granular (linguistic) variables and f^* is described by a collection of granular (linguistic) if–then rules:

$$f^*: \quad \{\text{if } X \text{ is } A_i \text{ then } Y \text{ is } B_i\}, \quad i=1,\ldots,m, \tag{4.47}$$

where A_i and B_i are granules of X and Y, respectively (Fig. 12). Equivalently, f^* may be expressed as a fuzzy graph

$$f^* = \sum_{i=1}^{m} A_i \times B_i, \tag{4.48}$$

where $A_i \times B_i$ is a cartesian granule in $U \times V$. Furthermore, we assume that the perception of $P(X)$ is described as

$$P^*(X) \text{ is } \sum_{j=1}^{n} p_j \backslash C_j, \tag{4.49}$$

where the C_j are granules of U, and

$$p_j = \text{Prob}\{X \text{ is } C_j\}. \tag{4.50}$$

Now, let $f^*(C_j)$ denote the image of C_j. Then, application of the extension principle yields

$$f^*(C_j) = \sum_{i=1}^{m} m_{ij} \wedge B_i, \tag{4.51}$$

where the matching coefficient, m_{ij}, is given by

$$m_{ij} = \sup(A_i \cap C_j), \tag{4.52}$$

with the understanding that

$$\sup(A_i \cap C_j) = \sup_u (\mu_{A_i}(u) \wedge \mu_{C_j}(u)), \tag{4.53}$$

where $u \in U$ and μ_{A_i} and μ_{C_j} are the membership functions of A_i and C_j, respectively.

In terms of $f^*(C_j)$, the probability distribution of Y may be expressed as

$$P^*(Y) \text{ is } \sum_{j=1}^{n} p_j \backslash f^*(C_j) \tag{4.54}$$

or, more explicitly, as

$$P^*(Y) \text{ is } \sum_{j=1}^{n} p_j \bigg\backslash \left(\sum_i m_{ij} \wedge B_i \right). \tag{4.55}$$

What these examples show is that computation with perception-based functions and probability distribution is both more general and more complex than computation with their measurement-based counterparts.

5. Concluding remarks

The perception-based theory of probabilistic reasoning which is outlined in this paper may be viewed as an attempt to add to probability theory a significant capability—a capability to operate on information which is perception-based. It is this capability that makes it possible for humans to perform a wide variety of physical and mental tasks without any measurements and any computations.

Perceptions are intrinsically imprecise, reflecting a fundamental limitation on the cognitive ability of humans to resolve detail and store information. Imprecision of perceptions places them well beyond the scope of existing meaning-representation and deductive systems. In this paper, a recently developed computational theory of perceptions is used for this purpose. Applicability of this theory depends in an essential way on the ability of modern computers to perform complex computations at a low cost and high reliability.

Natural languages may be viewed as systems for describing perceptions. Thus, to be able to operate on perceptions, it is necessary to have a means of representing the meaning of propositions drawn from a natural language in a form that lends itself to computation. In this paper, the so-called constraint-centered semantics of natural languages serves this purpose.

A conclusion which emerges from these observations is that to enable probability theory to deal with perceptions, it is necessary to add to it concepts and techniques drawn from semantics of natural languages. Without these concepts and techniques, there are many situations in which probability theory cannot answer questions that arise when everyday decisions have to be made on the basis of perception-based information. Examples of such questions are given in this paper.

A related point is that, in perception-based theory of probabilistic reasoning, imprecision can occur on may different levels—and not just on the level of imprecise probabilities. In particular, imprecision can occur on the level of events, counts and relations. More basically, it can occur on the level of definition of such basic concepts as random variable, causality, independence and stationarity. The concept of precisiated natural language may suggest a way of generalizing these and related concepts in a way that would enhance their expressiveness and operationality.

The confluence of probability theory and the computational theory of perceptions opens the door to a radical enlargement of the role of natural languages in probability theory. The theory outlined in this paper is merely a first step in this direction. Many further steps will have to be taken to develop the theory more fully. This will happen because it is becoming increasingly clear that real-world applications of probability theory require the capability to process perception-based information as a basis for rational decisions in an environment of imprecision, uncertainty and partial truth.

Acknowledgements

Research supported in part by ONR Contract N00014-99-C-0298, NASA Contract NCC2-1006, NASA Grant NAC2-1177, ONR Grant N00014-96-1-0556, ONR Grant FDN0014991035, ARO Grant DAAH 04-961-0341 and the BISC Program.

References

Barsalou, L.W., 1999. Perceptual symbol systems. Behav. Brain Sci. 22, 577–660.

Bowen, J., Lai, R., Bahler, D., 1992. Fuzzy semantics and fuzzy constraint networks. Proceedings of the First IEEE Conference on Fuzzy Systems, San Francisco, pp. 1009–1016.

Cresswell, M.J., 1973. Logic and Languages. Methuen, London.

Dempster, A.P., 1967. Upper and lower probabilities induced by a multivalued mapping. Ann. Math. Statist. 38, 325–339.

Dubois, D., Prade, H., 1988. Possibility Theory. Plenum Press, New York.

Dubois, D., Fargier, H., Prade, H., 1993. The calculus of fuzzy restrictions as a basis for flexible constraint satisfaction. Proceedings of the Second IEEE International Conference on Fuzzy Systems, San Francisco, pp. 1131–1136.

Fikes, R.E., Nilsson, N.J., 1971. STRIPS: a new approach to the application of theorem proving to problem solving. Artif. Intell. 2, 189–208.

Good, I.J., 1962. Subjective probability as the measure of a non-measurable set. In: Nagel, E., Suppes, P., Tarski, A. (Eds.), Logic, Methodology and Philosophy of Science. Stanford University Press, Stanford, pp. 319–329.

Grabisch, M., Nguyen, H.T., Walker, E.A., 1995. Fundamentals of Uncertainty Calculi with Applications to Fuzzy Inference. Kluwer, Dordrecht.

Hajek, P., 1998. Metamathematics of Fuzzy Logic. Kluwer, Dordrecht.

Katai, O., Matsubara, S., Masuichi, H., Ida, M., et al., 1992. Synergetic computation for constraint satisfaction problems involving continuous and fuzzy variables by using occam. In: Noguchi, S., Umeo, H. (Eds.), Transputer/Occam, Proceedings of the Fourth Transputer/Occam International Conference. IOS Press, Amsterdam, pp. 146–160.

Kaufmann, A., Gupta, M.M., 1985. Introduction to Fuzzy Arithmetic: Theory and Applications. Von Nostrand, New York.

Klir, G., Folger, T.A., 1988. Fuzzy Sets, Uncertainty, and Information. Prentice-Hall, Englewood Cliffs, NJ.

Klir, G., Yuan, B., 1995. Fuzzy Sets and Fuzzy Logic. Prentice-Hall, Englewood Cliffs, NJ.

Krause, P., Clark, D., 1993. Representing Uncertain Knowledge. Kluwer, Dordrecht.

Kruse, R., Meyer, D., 1987. Statistics with Vague Data. Kluwer, Dordrecht.
Kuipers, B.J., 1994. Qualitative Reasoning. MIT Press, Cambridge.
Mares, M., 1994. Computation over Fuzzy Quantities. CRC Press, Boca Raton.
Nguyen, H.T., Kreinovich, V., Wu, B., 1999. Fuzzy/probability-fractal/smooth. Internat. J. Uncertainty Fuzziness Knowledge-based Systems 7, 363–370.
Novak, V., 1991. Fuzzy logic, fuzzy sets, and natural languages. Int. J. Gen. Systems 20, 83–97.
Novak, V., Ramik, M., Cerny, M., Nekola, J. (Eds.), 1992. Fuzzy Approach to Reasoning and Decision-Making. Kluwer, Boston.
Pedrycz, W., Gomide, F., 1998. Introduction to Fuzzy Sets. MIT Press, Cambridge.
Ralescu, A.L., Bouchon-Meunier, B., Ralescu, D.A., 1995. Combining fuzzy quantifiers. Proceedings of the International Conference of CFSA/IFIS/SOFT'95 on Fuzzy Theory and Applications, Taipei, Taiwan.
Ralescu, D.A., 1995. Cardinality, quantifiers and the aggregation of fuzzy criteria. Fuzzy Sets and Systems 69, 355–365.
Reghis, M., Roventa, E., 1998. Classical and Fuzzy Concepts in Mathematical Logic and Applications. CRC Press, Boca Raton.
Shafer, G., 1976. A Mathematical Theory of Evidence. Princeton University Press, Princeton.
Shen, Q., Leitch, R., 1992. Combining qualitative simulation and fuzzy sets. In: Faltings, B., Struss, P. (Eds.), Recent Advances in Qualitative Physics. MIT Press, Cambridge.
Smithson, M., 1989. Ignorance and Uncertainty. Springer, Berlin.
Strat, T.M., 1992. Representative applications of fuzzy measure theory: decision analysis using belief functions. In: Wang, Z., Klir, G.J. (Eds.), Fuzzy Measure Theory. Plenum Press, New York, pp. 285–310.
Thomas, S.F., 1995. Fuzziness and Probability. ACG Press, Wichita, KS.
Vallee, R., 1995. Cognition et Systeme. l'Interdisciplinaire Systeme(s), Paris.
Walley, P., 1991. Statistical Reasoning with Imprecise Probabilities. Chapman & Hall, London.
Wang, Z., Klir, G., 1992. Fuzzy Measure Theory. Plenum Press, New York.
Yager, R.R., 1989. Some extensions of constraint propagation of label sets. Internat. J. Approx. Reasoning 3, 417–435.
Zadeh, L.A., 1955. General filters for separation of signals and noise. Proceedings of the Symposium on Information Networks, Polytechnic Institute of Brooklyn, New York, pp. 31–49.
Zadeh, L.A., 1968. Probability measures of fuzzy events. J. Math. Anal. Appl. 23, 421–427.
Zadeh, L.A., 1973. Outline of a new approach to the analysis of complex systems and decision processes. IEEE Trans. Systems Man Cybernet. SMC-3, 28–44.
Zadeh, L.A., 1975. The concept of a linguistic variable and its application to approximate reasoning. Part I: Inf. Sci. 8, 199–249; Part II: Inf. Sci. 8, 301–357; Part III: Inf. Sci. 9, 43–80.
Zadeh, L.A., 1978. Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets and Systems 1, 3–28.
Zadeh, L.A., 1979. Fuzzy sets and information granularity. In: Gupta, M., Ragade, R., Yager, R. (Eds.), Advances in Fuzzy Set Theory and Applications. North-Holland, Amsterdam, pp. 3–18.
Zadeh, L.A., 1981. Test-score semantics for natural languages and meaning representation via PRUF. In: Rieger, B. (Ed.), Empirical Semantics. Brockmeyer, Bochum, W. Germany, pp. 281–349.
Zadeh, L.A., 1983. A computational approach to fuzzy quantifiers in natural languages. Comput. Math. 9, 149–184.
Zadeh, L.A., 1986. Outline of a computational approach to meaning and knowledge representation based on a concept of a generalized assignment statement. In: Thoma, M., Wyner, A. (Eds.), Proceedings of the International Seminar on Artificial Intelligence and Man–Machine Systems. Springer, Heidelberg, pp. 198–211.
Zadeh, L.A., 1995. Probability theory and fuzzy logic are complementary rather than competitive. Technometrics 37, 271–276.
Zadeh, L.A., 1996. Fuzzy logic and the calculi of fuzzy rules and fuzzy graphs: a precis. Multivalued Logic 1, 1–38.
Zadeh, L.A., 1997. Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. Fuzzy Sets and Systems 90, 111–127.
Zadeh, L.A., 1999. From computing with numbers to computing with words—from manipulation of measurements to manipulation of perceptions. IEEE Trans. Circuits Systems 45, 105–119.
Zadeh, L.A., 2000. Outline of a computational theory of perceptions based on computing with words. In: Sinha, N.K., Gupta, M.M. (Eds.), Soft Computing and Intelligent Systems: Theory and Applications. Academic Press, London, pp. 3–22.