POSSIBILITY THEORY AND SOFT DATA ANALYSIS

by

L. A. Zadeh

Memorandum No. UCB/ERL M79/66

27 August 1979

(cover)

POSSIBILITY THEORY AND SOFT DATA ANALYSIS

by

L. A. Zadeh

Memorandum No. UCB/ERL M79/66

27 August 1979

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

Abstract

A thesis advanced in this paper is that much of the uncertainty which is associated with soft data is nonstatistical in nature. Based on this premise, an approach to the representation and manipulation of soft data--in which the recently developed theory of possibility plays a central role--is described and illustrated with examples.

1. Introduction

The term *soft data* does not have a universally agreed upon meaning. Some use it to characterize data that are imprecise or uncertain, while others attach the label "soft" to data whose credibility is open to question.

In dealing with soft data of the type encountered in such diverse fields as psychology, sociology, anthropology, medicine, economics, management science, operations research, pattern classification and systems analysis, it is a standard practice to rely almost entirely on the techniques provided by probability theory and statistics, especially in applications relating to parameter estimation, hypothesis testing and system identification. It can be argued, however, as we do in the present paper, that such techniques cannot cope

---

* To Professor Brian R. Gaines.
Supported by Naval Electronic Systems Command Contract N00039-78-C-0013 and National Science Foundation Grant ENG78-23143.

effectively with those problems in which the softness of data is nonstatistical in nature--in the sense that it relates, in the main, to the presence of fuzzy sets rather than to random measurement errors or data variability.

Needless to say, the inability of conventional statistical techniques to deal with problems of this type would not matter much if the predominance of fuzziness in softness were a rare phenomenon. In reality, the opposite is the case; for, upon closer examination, it becomes clear that much of the softness in data analysis is nonstatistical in the sense explicated above. Moreover, the same is true of most of the linguistic information that humans manipulate through an implicit use of what might be called *approximate* (or *fuzzy*) *reasoning* based on fuzzy rather than standard logic.

To make the latter point more concretely, it will be helpful to list--and subsequently analyze in greater detail--several typical examples of everyday type of questions which cannot be handled effectively by conventional probability-based methods. In these questions, the soft data are expressed as propositions appearing above the horizontal line; the italicized words are the labels of fuzzy sets; and the answers are expected to be in the form of a fuzzy proposition, that is, a proposition whose constituents may have a fuzzy denotation. Specifically:

(a) X is a *large number*
Y is *much larger* than X
____
How large is Y?

(b) *Most* Frenchmen are not *tall*
Elie is a Frenchman
____
How tall is Elie?

(c) It is *unlikely* that Andrea is very *young*
It is *likely* that Andrea is *young*
It is *very unlikely* that Andrea is *old*

How likely is it that Andrea is not old?

(d) It is *true* that Hans is not very *tall*
It is *very true* that Hans is not *short*

How tall is Hans?

(e) Brian is *much taller* than *most* of his *close friends*

How tall is Brian?

(f) If Bernadette lives in Versailles then she is *very rich*
If Bernadette lives in Monmartre then she is *poor*
It is *likely* that Bernadette lives in Versailles
It is *very unlikely* that Bernadette lives in Monmartre

How likely is it that Bernadette is not rich?

As will be seen in the sequel, our approach to the analysis of soft data of the type illustrated by the above examples is based on *fuzzy logic* [7,22,89] rather than on a combination of classical logic and probability-based methods —as is true of the conventional approaches to soft data analysis. In essence, our rationale for the use of fuzzy logic for soft data analysis rests on the premise that the denotations of imprecise terms which occur in a soft database are, for the most part, fuzzy sets rather than probability distributions. For example, in a proposition such as

$$p \triangleq X \text{ is a } \textit{large number} \qquad (1.1)$$

the softness of data is due to the fuzziness of the denotation of *large number*. Similarly, in the proposition

$$p \triangleq \text{It is } \textit{likely} \text{ that Andrea is } \textit{young} \qquad (1.2)$$

softness is due to: (a) the fuzziness of the denotation of *young*; and (b) the fuzziness of the term *likely*, which characterizes the probability of the fuzzy event "Andrea is young" [71,86]. As we shall see presently, the imprecision in (1.1) is possibilistic in nature, whereas in (1.2) it is partly probabilistic and partly possibilistic. Viewed in this perspective, then, a *soft datum* may be regarded, in general, as a proposition in which the uncertainty is due to a combination of probabilistic and possibilistic constituents.

When it is necessary to differentiate between a term and its denotation, the latter will be expressed in uppercase symbols. To illustrate, in (1.1) the term *large number* (or, simply, *large*) has as its denotation a fuzzy subset, LARGE, of the interval $U \triangleq [0,\infty)$. This subset is characterized by its membership function $\mu_{LARGE}: U \rightarrow [0,1]$ which associates with each number $u \in U$ the grade of membership of u in LARGE. For example, the grade of membership of $u = 100$ in LARGE might be 0.2 while that of 400 might be 0.9.

A basic aspect of a fuzzy proposition such as "X is small" is that it does not provide a precise characterization of the value of X. Instead, it defines a *possibility distribution* [92] of values of X which associates with each nonnegative real number u a number in the interval [0,1] which represents the *possibility* that X could take u as a value given the proposition "X is small." To express this in a symbolic form, we write

$$X \text{ is small} \rightarrow \Pi_X = \text{SMALL} \qquad (1.3)$$

which signifies that the proposition "X is small" translates into the assignment of the fuzzy set SMALL to the possibility distribution of X, Π_X. Equivalently, the proposition "X is small" will be said to *induce* the possibility distribution Π_X, with the right-hand member of (1.3) constituting a *possibility assignment equation*. For notational convenience, we shall write

$$\text{Poss}\{X = u\} \triangleq \pi_X(u) \qquad (1.4)$$

where the function $\pi_X\colon U \to [0,1]$ is the *possibility distribution function* and U is the domain of X.

Essentially, the possibility distribution of X is the collection of possible values of X, with the understanding that possibility is a matter of degree, so that the possibility that X could take u as a value may be any number in the interval [0,1] or, more generally, a point in a partially ordered set.

In general, a possibility distribution may be induced by a physical constraint or, alternatively, may be epistemic in origin. To illustrate the difference, let X be the number of passengers that can be carried in Carole's car, which is a five passenger Mercedes. In this case, by identifying $\pi_X(u)$ with the degree of ease with which u passengers can be put in Carole's car, the tabulation of π_X may assume the following form in which an entry such as (7,0.6) signifies that, by

X	1	2	3	4	5	6	7	8	9	10
π_X	1	1	1	1	1	0.8	0.6	0.9	0	0

some explicit or implicit criterion, the degree of ease with which 7 passengers can be carried in Carole's car is 0.6.

In the above example, the possibility distribution of X is induced by a physical constraint on the number of passengers that can be carried in Carole's car. To illustrate the case where the possibility distribution of X is epistemic in origin, i.e., reflects the state of knowledge about X, let X be Carole's age and let the information about Carole's age be conveyed by the proposition

$$p \triangleq \text{Carole is young} \qquad (1.5)$$

where young is the label of a specified fuzzy subset of the interval [0,100] which is characterized by its membership function μ_{YOUNG}, with $\mu_{YOUNG}(u)$ representing the degree to which a person who is u years old is young in a specified context.

The connection between π_X and μ_{YOUNG} is provided by the so-called *possibility postulate* of possibility theory [92,93] which asserts that, in the absence of any information about X other than that supplied by the proposition $p \triangleq$ Carole is young, the possibility that X = u is numerically equal to the grade of membership of u in YOUNG. Thus

$$\text{Poss}\{X=u\} = \pi_X(u) = \mu_{YOUNG}(u)\ , \quad u \in [0,100] \qquad (1.6)$$

or, equivalently,

$$\Pi_{Age(Carole)} = YOUNG \qquad (1.7)$$

with the understanding that the possibility assignment equation (1.7) is the translation of (1.5), i.e.,

$$\text{Carole is young} \longrightarrow \Pi_{Age(Carole)} = YOUNG\ . \qquad (1.8)$$

It is in this sense, then, that the epistemic possibility distribution of Carole's age is induced by the proposition $p \triangleq$ Carole is young.

What is the difference between probability and possibility? As the above examples indicate, the concept of possibility is an abstraction of our intuitive perception of ease of attainment or degree of compatibility, whereas the concept of probability is rooted in the perception of likelihood, frequency, proportion or strength of belief. Furthermore, as we shall see in Section 2, the rules governing the manipulation of possibilities are distinct from those which apply to probabilities.

An important aspect of the connection between probabilities and possibilities relates to the fact that they are *independent* characterizations of uncertainty in the sense that from the knowledge of the possibility distribution of a variable X we cannot deduce its probability distribution, and vice-versa. For example, from the knowledge of the possibility distribution of the number of passengers in Carole's car we cannot deduce its probability distribution. Nor can we deduce the possibility distribution from the probability distribution of the number of passengers. However, we can make a weaker assertion to the effect that if the possibility that $X=u$ is small, then it is likely that the probability that $X=u$ is also small. However, from this it does not follow that high possibility implies high probability, as is reflected in the commonly used statements of the form "It is possible but not probable that... ."[1]

In the present paper, we shall focus our attention on only a few of the basic aspects of possibility theory and its applications to the analysis of soft data. Thus, our main concern will be with the representation of soft data in linguistic form and with approximate inference from such data. In addition, we shall touch upon the issue of data granularity and its relation to the theory of evidence. We shall not consider, however, an issue that is of considerable relevance to the analysis of soft data, namely, the representation of imprecise relational dependencies in the form of linguistic decision tables and branching questionnaires [10, 97].

---

[1] It was brought to the author's attention by John E. Shively (Lawrence Berkeley Laboratory) that an interesting case of an interplay between probability and possibility occurs in the historical letter from Einstein to Roosevelt (dated August 2, 1939). In a passage in this letter, Einstein writes:

> In the course of the last four months it has been made probable--through the work of Joliot in France as well as Fermi and Szilard in America--that it may become possible to set up nuclear chain reactions in a large mass of uranium, by which vast amounts of power and large quantities of new radium-like elements would be generated. Now it appears almost certain that this could be achieved in the immediate future.

2. Basic Properties of Possibility Distributions

As we have indicated in the preceding section, the concept of a possibility distribution plays a central role in our approach to the representation and manipulation of soft data. In what follows, we shall discuss some of the basic properties of possibility distributions[2] and lay the groundwork for their application to soft data analysis in later sections.

Possibility Measure

Consider a variable X which takes values in a universe of discourse U, and let Π_X be the possibility distribution induced by a proposition of the form

$$p \triangleq X \text{ is } G \tag{2.1}$$

where G is a fuzzy subset of U which is characterized by its membership function μ_G. In consequence of the possibility postulate, we can assert that

$$\Pi_X = G \tag{2.2}$$

which implies that

$$\pi_X(u) = \mu_G(u) \ , \quad u \in U \tag{2.3}$$

where π_X is the possibility distribution function of X.

Now if F is a fuzzy subset of U, then the *possibility measure* of F is defined by the expression

$$\Pi(F) = \sup(F \cap G) \tag{2.4}$$

or, more explicitly,

$$\Pi(F) = \sup_U \left(\mu_F(u) \wedge \mu_G(u)\right) \tag{2.5}$$

where the supremum is taken over $u \in U$ and ^ represents the min operation. The number $\Pi(F)$, which ranges in value from 0

[2] In our exposition of the basic properties of possibility distributions and related concepts we shall draw on some of the definitions and examples in [91,94,98].

to 1, may be interpreted as the possibility that X is F given that X is G. Thus, in symbols,

$$\Pi(F) = \text{Poss}\{X \text{ is } F \mid X \text{ is } G\} = \sup(F \cap G) \tag{2.6}$$

In particular, if F is a nonfuzzy set A, then

$$\mu_A(u) = 1 \quad \text{if} \quad u \in A$$
$$\qquad = 0 \quad \text{if} \quad u \notin A$$

and hence

$$\Pi(A) = \text{Poss}\{X \text{ is } A \mid X \text{ is } G\} = \sup_A(G) \tag{2.7}$$
$$= \sup_A\left(\mu_G(u)\right), \quad u \in U$$

An important immediate consequence of (2.4) is the F-additivity of possibility measures expressed by

$$\Pi(F \cup H) = \Pi(F) \vee \Pi(H) \tag{2.8}$$

where F and H are arbitrary fuzzy subsets of U and V is the max operation. By contrast, the probability measures of F and H have the additive property expressed by

$$P(F \cup H) = P(F) + P(H) - P(F \cap H) \tag{2.9}$$

The fact that possibility measures are F-additive but not additive in the usual sense constitutes one of the basic differences between the concepts of possibility and probability [92].

As a simple illustration of (2.6), assume that the proposition "X is G" has the form

$$p \triangleq X \text{ is small} \tag{2.10}$$

where SMALL is a fuzzy set defined by[3]

$$\text{SMALL} = 1/0 + 0.8/2 + 0.6/3 + 0.4/4 + 0.25 \tag{2.11}$$

---

[3] The notation $F = \mu_1/u_1 + \cdots + \mu_n/u_n$ signifies that F is a collection of fuzzy singletons μ_i/u_i, $i = 1, \ldots, n$, with μ_i representing the grade of membership of u_i in F. More generally, F may be expressed as $F = \sum_i \mu_i/u_i$ or $F = \int_U \mu_F(u)/u$.

In this case, the possibility distribution induced by p is given by

$$\Pi_X = 1/0 + 0.8/2 + 0.6/3 + 0.4/4 + 0.1/5$$

and if the proposition X is F has the form

$$q \triangleq X \text{ is large} \tag{2.12}$$

where LARGE is defined by

$$\text{LARGE} \triangleq 0.2/4 + 0.4/5 + 0.6/6 + 0.8/7 + 1/8 + \cdots ,$$

then

$$\text{SMALL} \cap \text{LARGE} = 0.2/4 + 0.1/5$$

and hence

$$\text{Poss}\{X \text{ is large} | X \text{ is small}\} = 0.2$$

Joint, Marginal and Conditional Possibility Distributions

Let $X \triangleq (X_1,\ldots,X_n)$ be an n-ary variable which takes values in a universe of discourse $U = U_1 \times \cdots \times U_n$, with X_i, $i = 1,\ldots,n$, taking values in U_i. Furthermore, let F be an n-ary fuzzy relation in U which is characterized by its membership function μ_F. Then, the proposition

$$p \triangleq X \text{ is } F \tag{2.13}$$

induces an n-ary joint possibility distribution

$$\Pi_X \triangleq \Pi_{(X_1,\ldots,X_n)} \tag{2.14}$$

which is given by

$$\Pi_{(X_1,\ldots,X_n)} = F \tag{2.15}$$

Correspondingly, the possibility distribution function of X is expressed by

$$\begin{aligned}\pi_{(X_1,\ldots,X_n)}(u_1,\ldots,u_n) &= \mu_F(u_1,\ldots,u_n) , \quad u \triangleq (u_1,\ldots,u_n) \in U \\ &= \text{Poss}\{X_1 = u_1,\ldots,X_n = u_n\}\end{aligned}$$

As in the case of probabilities, we can define marginal and conditional possibilities. Thus, let $s \triangleq (i_1,\ldots,i_k)$ be a subsequence of the index sequence $(1,\ldots,n)$ and let s' denote the complementary subsequence $s' \triangleq (j_1,\ldots,j_m)$ (e.g., for $n = 5$, $s = (1,3,4)$ and $s' = (2,5)$). In terms of such sequences, a k-tuple of the form $(A_{i_1},\ldots,A_{i_k})$ may be expressed in an abbreviated form as $A_{(s)}$. In particular, the variable $X_{(s)} = (X_{i_1},\ldots,X_{i_k})$ will be referred to as a k-ary *subvariable* of $X \triangleq (X_1,\ldots,X_n)$, with $X_{(s')} = (X_{j_1},\ldots,X_{j_m})$ being a subvariable complementary to $X_{(s)}$.

The *projection* of $\Pi_{(X_1,\ldots,X_n)}$ *on* $U_{(s)} \triangleq U_{i_1} \times \cdots \times U_{i_k}$ is a k-ary possibility distribution denoted by

$$\Pi_{X_{(s)}} \triangleq \text{Proj}_{U_{(s)}} \Pi_{(X_1,\ldots,X_n)} \tag{2.16}$$

and defined by

$$\pi_{X_{(s)}}(u_{(s)}) \triangleq \sup_{u_{(s')}} \pi_X(u_1,\ldots,u_n) \tag{2.17}$$

where $\pi_{X_{(s)}}$ is the possibility distribution function of $\Pi_{X_{(s)}}$. For example, for $n = 2$,

$$\pi_{X_1}(u_1) \triangleq \sup_{u_2} \pi_{(X_1,X_2)}(u_1,u_2)$$

is the expression for the possibility distribution function of the projection of $\Pi_{(X_1,X_2)}$ on U_1. By analogy with the concept of a marginal probability distribution, $\Pi_{X_{(s)}}$ will be referred to as a *marginal possibility distribution.*

As a simple illustration, assume that $n = 3$, $U_1 = U_2 = U_3 = a+b$ or, more conventionally, $\{a,b\}$, and $\Pi_{(X_1,X_2,X_3)}$ is expressed as a linear form

$$\Pi_{(X_1,X_2,X_3)} = 0.8aaa + 1aab + 0.6baa + 0.2bab + 0.5bbb \tag{2.18}$$

in which a term of the form 0.6baa signifies that

$$\text{Poss}\{X_1 = b,\ X_2 = a,\ X_3 = a\} = 0.6$$

To derive $\Pi_{(X_1,X_2)}$ from (2.18), it is sufficient to replace the value of X_3 in each term in (2.18) by the null string Λ. This yields

$$\Pi_{(X_1,X_2)} = 0.8aa + 1aa + 0.6ba + 0.2ba + 0.5bb$$
$$= 1aa + 0.6ba + 0.5bb$$

and similarly

$$\Pi_{X_1} = 1a + 0.6b + 0.5b$$
$$= 1a + 0.6b$$

An n-ary possibility distribution is *particularized* by forming the conjunction of the propositions "X is F" and "$X_{(s)}$ is G," where $X_{(s)}$ is a subvariable of X. Thus,

$$\Pi_X[\Pi_{X_{(s)}} = G] \triangleq F \cap \bar{G} \tag{2.19}$$

where the right-hand member denotes the intersection of F with the cylindrical extension of G, i.e., a cylindrical fuzzy set defined by

$$\mu_{\bar{G}}(u_1,\ldots,u_n) = \mu_G(u_{i_1},\ldots,u_{i_k}) \ , \quad (u_1,\ldots,u_n) \in U_1 \times \cdots \times U_n \ . \tag{2.20}$$

As a simple illustration, consider the possibility distribution defined by (2.18), and assume that

$$\Pi_{(X_1,X_2)} = 0.4aa + 0.9ba + 0.1bb \ .$$

In this case,

$$\bar{G} = 0.4aaa + 0.4aab + 0.9baa + 0.9bab + 0.1bba + 0.1bbb$$
$$F \cap \bar{G} = 0.4aaa + 0.4aab + 0.6baa + 0.2bab + 0.1bbb$$

and hence

$$\Pi_{(X_1,X_2,X_3)}[\Pi_{(X_1,X_2)} = G]$$
$$= 0.4aaa + 0.4aab + 0.6baa + 0.2bab + 0.1bbb \ .$$

There are many cases in which the operations of particularization and projection are combined. In such cases it is convenient to use the simplified notation

$$_{X_{(r)}}\Pi[\Pi_{X_{(s)}} = G] \tag{2.21}$$

to indicate that the particularized possibility distribution (or relation) $\Pi[\Pi_{X_{(s)}} = G]$ is projected on $U_{(r)}$, where r, like s, is a subsequence of the index sequence (1,...,n). For example,

$$_{X_1 \times X_3}\Pi[\Pi_{(X_3,X_4)} = G]$$

would represent the projection of $\Pi[\Pi_{(X_3,X_4)} = G]$ on $U_1 \times U_3$. Informally, (2.21) may be interpreted as: Constrain the $X_{(s)}$ by $\Pi_{X_{(s)}} = G$ and read out the $X_{(r)}$. In particular, if the values of $X_{(s)}$--rather than their possibility distribution--are set equal to G, then (2.21) becomes

$$_{X_{(r)}}\Pi[X_{(s)} = G] \ .$$

We shall make use of (2.21) and its special cases in Section 3.

As we shall see in Section 3, if X and Y are variables taking values in U and V, respectively, then the *conditional possibility distribution* of Y given X is induced by a proposition of the form "If X is F then Y is G" and is expressed as $\Pi_{(Y|X)}$, with the understanding that

$$\pi_{(Y|X)}(v|u) \triangleq \text{Poss}\{Y = v | X = u\} \qquad (2.22)$$

where (2.21) defines the conditional possibility distribution function of Y given X.

If we know the distribution function of X and the conditional distribution function of Y given X, then we can construct the joint distribution function of X and Y by forming the conjunction ($\wedge \triangleq$ min)

$$\pi_{(X,Y)}(u,v) = \pi_X(u) \wedge \pi_{(Y|X)}(v|u) \ . \qquad (2.23)$$

However, unlike the identity that holds in the case of probabilities, we can also obtain $\pi_{(X,Y)}(u,v)$ by forming the conjunction of $\pi_{(X|Y)}(u|v)$ and $\pi_{(Y|X)}(v|u)$:

$$\pi_{(X,Y)}(u,v) = \pi_{(X|Y)}(u|v) \wedge \pi_{(Y|X)}(v|u) \ . \qquad (2.24)$$

In yet another deviation from parallelism with probabilities, the marginal possibility distribution function of X may be expressed in more than one way in terms of the joint and conditional possibility distribution functions. More specifically, we may have

(a) $$\pi_X(u) = v_v \pi_{(X,Y)}(u,v) \quad (2.25)$$

where v_v denotes the supremum over $v \in V$;

(b) $$\pi_X(u) = v_v \pi_{(X|Y)}(u|v) \quad (2.26)$$

and

(c) $$\pi_X(u) = \pi_{(X|Y)}(u,\tilde{v}(u)) \quad (2.27)$$

where, for a given u, $\tilde{v}(u)$ is the value of v at which $\pi_{(Y|X)}(v|u) = 1$, if $\tilde{v}(u)$ is defined for every $u \in U$.

Intuitively, (a) represents the possibility of assigning a value to X as perceived by an observer ((X,Y) observer) who observes the joint possibility distribution $\Pi_{(X,Y)}$. Similarly, (b) represents the perception of an observer ((X|Y) observer) who observes only the conditional possibility distribution $\Pi_{(X|Y)}$ and is unconcerned with or unaware of $\Pi_{(Y|X)}$. And (c) expresses the perception of an observer who assumes that v is assigned that value, if it exists, which makes $\pi_{(Y|X)}(v|u)$ equal to unity.

As will be seen in Section 3, the concept of a conditional possibility distribution plays a basic role in the formulation of a generalized form of *modus ponens* and in defining a measure of belief. What is as yet an unsettled issue revolves around the question of how to derive $\pi_{(X|Y)}$ and $\pi_{(Y|X)}$ from $\pi_{(X,Y)}$. Somewhat different answers to this question are presented in [92], [57] and [33]. It may well turn out to be the case that, in contrast to probabilities, there does not exist a unique solution to the problem and that, in general, the answer depends on the perspective of the observer.

The Extension Principle

Let f be a function from U to V. The extension principle--as its name implies--serves to extend the domain of definition of f from U to the set of fuzzy subsets of U. In particular, if F is a finite fuzzy subset of U expressed as

$$F = \mu_1/u_1 + \cdots + \mu_n/u_n$$

then f(F) is a finite fuzzy subset of V defined as

$$f(F) = f(\mu_1/u_1 + \cdots + \mu_n/u_n) \tag{2.28}$$
$$= \mu_1/f(u_1) + \cdots + \mu_n/f(u_n) \;.$$

More generally, if the support of F is a continuum, i.e.,

$$F = \int_U \mu_F(u)/u \tag{2.29}$$

then

$$f(F) = \int_U \mu_F(u)/f(u) \;. \tag{2.30}$$

Furthermore, if U is a cartesian product of $U_1,\ldots U_n$ and f is a mapping from $U_1 \times \cdots \times U_n$ to V, then

$$f(F) = \int_U \mu_F(u_1,\ldots,u_n)/f(u_1,\ldots,u_n) \;. \tag{2.31}$$

In connection with (2.31), it should be noted that there are many cases in which we have only partial information about μ_F, e.g., the knowledge of its projections on $U_1,\ldots,U_n$, which implies that the available information consists of the marginal membership functions $\mu_1,\ldots,\mu_n$, where

$$\mu_i(u_i) = \sup_{u_1,\ldots,u_{i-1},u_{i+1},\ldots,u_n} \mu_F(u_1,\ldots,u_n) \;, \quad i = 1,\ldots,n \;.$$

In such cases, the extension of the domain of definition of f is expressed by

$$f(F) = \int_U \mu_1(u_1) \wedge \cdots \wedge \mu_n(u_n)/f(u_1,\ldots,u_n) \tag{2.32}$$

with the understanding that, in replacing $\mu_F(u_1,\ldots,u_n)$ with

$\mu_1(u_1) \wedge \cdots \wedge \mu_n(u_n)$, we are tacitly invoking the principle of maximal restriction [95], which asserts that, in the absence of complete information about Π_X, we should equate Π_X to the maximal (i.e., least restrictive) possibility distribution which is consistent with the partial information about Π_X.

In applying the extension principle to the analysis of soft data, it is frequently convenient to employ a more explicit representation of f(F) which is equivalent to (2.32). Specifically, on denoting the membership function of f(F) by μ, we have

$$f(F) = \int_V \mu(v)/v \qquad (2.33)$$

where

$$\mu(v) \triangleq \text{Max}_{u_1,\ldots,u_n}\ \mu_1(u_1) \wedge \cdots \wedge \mu_n(u_n) \qquad (2.34)$$

subject to the constraint

$$v = f(u_1,\ldots,u_n)\ .$$

In this form, the extension principle will be employed in Section 4 to reduce the problem of inference from soft data to the solution of a variational problem in mathematical programming.

An important aspect of our approach to the analysis of soft data is the flexibility afforded by the assumption that the variables are allowed to be *linguistic* [90], that is, are allowed to have values that are represented as sentences in a natural or synthetic language, with each such value defining a possibility distribution in the domain of the variable. For example, if *Age* is a linguistic variable, its linguistic values might be of the form:

young	old
not young	not old
very young	very old
not very young	not very old
more or less young	more or less old
quite young	quite old
rather young	rather old

not young and not old
not very young and not very old
- - - - - - - - - - - - - - - - -

where *young* is a primary term which has to be calibrated in a specified context and *old* is its antonym. As we shall see in Section 3, the translation rules for propositions expressed in a natural language provide a method for computing the possibility distribution induced by a proposition of the form "X is ℓ," where ℓ is a linguistic value of X, from the knowledge of the membership functions of the primary term and its antonym.

3. Translation Rules and Meaning Representation

When soft data are represented in the form of propositions in a natural language, it is necessary to have, first, a system for translating such propositions into a more precise form; and second, a set of rules of inference which apply to the translated propositions and which may be employed to arrive at answers to questions regarding the data.

A meaning representation language which is well-suited for this purpose is PRUF [93]. In what follows, we shall state some of the relevant translation rules in PRUF and outline the associated rules of inference.

The translation rules in PRUF serve the purpose of facilitating the composition of the meaning of a complex proposition from the meanings of its constituents. For convenience, the rules in question are categorized into four basic types: Type 1: Rules pertaining to modification; Type II: Rules pertaining to composition; Type III: Rules pertaining to quantification; and Type IV: Rules pertaining to qualification.

Following a discussion of these rules and the associated rules of inference, we shall outline a general translation principle which forms the basis for PRUF, and sketch a general question-answering technique which reduces the problem of inference to the solution of a variational problem in mathematical programming.

Translation Rules

Modifier rule (Type I). Let X be a variable which takes values in a universe of discourse U and let F be a fuzzy subset of U. Consider the proposition

$$p \triangleq X \text{ is } F \qquad (3.1)$$

or, more generally,

$$p = N \text{ is } F \qquad (3.2)$$

where N is a variable, an object or a proposition. For example,

$$p \triangleq \text{Mary is young} \tag{3.3}$$

which may be expressed in the form (3.1), i.e.,

$$p \triangleq \text{Age(Mary) is young} \tag{3.4}$$

by identifying X with the variable Age(Mary).

Now, if in a particular context the proposition X is F translates into

$$X \text{ is } F \rightarrow \Pi_X = F \tag{3.5}$$

then in the same context

$$X \text{ is } mF \rightarrow \Pi_X = F^+ \tag{3.6}$$

where m is a modifier such as *not, very, more or less*, etc., and F^+ is a modification of F induced by m. More specifically: If m = not, then $F^+ = F' =$ complement of F, i.e.,

$$\mu_{F^+}(u) = 1 - \mu_F(u) \ , \quad u \in U \ . \tag{3.7}$$

If m = very, then $F^+ = F^2$, i.e.,

$$\mu_{F^+}(u) = \mu_F^{\,2}(u) \ , \quad u \in U \ . \tag{3.8}$$

If m = more or less, then $F^+ = \sqrt{F}$, i.e.,

$$\mu_{F^+}(u) = \sqrt{\mu_F(u)} \ , \quad u \in U \ . \tag{3.9}$$

As a simple illustration of (3.8), if SMALL is defined as in (2.11), then

$$X \text{ is very small} \rightarrow \Pi_X = F^2 \tag{3.10}$$

where

$$F^2 = 1/0 + 1/1 + 0.64/2 + 0.36/3 + 0.16/4 + 0.04/5 \ .$$

It should be noted that (3.7), (3.8) and (3.9) should be viewed as default rules which may be replaced by other translation rules in cases in which some alternative interpretations of the modifiers *very* and *more or less* are more appropriate.

<u>Conjunctive, Disjunctive and Implicational Rules (Type II)</u>. If

$$X \text{ is } F \rightarrow \Pi_X = F \quad \text{and} \quad Y \text{ is } G \rightarrow \Pi_Y = G \tag{3.11}$$

where F and G are fuzzy subsets of U and V, respectively, then

(a) $$X \text{ is } G \text{ and } Y \text{ is } G \rightarrow \Pi_{(X,Y)} = F \times G \tag{3.12}$$

where

$$\mu_{F\times G}(u,v) \triangleq \mu_F(u) \wedge \mu_G(v) . \tag{3.13}$$

(b) $$X \text{ is } F \text{ or } Y \text{ is } G \rightarrow \Pi_{(X,Y)} = \bar{F} \cup \bar{G} \tag{3.14}$$

where

$$\bar{F} \triangleq F \times V , \quad \bar{G} \triangleq U \times G \tag{3.15}$$

and

$$\mu_{\bar{F}\cup\bar{G}}(u,v) = \mu_F(u) \vee \mu_G(v) . \tag{3.16}$$

(c) $$\text{If } X \text{ is } F \text{ then } Y \text{ is } G \rightarrow \Pi_{(Y|X)} = \bar{F}' \oplus \bar{G} \tag{3.17}$$

where $\Pi_{(Y|X)}$ denotes the conditional possibility distribution of Y given X, and the bounded sum $\oplus$ is defined by

$$\mu_{\bar{F}'\oplus\bar{G}}(u,v) = 1 \wedge (1 - \mu_G(u) + \mu_G(v)) . \tag{3.18}$$

In stating the implicational rule in the form (3.17), we have merely chosen one of several alternative ways in which the conditional possibility distribution $\Pi_{(Y|X)}$ may be defined, each of which has some advantages and disadvantages depending on the application. Among the more important of these are the following [5,49,66]:

(c_2) $$\text{If } X \text{ is } F \text{ then } Y \text{ is } G \rightarrow \Pi_{(Y|X)} = \bar{F}' \cup G \tag{3.19}$$

(c_3) $$\text{If } X \text{ is } F \text{ then } Y \text{ is } G \rightarrow \Pi_{(Y|X)} = F\times G \cup F'\times V \tag{3.20}$$

(c_4) $$\text{If } X \text{ is } F \text{ then } Y \text{ is } G \rightarrow \pi_{(Y|X)}(v|u) \tag{3.21}$$

$$= 1 \text{ if } \mu_G(v) \geq \mu_F(u)$$

$$= \frac{\mu_G(v)}{\mu_F(u)} \text{ otherwise}$$

(c_5) If X is F then Y is G $\rightarrow \pi_{(Y|X)}(v|u)$ (3.22)

$$= 1 \text{ if } \mu_G(v) \geq \mu_F(u)$$
$$= \mu_G(v) \text{ otherwise}$$

As simple illustrations of (3.12), (3.14) and (3.17), if

$$F \triangleq \text{SMALL} = 1/1 + 0.6/2 + 0.1/3$$
$$G \triangleq \text{LARGE} = 0.1/1 + 0.6/2 + 1/3$$

then

X is small and Y is large $\rightarrow \Pi_{(X,Y)}$

$$= 0.1/(1,1) + 0.6/(1,2) + 1/(1,3) + 0.1(2,1) + 0.6/(2,2) + 0.6/(2,3) + 0.1/(3,1) + 0.1/(3,2) + 0.1/(3,3)$$

X is small or Y is large $\rightarrow \Pi_{(X,Y)}$

$$= 1/(1,1) + 1/(1,2) + 1/(1,3) + 0.6/(2,1) + 0.6/(2,2) + 1/(2,3) + 0.1/(3,1) + 0.6/(3,2) + 1/(3,3)$$

and

If X is small then Y is large $\rightarrow \Pi_{(Y|X)}$

$$= 0.1/(1,1) + 0.6/(1,2) + 1/(1,3) + 0.5/(2,1) + 1/(2,2) + 1/(2,3) + 1/(3,1) + 1/(3,2) + 1/(3,3) \; .$$

Quantification Rule (Type III). If $U = \{u_1, \ldots, u_N\}$, Q is a quantifier such as *many, few, several, all, some, most,* etc., and

$$X \text{ is } F \rightarrow \Pi_X = F \tag{3.23}$$

then the proposition "QX is F" (e.g., "many X's are large") translates into

$$\Pi_{\text{Count}(F)} = Q \tag{3.24}$$

where Count(F) denotes the number (or the proportion) of elements of U which are in F. By the definition of cardinality of F [90], if the fuzzy set F is expressed as

$$F = \mu_1/u_1 + \mu_2/u_2 + \cdots + \mu_N/u_N \tag{3.25}$$

then

$$\text{Count}(F) = \sum_{i=1}^{N} \mu_i \tag{3.26}$$

where the right-hand member is understood to be rounded-off to the nearest integer. As a simple illustration of (3.24), if the quantifier *several* is defined as

$$\text{SEVERAL} \triangleq 0/1 + 0.4/2 + 0.6/3 + 1/4 + 1/5 + 1/6 + 0.6/7 + 0.2/8 \tag{3.27}$$

then

$$\text{Several X's are large} \rightarrow \Pi_{\sum_{i=1}^{N} \mu_{LARGE}(u_i)} \tag{3.28}$$

$$= 0/1 + 0.4/2 + 0.6/3 + 1/4 + 1/5 + 1/6 + 0.6/7 + 0.2/8$$

where $\mu_{LARGE}(u_i)$ is the grade of membership of the i^{th} value of X in the fuzzy set LARGE.

Alternatively, and perhaps more appropriately, the cardinality of F may be defined as a fuzzy number, as is done in [91]. Thus, if the elements of F are sorted in descending order, so that $\mu_n \leq \mu_m$ if $n \geq m$, then the truth-value of the proposition

$$p \triangleq F \text{ has at least } n \text{ elements} \tag{3.29}$$

is defined to be equal to μ_n, while that of q,

$$q \triangleq F \text{ has at most } n \text{ elements} , \tag{3.30}$$

is taken to be $1 - \mu_{n+1}$. From this, then, it follows that the truth-value of the proposition r,

$$r \triangleq F \text{ has exactly } n \text{ elements} , \tag{3.31}$$

is given by $\mu_n \wedge (1-\mu_{n+1})$.

Let F↓ denote F sorted in descending order. Then (3.29) may be expressed compactly in the equivalent form

$$\text{FGCount}(F) = F{\downarrow} \tag{3.32}$$

which signifies that if the fuzzy cardinality of F is defined in terms of (3.29), with G standing for *greater than*, then the fuzzy count of elements in F is given by $F\downarrow$, with the understanding that $F\downarrow$ is regarded as a fuzzy subset of $\{0,1,2,\ldots\}$. In a similar fashion, (3.30) leads to the definition

$$\text{FLCount}(F) = (F\downarrow)' - 1 \qquad (3.33)$$

where L stands for *less than* and subtraction should be interpreted as translation to the left, while (3.31) leads to

$$\text{FECount}(F) = (F\downarrow) \cap ((F\downarrow)' - 1)$$

where E stands for *equal to*. For convenience, we shall refer to FGCount, FLCount and FECount as the FG cardinality, FL cardinality and FE cardinality, respectively. The concept of FG cardinality will be illustrated in Example 9, Section 5.

Remark. There may be some cases in which it may be appropriate to normalize the definition of FECount in order to convey a correct perception of the count of elements in a fuzzy set. In such cases, we may employ the definition

$$\text{FENCount}(F) = \frac{\text{FECount}(F)}{\text{Max}_n(\mu_n \wedge (1-\mu_{n+1}))} . \qquad (3.34)$$

<u>Truth Qualification Rule (Type IV)</u>. Let τ be a linguistic truth-value, e.g., *very true, quite true, more or less, true*, etc. Such a truth-value may be regarded as a fuzzy subset of the unit interval which is characterized by a membership function $\mu_\tau : [0,1] \rightarrow [0,1]$.

A truth-qualified proposition, e.g., "It is τ that X is F," is expressed as "X is F is τ." As shown in [89], the translation rule for such propositions is given by

$$X \text{ is } F \text{ is } \tau \longrightarrow \Pi_X = F^+ \qquad (3.35)$$

where

$$\mu_{F^+}(u) = \mu_\tau(\mu_F(u)) . \qquad (3.36)$$

As an illustration, consider the truth-qualified proposition

Yolanda is young is very true

which by (3.35), (3.36) and (3.8) translates into

$$\Pi_{Age(Yolanda)} = \mu_{TRUE^2}(\mu_{YOUNG}(u)) \quad . \tag{3.37}$$

Now, if we assume that

$$\mu_{YOUNG}(u) = (1 + (\frac{u}{25})^2)^{-1} \quad , \quad u \in [0,100] \tag{3.38}$$

and

$$\mu_{TRUE}(v) = v^2 \quad , \quad v \in [0,1]$$

then (3.36) yields

$$\Pi_{Age(Yolanda)} = (1 + (\frac{u}{25})^2)^{-4}$$

as the possibility distribution of the age of Yolanda.

Probability Qualification Rule (Type IV). This rule applies to propositions of the general form "X is F is λ," where X is a real-valued variable, F is a linguistic value of X, and λ is a linguistic value of likelihood (or probability), e.g., "X is small is not very likely." Unless stated to the contrary, λ is assumed to be a fuzzy subset of the unit interval [0,1] which is characterized by its membership function μ_λ, and the probability distribution of X is characterized by its probability density function p, i.e.,

$$\text{Prob}\{X \in [u,u+du]\} = p(u)du \quad . \tag{3.39}$$

As shown in [93], the translation rule for probability-qualified propositions is expressed by

$$X \text{ is } F \text{ is } \lambda \rightarrow \pi(p) = \mu_\lambda\left(\int_U \mu_F(u)p(u)du\right) \tag{3.40}$$

where $\pi(p)$ denotes the possibility that the probability density function of X is p, and the integral in the right-hand member of (3.40) represents the probability of the fuzzy event [86] "X is F." Thus, in the case of probability-qualified propositions, the proposition "X is F is λ" induces a possibility distribution of the probability density function of X.

As a simple illustration, consider the proposition

$$q \triangleq \text{Yolanda is young is very likely} \ . \qquad (3.41)$$

In this case, $X \triangleq \text{Age(Yolanda)}$ and the right-hand member of (3.40) becomes

$$\pi(p) = \mu^2_{LIKELY}\left(\int_0^{100} \mu_{YOUNG}(u)p(u)du\right) \ . \qquad (3.42)$$

Used in combination, the translation rules stated above provide a system for the determination of the possibility distributions induced by a fairly broad class of composite propositions. For example, by the use of (3.7), (3.8), (3.9), (3.12) and (3.18), the proposition

If X is not very large and Y is more or less small then Z is very very large.

can readily be found to induce the conditional possibility distribution described by

$$\pi_{(Z|X,Y)}(w|u,v) = 1 \wedge \left(1 - \left(1 - \mu^2_{LARGE}(u)\right) \wedge \mu^{0.5}_{SMALL}(v) + \mu^4_{LARGE}(w)\right) \ .$$

It is of interest to note that translation rules like those described above have found practical applications in the design of fuzzy logic controllers in steel plants, cement kilns and other types of industrial process control applications in which instructions expressed in a natural language are transformed into control signals [45,46,39,79].

A more general type of translation process in PRUF which subsumes the translation rules given above is the following.

Let $\mathcal{D} = \{D\}$ denote a collection of databases, with D representing a generic element of $\mathcal{D}$. For the purposes of our analysis, D will be assumed to consist of a collection of possibly time-varying relations. If R is a constituent relation in D, then by the *frame* of R is meant the name of R together with the names of its columns (i.e., attributes). For example, if a constituent of D is a relation labeled POPULATION whose tableau is comprised of columns labeled Name and Height then the frame of POPULATION is represented as POPULATION‖Name|Height| or, equivalently, as POPULATION[Name;Height].

If p is a proposition in a natural language, its translation into PRUF can assume one of three--essentially equivalent--forms.[4]

(a) p $\rightarrow$ a possibility assignment equation

(b) p $\rightarrow$ a procedure which yields for each D in $\mathcal{D}$ the possibility of D given p, i.e., Poss{D|p}

(c) p $\rightarrow$ a procedure which yields for each D in $\mathcal{D}$ the truth-value of p relative to D, i.e., Tr{p|D}

Remark. An important implicit assumption about the procedures involved in (b) and (c) is that they have a high degree of what might be called *explanatory effectiveness*, by which is meant a capability to convey the meaning of p to a human (or a machine) who is conversant with the meaning of the constituent terms in p but not with the meaning of p as a whole. For example, a procedure which merely tabulates the possibility of each D in $\mathcal{D}$ would, in general, have a low degree of explanatory effectiveness if it does not indicate in sufficient detail the way in which that possibility is arrived at. On the other extreme, a procedure which is excessively detailed and lacking in modularity would also have a low degree of explanatory effectiveness because the

[4] It should be noted that (b) and (c) are in the spirit of truth-conditional semantics and possible-world semantics, respectively [15,34]. In their conventional form, however, these semantics have no provision for fuzzy propositions and hence are not suitable for the analysis of soft data.

meaning of p might be obscured by the maze of unstructured steps in the body of the procedure.

The equivalence of (b) and (c) is a consequence of the way in which the concept of truth is defined in fuzzy logic

Thus, it can readily be shown that, under mildly restrictive assumptions on D, we have

$$Tr\{p|D\} = Poss\{D|p\}$$

which implies the equivalence of (b) and (c).

To illustrate (b) and show how (a) may be derived from (b), we shall consider first the relatively simple proposition

$$p \triangleq \text{Madan is not very tall} \,. \qquad (3.43)$$

In this case, it is convenient to assume that D contains two relations whose frames are:

POPULATION‖Name|Height|

TALL‖Height|μ|

In the relation TALL, each value of height is associated with the degree to which a person having that height is tall. In effect, then, the relation TALL defines the fuzzy set TALL.

The desired procedure involves the following steps.

1. Find Madan's height, h. In symbols, h is given by the expression (see (2.21))

$$h = {}_{\text{Height}}\ \text{POPULATION}[\text{Name} = \text{Madan}] \,.$$

2. Find the degree, δ, to which Madan is not very tall in D. Using the expression obtained in the preceding step, the answer is:

$$\delta = 1 - ({}_{\mu}\text{TALL}[\text{Height} = {}_{\text{Height}}\ \text{POPULATION}[\text{Name} = \text{Madan}]])^2 \,.$$

3. Equate the possibility of D to δ. This yields the desired translation of p into PRUF, namely

$$\pi(D) = 1 - ({}_{\mu}\text{TALL}[\text{Height} = {}_{\text{Height}}\ \text{POPULATION}[\text{Name=Madan}]])^2 \qquad (3.44)$$

To find the possibility distribution of Madan's height from (3.41), it is sufficient to observe that, for a fixed relation TALL, $\pi(D)$ depends only on Madan's height. From this it follows at once that

$$\Pi_{Height(Madan)} = (TALL^2)' \qquad (3.45)$$

or, equivalently,

$$\pi_{Height(Madan)}(u) = 1 - \mu^2_{TALL}(u) \qquad (3.46)$$

where u is a generic value of the variable Height. What should be noted is that the possibility assignment equation (3.45) could be obtained directly by applying to p the translation rules (3.7) and (3.8). Furthermore, the explanatory effectiveness of (3.45) is higher than that of (3.44).

Remark. In PRUF, it is important to differentiate between the meaning of a proposition and the information that is conveyed by it. Thus, if p is a proposition, then the procedure, P, into which it translates represents the meaning of p or, equivalently, its intension [15,41]. On the other hand, the possibility distribution which is induced by p constitutes the information, I(p), which is conveyed by p. Thus, in the foregoing example the possibility distribution defined by (3.45) represents the information conveyed by the proposition p $\triangleq$ Madan is not very tall. The meaning of p, then, is the procedure described by the right-hand member of (3.45).

If p and q are propositions such that

$$I(p) = I(q) \qquad (3.47)$$

then p and q are *semantically equivalent* [93], which is expressed as

$$p \leftrightarrow q \,. \qquad (3.48)$$

On the other hand, if

$$I(p) \leq I(q) \qquad (3.49)$$

then p *semantically entails* q [93], i.e.,

$$p \mapsto q \tag{3.50}$$

As we shall see in the next section, the concepts of semantic equivalence and semantic entailment play an important role in inference from soft data.

4. Inference from Soft Data and Mathematical Programming

By interpreting a soft datum as a fuzzy proposition, the problem of inference from soft data may be reduced to the problem of inference from a collection of fuzzy propositions.

Suppose that $E = \{p_1, \ldots, p_n\}$ (with E standing for *evidence*) is a collection of fuzzy propositions and let p be a proposition that is inferred from E. At this point, it is natural to raise two basic questions. First, what does it mean to say that p is inferred from E; and second, by what methods can p be inferred from E.

To answer the first question, it is convenient to make use of the concept of information, as defined in Section 3. More specifically, let $I(p_1 \wedge \cdots \wedge p_n)$ be the information conveyed by the conjunction of propositions $p_1, \ldots, p_n$ or, equivalently, the possibility distribution induced by $p_1 \wedge \cdots \wedge p_n$, and let $I(p_1 \wedge \cdots \wedge p_n \wedge p_n)$ be the information conveyed by the conjunction of $p_1 \wedge \cdots \wedge p_n$ and p. Then, we shall say, informally, that p may be *inferred from* $E = \{p_1, \ldots, p_n\}$ if

$$I(p_1 \wedge \cdots \wedge p_n) = I(p_1 \wedge \cdots \wedge p_n \wedge p) \ . \qquad (4.1)$$

In other words, p is inferrable from E if the addition of p to the evidence, E, does not affect the information conveyed by E.

As shown in [91], the above definition implies that the possibility distribution induced by the conjunction of $p_I, \ldots, p_n$ is contained in that induced by p. It is this containment property that underlies the *entailment principle* [91,93] which serves as a basis for the rules of inference stated in the sequel.

Remark. In speaking of entailment, it is necessary to differentiate between the entailment which obtains for particular denotations of the labels of fuzzy sets in

$p_1, \ldots, p_n, p$, and *strong entailment*, which results when (4.1) holds for all denotations. As an illustration, if *very* is interpreted as a squaring operation, then the proposition

$$p \triangleq \text{Veronica is intelligent}$$

is strongly entailed by

$$p_1 \triangleq \text{Veronica is very intelligent}$$

since

$$\text{INTELLIGENT}^2 \subset \text{INTELLIGENT}$$

regardless of the way in which INTELLIGENT, the denotation of *intelligent*, is defined. On the other hand, in the case of the propositions

$$\begin{array}{l} p_1 \triangleq \text{John is not young} \\ p_2 \triangleq \text{John is not old} \\ \hline p \triangleq \text{John is middle-aged} \end{array}$$

the conjunction of p_1 and p_2 may be expressed as (see (3.12))

$$p_1 \wedge p_2 \triangleq \text{John is not young and not old} . \qquad (4.2)$$

Consequently, if the denotations of *young, old* and *middle-aged* are such that the containment condition

$$\text{YOUNG}' \cap \text{OLD}' \subset \text{MIDDLE-AGED} \qquad (4.3)$$

is satisfied, then p is entailed by p_1 and p_2. However, since the question of whether or not (4.2) is satisfied depends on the denotations of the labels of fuzzy sets in p_1, p_2 and p, it follows that p is not strongly entailed by p_1 and p_2.

Remark. In some ways, the entailment principle appears to be counterintuitive because we generally expect a conclusion, p, to be sharper than the totality of data on which it is based. However, the reason for the apparent sharpness is that, in general, p involves only a small subset of the variables present in $p_1,\ldots,p_n$. More specifically, as we shall see in the sequel, in the process of inference we usually focus our attention on a small number of functionals defined on $\mathcal{D}$ and preceive the higher degree of focusing as a manifestation of sharpness of p. An example illustrating this and other aspects of the entailment principle is described in Section 5.

Rules of Inference

For purposes of inference from a collection of fuzzy propositions, it is convenient to have at one's disposal a system of basic rules which may be used singly or in combination to infer a fuzzy proposition p from a body of evidence $E = \{p_1,\ldots,p_n\}$. Several such rules, which constitute a subset of the inference rules in fuzzy logic, FL [89], are stated in a summary form in the following.

1. Projection Principle. Consider a fuzzy proposition whose translation is expressed as

$$p \rightarrow \Pi_{(X_1,\ldots,X_n)} = F \tag{4.4}$$

and let $X_{(s)}$ denote a subvariable of the variable $X \triangleq (X_1,\ldots,X_n)$, i.e.,

$$X_{(s)} = (X_{i_1},\ldots,X_{i_k}) \tag{4.5}$$

where the index sequence $s \triangleq (i_1,\ldots,i_k)$ is a subsequence of the sequence $(1,\ldots,n)$.

Furthermore, let $\Pi_{X_{(s)}}$ denote the marginal possibility distribution of $X_{(s)}$; that is,

$$\Pi_{X_{(s)}} = \mathrm{Proj}_{U_{(s)}} F \tag{4.6}$$

where U_i, $i = 1,\ldots,n$, is the universe of discourse associated with X_i;

$$U_{(s)} = U_{i_1} \times \cdots \times U_{i_k} \tag{4.7}$$

and the projection of F on $U_{(s)}$ is defined by the possibility distribution function (see (2.17))

$$\pi_{X_{(s)}}(u_{i_1},\ldots,u_{i_k}) = \sup_{u_{j_1},\ldots,u_{j_m}} \mu_F(u_1,\ldots,u_n) \qquad (4.8)$$

where $s' \triangleq (j_1,\ldots,j_m)$ is the index subsequence which is complementary to s, and μ_F is the membership function of F.

Now let q be a retranslation (i.e., reverse translation) of the possibility assignment equation

$$\Pi_{X_{(s)}} = \text{Proj}_{U_{(s)}} F \,. \qquad (4.9)$$

Then, the projection rule asserts that q may be inferred from p. In a schematic form, this assertion may be expressed more transparently as

$$\begin{array}{c} p \longrightarrow \Pi_{(X_1,\ldots,X_n)} = F \\ \downarrow \\ q \longleftarrow \Pi_{X_{(s)}} = \text{Proj}_{U_{(s)}} F \end{array} \qquad (4.10)$$

As was indicated in Section 2, the rule of inference represented by (4.10) is easy to apply when Π_X is expressed as a linear form. As an illustration, assume that $U_1 = U_2 = \{a,b\}$, and

$$\Pi_{(X_1,X_2)} = 0.8aa + 0.6ab + 0.4ba + 0.2bb$$

in which a term of the form 0.6ab signifies that

$$\text{Poss}\{X_1 = a,\ X_2 = b\} = 0.6 \,.$$

To obtain the projection of Π_X on, say, U_2 it is sufficient to replace the value of X_1 in each term by the null string Λ. Thus

$$\text{Proj}_{U_2} \Pi_{(X_1,X_2)} = 0.8a + 0.6b + 0.4a + 0.2b = 0.8a + 0.6b$$

and hence from the proposition

$$(X_1,X_2) \text{ is } 0.8aa + 0.6ab + 0.4ba + 0.2bb$$

we can infer by (4.10) that

$$X_2 \text{ is } 0.8a + 0.6b \ .$$

2. Conjunction Rule. Consider a proposition p which is an assertion concerning the possible values of, say, two variables X and Y which take values in U and V, respectively. Similarly, let q be an assertion concerning the possible values of the variables Y and Z, taking values in V and W. With these assumptions, the translations of p and q may be expressed as

$$\begin{array}{l} p \rightarrow \Pi^{p}_{(X,Y)} = F \\ q \rightarrow \Pi^{q}_{(Y,Z)} = G \end{array} \qquad (4.11)$$

Let $\bar{F}$ and $\bar{G}$ be, respectively, the cylindrical extensions of F and G in $U \times V \times W$. Thus,

$$\bar{F} = F \times W \qquad (4.12)$$

and

$$\bar{G} = U \times G \ . \qquad (4.13)$$

Using the conjunction rule, we can infer from p and q a proposition which is defined by the following scheme:

$$\begin{array}{l} r \rightarrow \Pi^{p}_{(X,Y)} = F \\ q \rightarrow \Pi^{q}_{(Y,Z)} = G \\ \hline r \leftarrow \Pi_{(X,Y,Z)} = \bar{F} \cap \bar{G} \end{array} \qquad (4.14)$$

On combining the projection and conjunction rules, we obtain the *compositional rule of inference* (4.17) which includes the classical *modus ponens* as a special case.

More specifically, on applying the projection rule to (4.14), we obtain the following inference scheme

$$\begin{array}{l} p \rightarrow \Pi^{p}_{(X,Y)} = F \\ q \rightarrow \Pi^{q}_{(Y,Z)} = G \\ \hline r \leftarrow \Pi^{r}_{(X,Z)} = F \circ G \end{array} \qquad (4.15)$$

where the composition of F and G is defined by

$$\mu_{F\circ G}(u,w) = \sup_v \left(\mu_F(u,v) \wedge \mu_G(v,w)\right) . \qquad (4.16)$$

In particular, if p is a proposition of the form "X is F" and q is a proposition of the form "If X is G then Y is H," then (4.15) becomes

$$\begin{array}{l} p \rightarrow \Pi_X = F \\ q \rightarrow \Pi_{(Y|X)} = \bar{G}' \oplus \bar{H} \\ \hline r \leftarrow \Pi_{(Y)} = F\circ(\bar{G}'\oplus\bar{H}) \end{array} \qquad (4.17)$$

The rule expressed by (4.17) may be viewed as a generalized form of *modus ponens* which reduces to the classical *modus ponens* when F = G and F, G, H are nonfuzzy sets.

Stated in terms of possibility distributions, the generalized *modus ponens* places in evidence the analogy between probabilistic and possibilistic inference. Thus, in the case of probabilities, we can deduce the probability distribution of Y from the knowledge of the probability distribution of X and the conditional probability distribution of Y given X. Similarly, in the case of possibility distributions, we can infer the possibility distribution of Y from the knowledge of the possibility distribution of X and the conditional possibility distribution of Y given X.

It is important to note that the generalized *modus ponens* as expressed by (4.17) may be used to enlarge significantly the area of applicability of rule-based systems of the type employed in MYCIN and other expert systems. This is due primarily to two aspects of (4.17) which are not present in conventional rule-based systems: (a) in the propositions "X is F" and "If X is G then Y is H," F, G and H may be fuzzy sets; and (b) F and G need not be identical. Thus, as a result of (a) and (b), a rule-based system employing (4.17) may be designed to have an interpolative capability [88,97].

In addition to the rules described above, there is an important method of inference through which the deduction of p is reduced to the solution of a variational problem in mathematical programming.

In general terms, suppose that we have a database D and that we wish to answer a question q which relates to the data resident in D. For example, we may have a database which contains a relation with the frame

POPULATION	Name	Age

and q may be: What is the average age of individuals in POPULATION?

In PRUF, the translation of q is expressed as the translation of the answer to q, with a symbol of the form $?\alpha$ identifying the variable whose value is to be determined. As an illustration, for the example under discussion the proposition to be inferred from D may be expressed as

$$p \triangleq \text{The average age of individuals in POPULATION is } ?f$$

where f is a function of the entries in D, say $X_1,\ldots,X_m$. Thus, to answer q we must compute the value of $f(X_1,\ldots,X_m)$ from whatever information is available about D.

To link the method under discussion to our earlier formulation of the problem of inference, we shall assume that the available information about D consists of the evidence $E = \{p_1,\ldots,p_n\}$, in which the p_i are fuzzy propositions.

Our definition of translation in Section 3 implies that each of the p_i in E induces a possibility distribution over $\mathcal{D}$. Thus, letting $\pi_i(X_1,\ldots,X_m)$ denote the possibility of D given p_i, we can assert that the possibility of D given $p_1,\ldots,p_n$ is given by the conjunction

$$\pi(D) \triangleq \pi_1(X_1,\ldots,X_m) \wedge \cdots \wedge \pi_n(X_1,\ldots,X_m) \qquad (4.18)$$

Thus, $\pi(D)$, as expressed by (4.18), may be viewed as an elastic constraint on D which is induced by the evidence $E \triangleq \{p_1,\ldots,p_n\}$.

From the knowledge of $\pi(D)$ we can infer the possibility distribution of the function

$$z = f(X_1,\ldots,X_m) \tag{4.19}$$

by invoking the extension principle, as shown in Section 3. In this way, the determination of the possibility distribution of f reduces, in principle, to the solution of the following variational problem in mathematical programming.

$$\mu(z) \triangleq \mathrm{Max}_{X_1,\ldots,X_m} \pi_1(X_1,\ldots,X_m) \wedge \cdots \wedge \pi_n(X_1,\ldots,X_m) \tag{4.20}$$

subject to

$$z = f(X_1,\ldots,X_m) \ .$$

In terms of $\mu(z)$, the possibility distribution of f may be expressed in the form

$$\Pi_f = \int_V \mu(z)/z \tag{4.21}$$

where V is the range of z. An example illustrating the application of this technique will be discussed in Section 5.

As a further example, consider the proposition which occurs in Example (e), Section 1, namely:

p ≜ Brian is much taller than most of his close friends

For the purpose of representing the meaning of p, it is expedient to assume that D is comprised of the relations

POPULATION‖Name|Height|

FRIENDS‖Name1|Name2|μ|

MUCH TALLER‖Height1|Height2|μ|

MOST‖ρ|μ|

In the relation FRIENDS, μ represents the degree to which an individual whose name is Name2 is a friend of Name1. Similarly, in the relation MUCH TALLER, μ represents the degree to which an individual whose height is HEIGHT1 is much taller than one whose height is HEIGHT2. In MOST, μ represents the degree to which a proportion, ρ, fits the definition of MOST as a fuzzy subset of the unit interval.

To represent the meaning of p we shall translate p--in the spirit of (c) (Section 3)--into a procedure which computes the truth-value of p relative to a given D. The procedure--as described below--may be viewed as a sequence of computations which, in combination, yield the truth-value of p.

1. Obtain Brian's height from POPULATION. Thus,

$$\text{Height(Brian)} =_{\text{Height}} \text{POPULATION[Name} = \text{Brian]}$$

2. Determine the fuzzy set, MT, of individuals in POPULATION in relation to whom Brian is much taller.

Let Name_i be the name of the i^{th} individual in POPULATION. The height of Name_i is given by

$$\text{Height}(\text{Name}_i) =_{\text{Height}} \text{POPULATION}[\text{Name} = \text{Name}_i]$$

Now the degree to which Brian is much taller than Name_i is given by

$$\delta_i =_{\mu} \text{MUCH TALLER}[\text{Height(Brian)}, \text{Height}(\text{Name}_i)]$$

and hence MT may be expressed as

$$\text{MT} = \textstyle\sum_i \delta_i/\text{Name}_i \;, \quad \text{Name}_i \in_{\text{Name}} \text{POPULATION}$$

where $_{\text{Name}}$POPULATION is the list of names of individuals in POPULATION, δ_i is the grade of membership of Name_i in MT, and $\sum_i$ is the union of singletons δ_i/Name_i (see footnote 3).

3. Determine the fuzzy set, CF, of individuals in POPULATION who are close friends of Brian.

To form the relation CLOSE FRIENDS from FRIENDS we intensify FRIENDS by squaring it (i.e., by replacing μ with μ^2). Then, the fuzzy set of close friends of Brian is given by

$$\text{CF} =_{\mu \times \text{Name2}} \text{FRIENDS}^2[\text{Name1} = \text{Brian}]$$

4. Form the count of elements of CF:

$$\text{Count}(CF) = \sum_i \mu_{CF}(\text{Name}_i)$$

where $\mu_{CF}(\text{Name}_i)$ is the grade of membership of Name_i in CF and $\sum_i$ is the arithmetic sum. More explicitly

$$\text{Count}(F) = \sum_i \mu^2_{FRIENDS}(\text{Brian}, \text{Name}_i)$$

5. Form the intersection of CF and MT, that is, the fuzzy set of those close friends of Brian in relation to whom he is much taller.

$$H \triangleq CF \cap MT$$

6. Form the count of elements of H.

$$\text{Count}(H) = \sum_i \mu_H(\text{Name}_i)$$

where $\mu_H(\text{Name}_i)$ is the grade of membership of Name_i in H and $\sum_i$ is the arithmetic sum.

7. Form the ratio

$$r = \frac{\text{Count}(MT \cap CF)}{\text{Count}(CF)}$$

which represents the proportion of close friends of Brian in relation to whom he is much taller.

8. Compute the grade of membership of r in MOST

$$\tau = {}_{\mu}\text{MOST}[\rho = r]$$

The value of τ is the desired truth-value of p with respect to D and, equivalently, the possibility of D given p. In terms of the membership functions of FRIENDS, MUCH TALLER and MOST, the value of τ is given explicitly by the expression

$$\tau = \mu_{MOST}\left(\frac{\sum_i \mu_{MT}(\text{Height(Brian)},\text{Height}(\text{Name}_i)) \wedge \mu^2_{CF}(\text{Brian},\text{Name}_i)}{\sum_i \mu^2_{CF}(\text{Brian},\text{Name}_i)}\right) \tag{4.22}$$

In summary, the procedure in question serves to represent the meaning of p by describing the operations that must be performed on D in order to compute the truth-value of p with respect to D. Thus, viewed as an expression in PRUF, (4.22) is in effect a mathematical description of a procedure which defines τ as a function of D. However, as was stressed in Section 3, the meaning of p is the procedure itself rather than the value of τ which it returns for a given D.

5. Examples of Inference from Soft Data

To illustrate the application of some of the techniques described in the preceding sections, we shall consider several simple examples, including Examples (a), (b), (c) and (e) of Section 2. As is generally the case in inference from soft data, the chains of inference in these examples are short.

Example 1 (Example (a), Section 1).

X is a large number
Y is much larger than X

How large is Y?

Solution. On applying the compositional rule of inference (4.15), we obtain the following expression for the possibility distribution of Y

$$\Pi_Y = \text{LARGE} \circ \text{MUCH LARGER} \tag{5.1}$$

or, more explicitly,

$$\pi_Y(v) = \sup_u(\mu_{LARGE}(u) \wedge \mu_{MUCH\ LARGER}(u,v)) \tag{5.2}$$

where LARGE and MUCH LARGER are the fuzzy denotations of *large* and *much larger*, respectively.

Example 2.

X is small
Y is approximately equal to X
Z is much larger than both X and Y

How large is Z?

Solution. Proceeding as in Example 1, we obtain the following expression for the possibility distribution of Z

$$\Pi_Z = (\text{MUCH LARGER THAN} \circ \text{APPROXIMATELY EQUAL} \circ \text{SMALL}) \cap \text{MUCH LARGER THAN} \circ \text{SMALL} \tag{5.3}$$

in which the intersection implies that Z is much larger than

X, and Z is much larger than Y.

Example 3 (Example (b), Section 1).

Most Frenchmen are not tall
Elie is a Frenchman

How tall is Elie?

Solution. First, we interpret the question as follows:

Most Frenchmen are not tall
Elie is a Frenchman picked at random

What is the probability that Elie is tall?

Second, we assume that the database consists of a single relation of the form

POPULATION	Name	μ

in which μ_i is the degree to which Name_i is tall, and i ranges from 1 to N.

Now, the constraint on the database induced by the proposition

$$p \triangleq \text{Most Frenchmen are not tall}$$

gives rise to the possibility distribution expressed by

$$\pi_p(\text{POPULATION}) = \mu_{MOST}\left(\frac{1}{N}\sum_i (1-\mu_i)\right) \qquad (5.4)$$

in which the argument of μ_{MOST} represents the proportion of Frenchmen who are not tall.

Furthermore, if a Frenchman is chosen at random, then the probability that he is tall is given by (see (3.40))

$$\text{Prob}\{\text{Frenchman is tall}\} = \frac{1}{N}\sum_i \mu_i \ . \qquad (5.5)$$

Thus, the proposition (in which λ is a linguistic probability)

$$q \triangleq \text{The probability that a Frenchman is tall is } \lambda$$

induces the possibility distribution

$$\pi_q(\text{POPULATION}) = \mu_\lambda\left(\frac{1}{N}\sum_i \mu_i\right) \ . \qquad (5.6)$$

To apply the entailment principle to the problem in hand, we have to find a λ such that

$$\mu_\lambda(\frac{1}{N}\sum_i \mu_i) \geq \mu_{MOST}(\frac{1}{N}\sum_i \mu_i) \quad . \tag{5.7}$$

Furthermore, to be as informative as possible, the λ in q should be as small as possible in the sense that there should be no λ' such that

$$\lambda'(v) \leq \lambda(v) \tag{5.8}$$

for all v in [0,1] and $\lambda'(v) < \lambda(v)$ for at least some v in [0,1].

With this as our objective, we first note that (5.4) may be rewritten as

$$\begin{aligned}\pi_p(\text{POPULATION}) &= \mu_{MOST}(1 - \frac{1}{N}\sum_i \mu_i) \\ &= \mu_{ANT\ MOST}(\frac{1}{N}\sum_i \mu_i)\end{aligned} \tag{5.9}$$

where ANT MOST stands for the denotation of the antonym of *most*, i.e.,

$$\mu_{ANT\ MOST}(v) = \mu_{MOST}(1-v) \quad , \quad v \in [0,1] \tag{5.10}$$

which signifies that the membership function of ANT MOST is the mirror image of that of MOST.

At this juncture, then, we can assert that

$$\begin{aligned}p &\triangleq \text{Most Frenchmen are not tall} \\ &\longrightarrow \pi_p(\text{POPULATION}) = \mu_{ANT\ MOST}(\frac{1}{N}\sum_i \mu_i)\end{aligned} \tag{5.11}$$

while

$$\begin{aligned}r &\triangleq \text{Prob\{Frenchman is tall\} is } \gamma \\ &\longrightarrow \pi_r(\text{POPULATION}) = \mu_\gamma(\frac{1}{N}\sum_i \mu_i)\end{aligned} \tag{5.12}$$

where γ is a linguistic probability.

On comparing (5.11) with (5.12), we note that if the fuzzy set LIKELY is defined to be equal to MOST, i.e.,

$$\mu_{LIKELY}(v) = \mu_{MOST}(v) \quad , \quad v \in [0,1] \tag{5.13}$$

so that

$$\mu_{UNLIKELY}(v) = \mu_{ANT\ LIKELY}(v) = \mu_{ANT\ MOST}(v) \quad (5.14)$$

then we can infer from (5.11) and (5.12) the semantic equivalence (3.48)

$p \triangleq$ Most Frenchmen are not very tall $\leftrightarrow$
$r \triangleq$ Prob{Frenchman is tall} is unlikely

Consequently, as the answer to the posed question, we have

Most Frenchmen are not tall
Elie is a Frenchman

It is unlikely that Elie is tall

In essence, then, what we have shown is that, under the assumption that the fuzzy sets MOST and LIKELY are equal, we can infer from the premise

$p \triangleq$ Most Frenchmen are not tall

the semantically equivalent proposition

$r \triangleq$ It is unlikely that a Frenchman picked at random is tall

from which it follows that "It is unlikely that Elie is tall."

Example 4.

Most Swedes are tall

How many Swedes are very tall?

Solution. Suppose that the answer is of the form

$r \triangleq$ Q Swedes are very tall

where Q is a fuzzy quantifier. Then, proceeding as in Example 3, we have

$$p \triangleq \text{Most Swedes are tall} \rightarrow \pi_p(\text{POPULATION}) = \mu_{MOST}\left(\frac{1}{N}\sum_i \mu_i\right) \quad (5.15)$$

and

$$r \triangleq Q \text{ Swedes are very tall} \rightarrow \pi_r(\text{POPULATION}) = \mu_Q\left(\frac{1}{N}\sum_i \mu_i^2\right) \quad (5.16)$$

Consequently, what we have to find is the "smallest" Q such that

$$\mu_Q\left(\frac{1}{N}\sum_i \mu_i^2\right) \geq \mu_{MOST}\left(\frac{1}{N}\sum_i \mu_i\right) . \quad (5.17)$$

It can easily be verified that such a Q is given by[5]

$$Q = {}^2MOST \quad (5.18)$$

where the "left-square" of MOST is defined by

$$\mu_{{}^2MOST}(F) = \mu_{MOST}(\sqrt{v}) \ , \quad v \in [0,1] . \quad (5.19)$$

For, from the elementary inequality

$$\sqrt{1/N \, \Sigma_i \mu_i^2} \geq \frac{1}{N}\sum_i \mu_i \quad (5.20)$$

and the monotonicity of μ_{MOST} it follows that

$$\mu_{MOST}\left(\sqrt{1/N \, \Sigma_i \mu_i^2}\right) \geq \mu_{MOST}\left(\frac{1}{N}\sum_i \mu_i\right) \quad (5.21)$$

which, in view of (5.19), implies that

$$\mu_{{}^2MOST}\left(\frac{1}{N}\sum_i \mu_i^2\right) \geq \mu_{MOST}\left(\frac{1}{N}\sum_i \mu_i\right) \quad (5.22)$$

and hence that the proposition

$p \triangleq$ Most Swedes are tall

entails

$q \triangleq {}^2$Most Swedes are very tall

Example 5.

Naomi is not very tall is true

How true is it that Naomi is tall?

Solution. Suppose that the answer to the question is expressed as a proposition q:

$q \triangleq$ Naomi is tall is τ

[5] If MOST is interpreted as a fuzzy number [90,18,20] then ^{2}MOST may be expressed as the product of MOST with itself.

where τ is a linguistic truth-value, e.g., very true, more or less true, etc.

To determine τ, we set q semantically equal to p (see (3.49)), i.e., we assert that the possibility distributions induced by p and q are equal. Now, by (3.8) and (3.36), we have

$$\text{Naomi is not very tall is true} \rightarrow \Pi_{\text{Height(Naomi)}} = F \tag{5.23}$$

where

$$\mu_F(u) = \mu_{TRUE}(1 - \mu^2_{TALL}(u)) \tag{5.24}$$

and

$$\text{Naomi is tall is } \tau \rightarrow \mu_\tau(\mu_{TALL}(u)) \tag{5.25}$$

where μ_{TALL} and μ_{TRUE} are the membership functions of TALL and TRUE, respectively. Consequently, for all u in the domain of the variable Height(Naomi), we have

$$\mu_{TRUE}(1 - \mu^2_{TALL}(u)) = \mu_\tau(\mu_{TALL}(u)) \tag{5.26}$$

from which it follows that the membership function of τ is given by

$$\mu_\tau(v) = 1 - v^2 \ , \quad v \in [0,1] \ . \tag{5.27}$$

Thus, if μ_{TRUE} is defined by

$$\mu_{TRUE}(v) = v^2 \tag{5.28}$$

then

$$\mu_\tau(v) = 1 - \mu_{TRUE}(v) \tag{5.29}$$

and hence

$$\tau = \text{not true} \ . \tag{5.30}$$

On the other hand, if

$$\mu_{TRUE}(v) = v \tag{5.31}$$

then

$$\mu_\tau(v) = 1 - \mu^2_{TRUE}(v) \tag{5.32}$$

and

$$\tau = \text{not very true} \ . \tag{5.33}$$

Example 6.

Marvin lives near MIT
Lucia lives near MIT

What is the distance between the residences of Marvin and Lucia?

Solution. Let (X_M,Y_M) and (X_L,Y_L) be the coordinates of the residences of Marvin and Lucia, respectively. Furthermore, let $\Pi_{(X_M,Y_M)}$ and $\Pi_{(X_L,Y_L)}$ be the possibility distributions induced by p and q, that is, derived from the definition of the binary fuzzy relation NEAR.

Now, the distance between the residences of Marvin and Lucia is expressed by

$$d = \sqrt{(X_M-X_L)^2+(Y_M-Y_L)^2}\ . \tag{5.34}$$

Using (5.34) and applying the extension principle (2.34), the possibility distribution function of d is found to be given by

$$\pi_d(w) = \sup_{u_1,v_1,u_2,v_2}\left(\pi_{(X_M,Y_M)}(u_1,v_1)\wedge\pi_{(X_L,Y_L)}(u_2,v_2)\right) \tag{5.35}$$

subject to

$$w = \sqrt{(u_1-u_2)^2+(v_1-v_2)^2} \tag{5.36}$$

where the supremum is taken over all possible values of X_M, Y_M, X_L and Y_L subject to the constraint (5.36). Generally, π_d as defined by (5.35) will be a monotone decreasing function of w, with $\pi_d(w) = 1$ for sufficiently small values of w.

Example 7 (Example (c), Section 1).

$p_1 \triangleq$ It is unlikely that Andrea is very young
$p_2 \triangleq$ It is likely that Andrea is young
$p_3 \triangleq$ It is very unlikely that Andrea is old

$q \triangleq$ How likely is it that Andrea is not old?

Solution. To find the answer to the posed question, we shall reduce the stated problem to the solution of a mathematical program, as described in Section 4.

First, each of the premises is translated into a constraint on the probability density, p, of Andrea's age. Thus, using (3.8), (5.14) and (3.40), we have

$p_1 \triangleq$ It is unlikely that Andrea is very young $\rightarrow$

$$\pi_1(p) = \mu_{LIKELY}\left(1 - \int_0^{100} \mu^2_{YOUNG}(u)p(u)du\right) \tag{5.37}$$

$$\pi_2(p) = \mu_{LIKELY}\left(\int_0^{100} \mu_{YOUNG}(u)p(u)du\right) \tag{5.38}$$

$$\pi_3(p) = \mu^2_{LIKELY}\left(1 - \int_0^{100} \mu_{OLD}(u)p(u)du\right) \tag{5.39}$$

where $\int_0^{100} \mu_{YOUNG}(u)p(u)du$ represents the probability of the fuzzy event "Andrea is young," with the understanding that the range of the variable Age(Andrea) is the interval [0,100].

Next, we must translate the answer to the posed question, which we assume to be of the form "It is λ that Andrea is not old," where λ is a linguistic probability. Thus

$$q \rightarrow \pi_q(p) = \mu_\lambda\left(\int_0^{100} \left(1 - \mu_{OLD}(u)\right)p(u)du\right) \tag{5.40}$$

where μ_λ is the unknown membership function of λ.

Finally, by using (4.20), the problem in question is reduced to the solution of the variational problem

$$\mu_\lambda(\gamma) \triangleq \max_p\left\{\left[\mu_{LIKELY}\left(1 - \int_0^{100} \mu^2_{YOUNG}(u)p(u)du\right) \wedge \mu_{LIKELY}\left(\int_0^{100} \mu_{YOUNG}(u)p(u)du\right) \wedge \mu^2_{LIKELY}\left(1 - \int_0^{100} \mu_{OLD}(u)p(u)du\right)\right]\right\} \tag{5.41}$$

subject to

$$\gamma = \int_0^{100} \left(1 - \mu_{OLD}(u)\right)p(u)du$$

where γ is the numerical probability of the fuzzy event

"Andrea is not old."

Example 8 (Example (e), Section 1).

Brian is much taller than most of his close friends

How tall is Brian?

Solution. Let x denote Brian's height. In Section 3, we have found that, relative to a given database D, the truth of p is given by

$$\tau = \mu_{MOST}\left(\frac{\sum_i \mu_{MT}(x,\text{Height}(\text{Name}_i)) \wedge \mu^2_{CF}(\text{Brian},\text{Name}_i)}{\sum_i \mu^2_{CF}(\text{Brian},\text{Name}_i)}\right) \quad (5.42)$$

where $\mu_{MT}(x,\text{Height}(\text{Name}_i))$ is the degree to which Brian is much taller than Name_i and μ_F is the degree to which Name_i is Brian's close friend.

Now, for a given value of x and a given D, the value of τ may be interpreted as the possibility of x given D. Thus, the possibility distribution function of Brian's height is given by the same expression as τ, and hence

$$\text{Poss}\{\text{Height}(\text{Brian}) = x\} \quad (5.43)$$
$$= \mu_{MOST}\left(\frac{\sum_i \mu_{MT}(x,\text{Height}(\text{Name}_i)) \wedge \mu^2_{CF}(\text{Brian},\text{Name}_i)}{\sum_i \mu^2_{CF}(\text{Brian},\text{Name}_i)}\right)$$

Example 9. Find the consistency of the proposition

$$p \triangleq \text{Sharon has more than a few good friends}$$

with the database

$$GF_{Sharon} = \text{Mary} + 0.9\text{Valya} + 0.9\text{Doris} + 0.8\text{John} + 0.7\text{Chris} + 0.6\text{Pat} + 0.5\ \text{Denise} + \cdots \quad (5.44)$$

$$FEW = 0.8/1 + 0.9/2 + 1/3 + 1/4 + 0.8/5 + 0.5/6 + 0.2/7 \quad (5.45)$$

where GF_{Sharon} is the fuzzy set of Sharon's good friends (arranged in order of decreasing degree of friendship) and

FEW is the fuzzy denotation of *few*.[6]

Solution. If FEW is defined by (5.45), then *at least few* is expressed by

$$\underline{>}\circ FEW = 0.8/1 + 0.9/2 + 1/3 + 1/4 + \cdots \qquad (5.46)$$

where $\underline{>}\circ FEW$ is the composition of the binary relation $\underline{>}$ with the unary relation FEW.

The FG cardinality of the fuzzy set GF_{Sharon} is given by

$$FGCount(GF_{Sharon}) = 1/1 + 0.9/2 + 0.9/3 + 0.8/4 + 0.7/5 + 0.6/6 + 0.5/7 + \cdots \qquad (5.47)$$

and hence the degree of consistency of p with the database is given by

$$\begin{aligned}\gamma &= \sup(FGCount(GF_{Sharon}) \cap \underline{>}\circ FEW) \qquad (5.48)\\ &= \sup(0.8/1 + 0.9/2 + 0.9/3 + 0.8/4 + \cdots)\\ &= 0.9\end{aligned}$$

[6] A fuller discussion of problems of this type may be found in [11].

6. Evidence, Certainty and Possibility

An important issue that arises in the analysis of soft data relates to the need for a way of assessing the degree of credibility of a conclusion which is inferred from a body of evidence.

For our purposes, it will be convenient to regard a body of evidence--or simply evidence, E--as a collection of fuzzy propositions, $E = \{g_1,\ldots,g_n\}$. Furthermore, we shall assume that the evidence is *granular* in nature, that is, each g_i, $i = 1,\ldots,n$, is a granule of the form[7]

(a) $g_i \triangleq Y \text{ is } G_i \text{ is } \lambda_i$ (6.1)

and/or

(b) $g_i \triangleq \text{If } X \text{ is } F_i \text{ then } Y \text{ is } G_i$ (6.2)

and/or

(c) $g_i \triangleq \text{If } X \text{ is } F_i \text{ then } Y \text{ is } G_j \text{ is } \lambda_j$, $j = 1,\ldots,m$ (6.3)

and/or

(d) $g_i \triangleq X \text{ is } F_i \text{ is } p_i$ (6.4)

where X and Y are variables taking values in U and V, respectively; F_i, $i = 1,\ldots,n$ and G_j, $j = 1,\ldots,m$, are fuzzy subsets of U and V; and p_i and λ_j are linguistic probabilities.

Although E may comprise a mixture of granules of the form (a), (b), (c) and (d), there are two special cases which are typical of the problems encountered in practice. In one, which we shall label Type I, all of the granules in E are of the form (a), and E may be regarded as the conjunction of $g_1,\ldots,g_n$. In the other, all of the granules in E are of the form (b) and (d), and the evidence is said to be of Type II. In the latter case, we shall assume for simplicity that X ranges over a finite set which for convenience may be taken to be the set of integers $\{1,\ldots,n\}$.

[7] A more detailed discussion of the concept of information granularity may be found in [94].

As a simple illustration of evidence of Type I, assume that we are interested in Penny's age and that the available evidence about her age is comprised of the following soft data granules:

(a) $g_1 \triangleq$ Penny is very young is unlikely
$g_2 \triangleq$ Penny is young is very likely
$g_3 \triangleq$ Penny is not young is unlikely

As an illustration of evidence of Type II, we may have, as in Example (f) in Section 1:

(b) $g_1 \triangleq$ If Penny is an undergraduate student, then she is very young
$g_2 \triangleq$ If Penny is a graduate student, then she is young
$g_3 \triangleq$ If Penny is a doctor then she is not very young
$g_4 \triangleq$ Penny is an undergraduate student is unlikely
$g_5 \triangleq$ Penny is a graduate student is likely
$g_6 \triangleq$ Penny is a doctor is not likely

Given a collection of data granules such as those appearing in (a) and (b), we wish to infer from E an answer to a question of the general form:

$$q \triangleq Y \text{ is } Q \text{ is } ?\alpha \qquad (6.5)$$

where Q is a specified fuzzy subset of V and $?\alpha$ is the desired linguistic probability. For example:

$$q \triangleq \text{Penny is not very young is } ?\alpha \qquad (6.6)$$

to which the answer might be, say,

$$?\alpha \triangleq \text{not very likely} .$$

In the case of evidence of Type I, an answer to a question of the form (6.5) may be obtained, in principle, by using the mathematical programming technique employed in Example 7, Section 5. In the case of evidence of Type II, however, we shall use a different approach involving a replacement of the posed question with a *surrogate* question, q_s, that is, a question which, unlike q, may be answerable based on the information contained in E. Such a question in the case of (6.6), for example, might be

$q_s \triangleq$ What is the degree of certainty that Penny is not very young?

or

$q_s \triangleq$ What is the degree of possibility that Penny is not very young?

The approach described in the sequel is based on a generalization of the concepts of upper and lower probabilities [17,29] which serve as a point of departure for Shafer's theory of evidence [67]. Viewed from the perspective of our approach, the latter theory is concerned with the special case where (a) the evidence is of Type II; (b) the G_i and Q are nonfuzzy sets; and (c) the p_i are numerical probabilities.

Assuming, first, that the G_i are fuzzy sets but the p_i are numerical probabilities, we define the *conditional possibility* and the *conditional certainty* of the proposition "Y is Q" (or, equivalently, the event "Y is Q") given that "Y is G_i" as follows:

$$\mathrm{Poss}\{Y \text{ is } Q \mid Y \text{ is } G_i\} = \sup(Q \cap G_i) \tag{6.7}$$

$$\mathrm{Cert}\{Y \text{ is } Q \mid Y \text{ is } G_i\} = \inf(G_i \Rightarrow Q) \tag{6.8}$$

where

$$\sup(Q \cap G_i) = \sup_v(\mu_Q(v) \wedge \mu_{G_i}(v)), \quad v \in V \tag{6.9}$$

$$\inf(G_i \Rightarrow Q) = \inf_v((1 - \mu_{G_i}(v)) \vee \mu_Q(v)) \tag{6.10}$$

and μ_Q and μ_{G_i} are the membership functions of Q and G_i, respectively.

In effect, the right-hand members of (6.7) and (6.8) serve as measures of the degree to which the proposition "Y is G_i" influences one's belief in the proposition "Y is Q."

In particular, (6.7) serves as a measure of the degree of possibility while (6.8) plays the same role in relation to the degree of certainty. Note that when Q and G_i are nonfuzzy, we have

$$\begin{aligned} \sup(Q \cap G_i) &= 1 \text{ if } Q \cap G_i \text{ is nonempty} \\ &= 0 \text{ if } Q \cap G_i = \theta \end{aligned} \tag{6.11}$$

and

$$\begin{aligned} \inf(G_i \Rightarrow Q) &= 1 \text{ if } G_i \subset Q \\ &= 0 \text{ otherwise} \end{aligned} \tag{6.12}$$

Now since X is assumed to be a random variable which takes the values 1,...,n with respective probabilities $p_1,\ldots,p_n$, the conditional possibility and conditional certainty of the proposition "Y is Q" are also random variables whose respective expectations are given by

$$\begin{aligned} E\Pi(Q) &= \textstyle\sum_i p_i \sup(Q \cap G_i) \\ &= \textstyle\sum_i p_i \sup_v(\mu_Q(v) \wedge \mu_{G_i}(v)) \end{aligned} \tag{6.13}$$

$$\begin{aligned} EC(Q) &= \textstyle\sum_i p_i \inf(G_i \Rightarrow Q) \\ &= \textstyle\sum_i p_i \inf_v((1-\mu_{G_i}(v)) \vee \mu_Q(v)) \\ &= 1 - E\Pi(Q') \end{aligned} \tag{6.14}$$

We shall refer to $E\Pi(Q)$ and $EC(Q)$ as the *expected possibility* and the *expected certainty*, respectively, of the proposition "Y is Q." When Q and $G_1,\ldots,G_n$ are nonfuzzy, EC(Q) and $E\Pi(Q)$ reduce to the Shafer's *degree of belief* and *degree of plausibility*, respectively, which correspond to the lower and upper probabilities in Dempster's work [17].[8] Our feeling is that Shafer's identification of "degree of belief" with the lower rather than the upper probability (or, more generally, with E(Q) rather than $E\Pi(Q)$) is open to question,

[8] It should be remarked that EC(Q) and $E\Pi(Q)$ are not normalized--as are the lower and upper probabilities in the work of Dempster and Shafer. As is pointed out in [95], the normalization in question leads to counterintuitive results in application to combination of bodies of evidence.

since there is no particular reason for singling out EC(Q) or EΠ(Q) or, for that matter, any convex combination of them as a universal measure of the degree of belief.

Having defined the concepts of expected certainty and expected possibility, we are in a position to see the rationale for employing the technique of surrogate questions in the case of evidence of Type II. Taking for simplicity the case where the G_i and Q are nonfuzzy and the p_i are numerical probabilities, the evidence can be expressed in the form

$$g_1 \triangleq Y \in G_1 \text{ or } g_2 \triangleq Y \in G_2 \text{ or } \cdots \text{ or } g_n \triangleq Y \in G_n$$

$$\text{Prob}\{g_1\} = p_1 \text{ and } \text{Prob}\{g_2\} = p_2 \text{ and } \cdots \text{ and } \text{Prob}\{g_n\} = p_n$$

Now let us assume that the original question is: What is the numerical probability that $Y \in Q$? It is easy to see that the granularity of available evidence makes it infeasible to answer questions of this type for arbitrary Q. Thus, we are compelled to replace the original unanswerable question with a surrogate answerable question which in some sense is close to the original question. In the case under discussion, such questions would be:

(a) What is the expected certainty (or, equivalently, the degree of belief (Shafer) or the lower probability (Dempster)) that $Y \in Q$?

(b) What is the expected possibility (or, equivalently, the degree of plausibility (Shafer) or the upper probability (Dempster)) that $Y \in Q$?

Based on the available evidence, the answers to (a) and (b) are:

$$EC(Q) = \sum_i p_i \inf(G_i \Rightarrow Q) \quad , \quad i = 1,\ldots,n$$

and

$$E\Pi(Q) = \sum_i p_i \sup(G_i \cap Q)$$

where (see (6.11) and (6.12))

$$\inf(G_i \Rightarrow Q) = 1 \text{ if } G_i \subset Q$$
$$= 0 \text{ otherwise}$$

and

$$\sup(G_i \cap Q) = 1 \text{ if } G_i \cap Q = \theta$$
$$= 0 \text{ otherwise}$$

A serious shortcoming of the Shafer-Dempster approach is that if G_i and Q are nonfuzzy and the condition

$$G_i \subset Q$$

is not satisfied exactly, then no matter how small the error might be the contribution of the term $p_i \inf(G_i \Rightarrow Q)$ to the value of EC(Q) in the summation

$$EC(Q) = \sum_i p_i \inf(G_i \Rightarrow Q)$$

would be zero. In intuitive terms, what this means is that a piece of evidence will be disregarded so long as there is the slightest doubt about its perfect validity. We avoid this extreme degree of conservatism in our approach by (a) allowing the G_i and Q to be fuzzy; and (b) fuzzifying the concept of containment, with the expression $\inf(G_i \Rightarrow Q)$ in (6.14) representing, in effect, the degree to which G_i is contained in Q. Thus, if G is regarded as a random variable which takes the values $G_1,\ldots,G_n$ with respective probabilities $p_1,\ldots,p_n$, then we can write

$$EC(Q) = \text{Prob}\{G \subset Q\} \qquad (6.15)$$

with the understanding that $G \subset Q$ is a fuzzy event [86] and that the degree to which $G \subset Q$ is satisfied is expressed by

$$\text{degree}\{G \subset Q\} = \inf(G \Rightarrow Q) \ .$$

Viewed in this perspective, (6.15) may be regarded as a natural generalization of Dempster's lower probability and Shafer's degree of belief.

For the purpose of illustration, we shall conclude this section by describing the application of our approach to

Example (b). In this example, the G_i and Q are fuzzy and the p_i are linguistic probabilities. More specifically, we have

$$G_1 \triangleq YOUNG^2$$
$$G_2 \triangleq YOUNG$$
$$G_2 \triangleq (YOUNG^2)'$$
$$Q \triangleq YOUNG'$$
$$p_1 \triangleq UNLIKELY = ANT\ LIKELY$$
$$p_2 \triangleq LIKELY$$
$$p_3 \triangleq (LIKELY)'$$

where YOUNG is the denotation of *young*, $YOUNG^2$ is the denotation of *very young*, ANT is the antonym, i.e., (see (3.40))

$$\mu_{ANT\ LIKELY}(v) = \mu_{LIKELY}(1-v) \ , \quad v \in [0,1] \quad (6.16)$$

and the prime represents the complement.

Now let

$$\alpha_1 = \sup(YOUNG^2 \cap YOUNG') \quad (6.17)$$

$$\alpha_2 = \sup(YOUNG \cap YOUNG') \quad (6.18)$$

$$\alpha_3 = \sup((YOUNG^2)' \cap YOUNG') \quad (6.19)$$

where the α_i are numbers in the interval [0,1]. (From (6.18) it follows that $\alpha_2 = 0.5$ but we shall not make use of this fact.) Then, using (6.13) we can express $E\Pi(Q)$ as

$$E\Pi(Q) = \alpha_1 UNLIKELY \oplus \alpha_2 LIKELY \oplus \alpha_3 LIKELY' \quad (6.20)$$

where $\oplus$ denotes the sum of fuzzy numbers [90,50,18].

To compute $E\Pi(Q)$ as a fuzzy number, we have to take into consideration the fact that the numerical probabilities must sum up to unity. Thus, on denoting these probabilities by v_1, v_2, and v_3, and applying the extension principle (4.20), the determination of the membership function of $E\Pi(Q)$ is reduced to the solution of the following variational problem:

$$\mu(z) \triangleq \max_{v_1,v_2,v_3}\left(\mu_{LIKELY}\left(\frac{1-v_1}{\alpha_1}\right) \wedge \mu_{LIKELY}\left(\frac{v_2}{\alpha_2}\right) \wedge \left(1-\mu_{LIKELY}\left(\frac{v_3}{\alpha_3}\right)\right)\right) \qquad (6.21)$$

subject to

$$z = \alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3 \qquad (6.22)$$

$$1 = v_1 + v_2 + v_3$$

Thus, expressed as a fuzzy set, we have

$$E\Pi(Q) = \int_{[0,1]} \mu(z)/z \qquad (6.23)$$

where $\mu(z)$ is given by (6.21). To compute EC(Q), then, we can make use of the identity (6.14)

$$EC(Q) = 1 - E\Pi(Q') \qquad (6.24)$$

From our definitions of $E\Pi(Q)$ and EC(Q) it is a simple matter to derive a basic rule of conditioning which may be regarded as a generalization of those given by Dempster and Shafer. Specifically, assume that the evidence has the form:

If $X = i$ then Y is G_i , $i = 1,\ldots,n$
$\text{Prob}\{X=i\} = p_i$

and, in addition, we know that

$$g_0 \triangleq Y \text{ is } G_0$$

where G_0 is a given fuzzy subset of V.

Clearly, the available evidence may be expressed in the equivalent form:

If $X = i$ then Y is $G_i \cap G_0$, $i = 1,\ldots,n$
$\text{Prob}\{X=i\} = p_i$

which implies that

$$E\Pi(Q) \text{ conditioned on "Y is } G_0\text{"} = E\Pi(Q \cap G_0) \qquad (6.25)$$

and correspondingly

$$EC(Q) \text{ conditioned on "Y is } G_0\text{"} = 1 - E\Pi(Q' \cup G_0') \qquad (6.26)$$

Remark. The connection between the definition of expected possibility--as expressed by (6.13)--with that of the upper probability in [17] and [67]--may be made more transparent by interpreting $E\Pi(Q)$ as the probability of a fuzzy event--in the manner of (6.15). More specifically, if $\sup(G \cap Q)$ is regarded as the degree of occurrence of the fuzzy event $G \cap Q?$, in which the question mark serves to signify that we are concerned with the degree to which G intersects Q rather than with the intersection of G and Q, then we can write

$$E\Pi(Q) = \text{Prob}\{G \cap Q?\} \qquad (6.27)$$

with the understanding that G is a random variable which takes the values $G_1, \ldots, G_n$ with respective probabilities $p_1, \ldots, p_n$.

In summary, then, the expected possibility and expected certainty may be expressed in the form

$$E\Pi(Q) = \text{Prob}\{G \cap Q?\} \qquad (6.28)$$

and

$$EC(Q) = \text{Prob}\{G \subset Q\} \qquad (6.29)$$

which clarifies the sense in which $E\Pi(Q)$ and $EC(Q)$ may be viewed, respectively, as generalizations of the concepts of upper and lower probabilities--concepts which are defined in [17] and [67] under the assumption that the G_i and Q are nonfuzzy sets.

7. Concluding Remark

The approach to the analysis of soft data described in this paper represents a substantive departure from the conventional probability-based methods.

The main thesis underlying our approach is that, in general, the uncertainty which is intrinsic in soft data is a mixture of probabilistic and possibilistic constituents and, as such, must be dealt with by a combination of probabilistic and possibilistic methods. We have indicated, in general terms, how this can be done through the use of the concept of a possibility distribution and the related concepts of a linguistic variable, semantic entailment, semantic equivalence, and the extension principle. Finally, we have shown how the concepts of expected possibility and expected certainty relate to the important issue of credibility analysis, and indicated a way of reducing many of the problems in inference from soft data to the solution of nonlinear programs.

The issues associated with soft data analysis are varied and complex. Clearly, we have--at this juncture--only a partial understanding of the basic problem of inference from soft data and the associated problem of credibility assessment. What is likely, however, is that, in the years to come, our understanding of these and related problems will be enhanced through a further development of possibility-based methods for the representation and manipulation of soft data.

8. References and Related Papers

1. B. Adams, *Math. Biosci.* 32, 177 (1976).
2. J. F. Baldwin, Report EM/FS 10, University of Bristol (1978).
3. J. F. Baldwin, N. C. F. Guild and B. W. Pilsworth, Report EM/FS 5, University of Bristol (1978).
4. J. F. Baldwin and B. W. Pilsworth, Report EM/FS 9, University of Bristol (1978).
5. W. Bandler and L. Kohout, *Proc. 9th Int. Symp. on Multiple-Valued Logic*, 200 (1979).
6. G. Banon, *Proc. Int. Colloq. Theory Appl. of Fuzzy Sets* (1978).
7. R. E. Bellman and L. A. Zadeh, in *Modern Uses of Multiple-Valued Logic*, J. M. Dunn, G. Epstein, Eds. (D. Reidel, Dordrecht-Holland, 1977), p. 103.
8. M. Black, *Dialogue* 2, 1 (1963).
9. D. Bobrow and A. Collins, *Representation and Understanding* (Academic Press, New York, 1975).
10. B. Bouchon, to appear in *Fuzzy Sets and Systems*, 1979.
11. B. Cerny, Doctoral Dissertation, Dept. of Educ. Psych., Univ. of Calif., Berkeley, 1978.
12. L. J. Cohen, *The Implications of Induction* (Methuen, London, 1970).
13. A. Collins, *Proc. TINLAP-2*, 194 (1978).
14. A. Collins, E. Warnock, N. Aiello and M. Miller, in *Representation and Understanding*, D. Bobrow, A. Collins, Eds. (Academic Press, New York, 1975).
15. M. J. Cresswell, *Logics and Languages* (Methuen, London, 1973).
16. D. Davidson, *Synthese* 17, 304 (1967).
17. A. P. Dempster, *Ann Math. Statist.* 38, 325 (1967).
18. D. Dubois and H. Prade, *Int. J. System Sci.* 9, 613 (1979).
19. D. Dubois and H. Prade, *Fuzzy Sets and Systems* (Academic Press, New York, 1980).
20. R. Duda et al., Ann. Report SRI International, Nos. 5821, 6415 (1977).
21. K. Fine, *Synthese* 30, 265 (1975).
22. B. R. Gaines, *Int. J. Man-Machine Studies* 6, 623 (1975).
23. B. R. Gaines, *Proc. 6th Int. Symp. on Multiple-Valued Logic* IEEE76CH 1111-4C, 179 (1976).
24. B. R. Gaines and L. J. Kohout, *Proc. Int. Symp. on Multiple-Valued Logic*, 183 (1975).
25. B. R. Gaines and L. J. Kohout, *Int. J. Man-Machine Studies* 9, 1 (1977).
26. R. Giles, *Int. J. Man-Machine Studies* 8, 313 (1976).
27. J. A. Goguen, *Synthese* 19, 325 (1969).
28. J. A. Goguen, *Int. J. Man-Machine Studies* 6, 513 (1974).

29. I. J. Good, in *Logic, Methodology and Philosophy of Science: Proceedings of the 1960 International Congress*, E. Nagel, P. Suppes, A. Tarski, Eds. (Stanford University Press, Stanford, 1962).
30. S. Haack, *Philosophy of Logics* (Cambridge University Press, Cambridge, 1978).
31. I. Hacking, *The Emergence of Probability* (Cambridge University Press, Cambridge, 1975).
32. H. M. Hersh and A. Caramazza, *J. Exp. Psych.* 105, 254 (1976).
33. E. Hisdal, *Fuzzy Sets and Systems* 1, 283 (1978).
34. G. E. Hughes and M. J. Cresswell, *An Introduction to Modal Logic* (Methuen, London, 1968).
35. E. T. Jaynes, *Phys. Rev.* 106, 620 (1957); 108, 171 (1957).
36. A. Joshi, in *Pattern Directed Inference*, F. Hays-Roth, D. Waterman, Eds. (Academic Press, New York, 1978).
37. J. Kampé de Feriet and B. Forte, *Comptes Rendus Acad. Sci.* A-265, 110, 142, 350 (1967).
38. A. Kaufmann, *Introduction to the Theory of Fuzzy Subsets* (Masson and Co., Paris, 1975).
39. W. J. M. Kickert and H. R. van Nauta Lemke, *Automatica* 12, 301 (1976).
40. S. Kullback, *Information Theory and Statistics* (John Wiley, New York, 1959).
41. H. E. Kyburg, *Synthese* 36, 87 (1977).
42. G. Lakoff, *J. Phil. Logic* 2, 458 (1973). Also in *Contemporary Research in Philosophical Logic and Linguistic Semantics*, D. Hockney, W. Harper, B. Freed, Eds. (D. Reidel, The Netherlands, 1973), p. 221.
43. R. A. LeFaivre, *J. Cybernetics* 4, 57 (1974).
44. R. C. Lyndon, *Notes on Logic* (D. van Nostrand, New York, 1966).
45. M. Mamdani, in *Fuzzy Automata and Decision Processes*, M. M. Gupta, G. N. Saridis, B. R. Gaines, Eds. (Elsevier North-Holland, New York, 1977).
46. E. H. Mamdani and S. Assilian, *Int. J. Man-Machine Studies* 7, 1 (1975).
47. G. A. Miller and P. N. Johnson-Laird, *Language and Perception* (Harvard University Press, Cambridge, 1976).
48. M. Minsky, in *The Psychology of Computer Vision*, P. Winston, Ed. (McGraw-Hill, New York, 1975), p. 211.
49. M. Mizumoto, S. Fukame and K. Tanaka, *Proc. 3rd Workshop on Fuzzy Reasoning* (1978).
50. M. Mizumoto and K. Tanaka, *Proc. Int. Conf. on Cybernetics and Society* (1976).
51. G. C. Moisil, Lectures on the logic of fuzzy reasoning, *Scientific Editions* (Bucharest, 1975).
52. R. E. Moore, *Interval Analysis* (Prentice-Hall, Englewood Cliffs, N.J., 1966).
53. S. Nahmias, *Fuzzy Sets and Systems* 1, 97 (1978).

54. C. V. Negoita and D. A. Ralescu, *Applications of Fuzzy Sets to Systems Analysis* (Birkhauser Verlag, Basel, 1975).
55. A. Newell and H. A. Simon, *Human Problem Solving* (Prentice-Hall, Englewood Cliffs, N. J., 1972).
56. H. T. Nguyen, *J. Math. Anal. Appl.* 65, 531 (1978).
57. H. T. Nguyen, *Fuzzy Sets and Systems* 1, 283 (1978).
58. J. J. Ostergaard, Pub. 7601, Technical University of Denmark, Lyngby (1976).
59. T. J. Procyk, Fuzzy Logic Working Group, Queen Mary College, London (1976).
60. N. Rescher, *Plausible Reasoning* (Van Gorcum, Amsterdam, 1976).
61. C. Rieger, *Proc. Georgetown University Linguistics Roundtable* (1976).
62. R. D. Rosenkrantz, *Inference, Method and Decision* (D. Reidel, Dordrecht, 1977).
63. E. Sanchez, Fuzzy relations, University of Marseille (1974).
64. E. Sanchez, M77/28, Univ. of Calif. Berkeley (1977).
65. R. C. Schank, *Conceptual Information Processing* (Academic Press, New York, 1975).
66. B. S. Sembi and E. H. Mamdani, *Proc. 9th Int. Symp. on Multiple-Valued Logic*, 143 (1979).
67. G. Shafer, *A Mathematical Theory of Evidence* (Princeton University Press, Princeton, 1976).
68. J. E. Shore and R. W. Johnson, NRL Memo 3-898, Naval Research Laboratory (1978).
69. E. Shortliffe, *MYCIN: Computer-Based Medical Consultations* (American Elsevier, New York, 1976).
70. E. H. Shortliffe and B. G. Buchanan, *Math. Biosci.* 23, 351 (1975).
71. P. Smets, Un modele mathematico-statistique simulant le processus du diagnostic medicale, Free University of Brussels (1978).
72. M. Sugeno, Ph.D. Dissertation, Tokyo Institute of Technology (1974).
73. P. Suppes, *J. Roy. Statist. Soc. B* 36, 160 (1974).
74. P. Suppes and M. Zanotti, *Synthese* 36, 427 (1977).
75. R. G. Swinburne, *An Introduction to Confirmation Theory* (Methuen, London, 1973).
76. P. Szolovits, L. B. Hawkinson and W. A. Martin, LCS Memo, M.I.T. (1977).
77. A. Tarksi, *Logic, Semantics, Metamathematics* (Clarendon, Press, London, 1956).
78. T. Terano and M. Sugeno, in *Fuzzy Sets and Their Applications to Cognitive and Decision Processes*, L. A. Zadeh, K. S. Fu, K. Tanaka, M. Shimura, Eds. (Academic Press, New York, 1975), p. 151.
79. R. M. Tong, *Int. J. General Systems* 4, 143 (1978).
80. M. Tribus, *Rational Descriptions, Decisions and Designs* (Pergamon Press, New York, 1969).

81. A. Tversky and D. Kahneman, *Science* 185, 1124 (1974).
82. S. Watanabe, *Information and Control* 15, 1 (1969).
83. H. Wechsler, *Proc. Int. Symp. on Multiple-Valued Logic*, 162 (1975).
84. F. Wenstop, *Int. J. Man-Machine Studies* 8, 293 (1976).
85. P. M. Williams, Bayesian conditionalization and the principle of minimum information, University of Sussex (1978).
86. L. A. Zadeh, *J. Math. Anal. Appl.* 23, 421 (1968).
87. L. A. Zadeh, *J. Cybernetics* 2, 4 (1972).
88. L. A. Zadeh, *IEEE Trans. Systems, Man and Cybernetics* SMC-3, 28 (1973).
89. L. A. Zadeh, *Synthese* 30, 407 (1975).
90. L. A. Zadeh, *Information Sciences* 8, 199 (1975); 8, 301 (1975); 9, 43 (1975).
91. L. A. Zadeh, Memo M77/58, Univ. of Calif. Berkeley (1977).
92. L. A. Zadeh, *Fuzzy Sets and Systems* 1, 3 (1978).
93. L. A. Zadeh, *Int. J. Man-Machine Studies* 10, 395 (1978).
94. L. A. Zadeh, in *Advances in Fuzzy Sets Theory and Applications*, M. Gupta, R. Ragade, R. Yager, Eds. (North-Holland, Amsterdam, 1979).
95. L. A. Zadeh, M79/24, Univ. of Calif. Berkeley (1979).
96. L. A. Zadeh, K. S. Fu, K. Tanaka and M. Shimura, *Fuzzy Sets and Their Application to Cognitive and Decision Processes* (Academic Press, New York, 1975).
97. L. A. Zadeh, *Int. J. Man-Machine Studies* 8, 249 (1976).
98. L. A. Zadeh, M79/32, Univ. of Calif. Berkeley (1979).