FUZZY SETS AND INFORMATION GRANULARITY

by

L. A. Zadeh

Memorandum No. UCB/ERL M79/45

19 July 1979

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

FUZZY SETS AND INFORMATION GRANULARITY*

L.A. Zadeh

Computer Science Division
Department of Electrical Engineering and Computer Sciences
and the Electronics Research Laboratory
University of California at Berkeley

1. Introduction

Much of the universality, elegance and power of classical mathematics derives from the assumption that real numbers can be characterized and manipulated with infinite precision. Indeed, without this assumption, it would be much less simple to define what is meant by the zero of a function, the rank of a matrix, the linearity of a transformation or the stationarity of a stochastic process.

It is well-understood, of course, that in most real-world applications the effectiveness of mathematical concepts rests on their robustness, which in turn is dependent on the underlying continuity of functional dependencies [1]. Thus, although no physical system is linear in the idealized sense of the term, it may be regarded as such as an approximation. Similarly, the concept of a normal distribution has an operational meaning only in an approximate and, for that matter, not very well-defined sense.

There are many situations, however, in which the finiteness of the resolving power of measuring or information gathering devices cannot be dealt with through an appeal to continuity. In such cases, the information may be said to be <u>granular</u> in the sense that the data points within a granule have to be dealt with as a whole rather than individually.

Taken in its broad sense, the concept of information granularity occurs under various guises in a wide variety of fields. In particular, it bears a close relation to the concept of aggregation in economics; to decomposition and partition--in the theory of automata and system theory; to bounded uncertainties--in optimal control [2], [3]; to locking granularity--in the analysis of concurrencies in data base management systems [4]; and to the manipulation of numbers as intervals--as in interval analysis [5]. In the present paper, however, the concept of information granularity is employed in a stricter and somewhat narrower sense which is defined in greater detail in Sec. 2. In effect, the main motivation for our approach is to define the concept of information granularity in a way that relates it to the theories of evidence of Shafer [6], Dempster [7], Smets [8], Cohen [9], Shackle [10] and others, and provides a basis for the construction of more general theories in which the evidence is allowed to be fuzzy in nature.

More specifically, we shall concern ourselves with a type of information granularity in which the data granules are characterized by propositions of the general form

$$g \triangleq X \text{ is } G \text{ is } \lambda \tag{1.1}$$

---

*To Professor J. Kampe de Feriet.

Research supported by Naval Electronic Systems Command Contract N00039-78-G0013 and National Science Foundation Grant ENG-78-23143.

To appear in: <u>Advances in Fuzzy Sets Theory and Applications</u>, M. Gupta, R. Ragade and R. Yager, eds., North-Holland Publishing Co., 1979.

in which X is a variable taking values in a universe of discourse U, G is a fuzzy subset of U which is characterized by its membership function μ_G, and the qualifier λ denotes a fuzzy probability (or likelihood). Typically, but not universally, we shall assume that U is the real line (or R^n), G is a convex fuzzy subset of U and λ is a fuzzy subset of the unit interval. For example:

$g \triangleq$ X is small is likely

$g \triangleq$ X is not very large is very unlikely

$g \triangleq$ X is much larger than Y is unlikely

We shall not consider data granules which are characterized by propositions in which the qualifier λ is a fuzzy possibility or fuzzy truth-value.

In a general sense, a <u>body of evidence</u> or, simply, <u>evidence</u>, may be regarded as a collection of propositions. In particular, the evidence is <u>granular</u> if it consists of a collection of propositions,

$$E = \{g_1,\ldots,g_N\} , \tag{1.2}$$

each of which is of the form (1.1). Viewed in this perspective, Shafer's theory relates to the case where the constituent granules in (1.2) are crisp (nonfuzzy) in the sense that, in each g_i, G_i is a nonfuzzy set and λ_i is a numerical probability, implying that g_i may be expressed as

$$g_i \triangleq \text{"Prob}\{X \in G_i\} = p_i\text{"} \tag{1.3}$$

where p_i, $i = 1,\ldots,N$, is the probability that the value of X is contained in G. In the theories of Cohen and Shackle, a further restriction is introduced through the assumption that the G_i are nested, i.e., $G_1 \subset G_2 \subset \cdots \subset G_N$. As was demonstrated by Suppes and Zanotti [11] and Nguyen [12], in the analysis of evidence of the form (1.3) it is advantageous to treat E as a random relation.

Given a collection of granular bodies of evidence $E = \{E_1,\ldots,E_K\}$, one may ask a variety of questions the answers to which are dependent on the data resident in E. The most basic of these questions--which will be the main focus of our attention in the sequel--is the following:

Given a body of evidence $E = \{g_1,\ldots,g_N\}$ and an arbitrary fuzzy subset Q of U, what is the probability--which may be fuzzy or nonfuzzy--that X is Q? In other words, from the propositions

$$\begin{array}{l} g_1 \triangleq X \text{ is } G_1 \text{ is } \lambda_1 \\ - - - - - - - - \\ g_N \triangleq X \text{ is } G_N \text{ is } \lambda_N \end{array} \tag{1.4}$$

we wish to deduce the value of $?\lambda$ in the question

$$q \triangleq X \text{ is } Q \text{ is } ?\lambda \tag{1.5}$$

As a concrete illustration, suppose that we have the following granular information concerning the age of Judy ($X \triangleq$ Age(Judy))

$$\begin{array}{l} g_1 \triangleq \text{Judy is very young is unlikely} \\ g_2 \triangleq \text{Judy is young is likely} \\ g_3 \triangleq \text{Judy is old is very unlikely} \end{array} \tag{1.6}$$

The question is: What is the probability that Judy is not very young; or, equivalently: What is the value of $?\lambda$ in

$$q \triangleq \text{Judy is not very young is } ?\lambda$$

In cases where E consists of two or more distinct bodies of evidence, an important issue relates to the manner in which the answer to (1.5)--based on the information resident in E --may be composed from the answers based on the information resident in each of the constituent bodies of evidence $E_1,\ldots,E_K$. We shall consider this issue very briefly in Sec. 3.

In the theories of Dempster and Shafer, both the evidence and the set Q in (1.5) are assumed to be crisp, and the question that is asked is: What are the bounds on the probability λ that $X \in Q$? The lower bound, λ_*, is referred to as the <u>lower probability</u> and is defined by Shafer to be the <u>degree of belief</u> that $X \in Q$, while the upper bound, λ^*, is equated to the <u>degree of plausibility</u> of the proposition $X \in Q$. An extension of the concepts of lower and upper probabilities to the more general case of fuzzy granules will be described in Sec. 3.

As will be seen in the sequel, the theory of fuzzy sets and, in particular, the theory of possibility, provides a convenient conceptual framework for dealing with information granularity in a general setting. Viewed in such a setting, the concept of information granularity assumes an important role in the analysis of imprecise evidence and thus may aid in contributing to a better understanding of the complex issues arising in credibility analysis, model validation and, more generally, those problem areas in which the information needed for a decision or system performance evaluation is incomplete or unreliable.

2. Information Granularity and Possibility Distributions

Since the concept of information granularity bears a close relation to that of a possibility distribution, we shall begin our exposition with a brief review of those properties of possibility distributions which are of direct relevance to the concepts introduced in the following sections.

Let X be a variable taking values in U, with a generic value of X denoted by u. Informally, a possibility distribution, Π_X, is a fuzzy relation in U which acts as an elastic constraint on the values that may be assumed by X. Thus, if π_X is the membership function of Π_X, we have

$$\text{Poss}\{X = u\} = \pi_X(u) \ , \quad u \in U \tag{2.1}$$

where the left-hand member denotes the possibility that X may take the value u and $\pi_X(u)$ is the grade of membership of u in Π_X. When used to characterize Π_X, the function π_X: $U \rightarrow [0,1]$ is referred to as a <u>possibility distribution function</u>.

A possibility distribution, Π_X, may be induced by physical constraints or, alternatively, it may be epistemic in nature, in which case Π_X is induced by a collection of propositions--as described at a later point in this section.

A simple example of a possibility distribution which is induced by a physical constraint is the number of tennis balls that can be placed in a metal box. In this case, X is the number in question and $\pi_X(u)$ is a measure of the degree of ease (by some specified mechanical criterion) with which u balls can be squeezed into the box.

As a simple illustration of an epistemic possibility distribution, let X be a real-valued variable and let p be the proposition

$$p \triangleq a \leq X \leq b$$

where [a,b] is an interval in R^1.[1] In this case, the possibility distribution

[1] A more detailed discussion of possibility theory may be found in [13]-[15].

induced by p is the uniform distribution defined by

$$\pi_X(u) = 1 \quad \text{for} \quad a \le u \le b$$
$$= 0 \quad \text{elsewhere.}$$

Thus, given p we can assert that

$$\mathrm{Poss}\{X = u\} = 1 \quad \text{for u in } [a,b]$$
$$= 0 \quad \text{elsewhere.}$$

More generally, as shown in [16], a proposition of the form

$$p \triangleq N \text{ is } F \tag{2.2}$$

where F is a fuzzy subset of the cartesian product $U = U_1 \times \cdots \times U_n$ and N is the name of a variable, a proposition or an object, induces a possibility distribution defined by the <u>possibility assignment equation</u>

$$N \text{ is } F \rightarrow \Pi_{(X_1,\ldots,X_n)} = F \tag{2.3}$$

where the symbol $\rightarrow$ stands for "translates into," and $X \triangleq (X_1,\ldots,X_n)$ is an n-ary variable which is implicit or explicit in p. For example,

(a) $$X \text{ is small} \rightarrow \Pi_X = \mathrm{SMALL} \tag{2.4}$$

where SMALL, the denotation of <u>small</u>, is a specified fuzzy subset of $[0,\infty)$. Thus, if the membership function of SMALL is expressed as μ_{SMALL}, then (2.4) implies that

$$\mathrm{Poss}\{X = u\} = \mu_{\mathrm{SMALL}}(u) \ , \quad u \in [0,\infty) \ . \tag{2.5}$$

More particularly, if--in the usual notation--

$$\mathrm{SMALL} = 1/0 + 1/1 + 0.8/2 + 0.6/3 + 0.5/4 + 0.3/5 + 0.1/6 \tag{2.6}$$

then

$$\mathrm{Poss}\{X = 3\} = 0.6$$

and likewise for other values of u.

Similarly,

(b) $$\text{Dan is tall} \rightarrow \Pi_{\mathrm{Height(Dan)}} = \mathrm{TALL} \tag{2.7}$$

where the variable Height(Dan) is implicit in the proposition "Dan is tall" and TALL is a fuzzy subset of the interval [0,220] (with the height assumed to be expressed in centimeters).

(c) $$\text{John is big} \rightarrow \Pi_{(\mathrm{Height(John)},\mathrm{Weight(John)})} = \mathrm{BIG} \tag{2.8}$$

where BIG is a fuzzy binary relation in the product space $[0,220] \times [0,150]$ (with height and weight expressed in centimeters and kilograms, respectively) and the variables $X_1 \triangleq$ Height(John), $X_2 \triangleq$ Weight(John) are implicit in the proposition "John is big."

In a more general way, the translation rules associated with the meaning representation language PRUF [16] provide a system for computing the possibility distributions induced by various types of propositions. For example

$$X \text{ is not very small} \rightarrow \Pi_X = (\mathrm{SMALL}^2)' \tag{2.9}$$

where SMALL^2 is defined by

$$\mu_{\mathrm{SMALL}^2} = (\mu_{\mathrm{SMALL}})^2 \tag{2.10}$$

and ' denotes the complement. Thus, (2.10) implies that the possibility distribution function of X is given by

$$\pi_X(u) = 1 - \mu_{\mathrm{SMALL}}^2(u) \ . \tag{2.11}$$

In the case of conditional propositions of the form $p \triangleq$ If X is F then Y is G, the possibility distribution that is induced by p is a <u>conditional possibility distribution</u> which is defined by[2]

$$\text{If X is F then Y is G} \rightarrow \Pi_{(Y|X)} = \bar{F}' \cup \bar{G} \tag{2.12}$$

where $\Pi_{(Y|X)}$ denotes the conditional possibility distribution of Y given X, F and G are fuzzy subsets of U and V, respectively, $\bar{F}$ and $\bar{G}$ are the cylindrical extensions of F and G in $U \times V$, $\cup$ is the union, and the conditional possibility distribution function of Y given X is expressed by

$$\pi_{(Y|X)}(v|u) = (1 - \mu_F(u)) \vee \mu_G(v) \ , \quad u \in U, \ v \in V \tag{2.13}$$

where μ_F and μ_G are the membership functions of F and G, and $\vee \doteq \max$. In connection with (2.12), it should be noted that

$$\pi_{(Y|X)}(v|u) = \mathrm{Poss}\{Y = v | X = u\} \tag{2.14}$$

whereas

$$\pi_{(X,Y)}(u,v) = \mathrm{Poss}\{X = u, \ Y = v\} \ . \tag{2.15}$$

A concept which is related to that of a conditional possibility distribution is the concept of a <u>conditional possibility measure</u> [13]. Specifically, let Π_X be the possibility distribution induced by the proposition

$$p \triangleq X \text{ is } G \ ,$$

and let F be a fuzzy subset of U. Then, the <u>conditional possibility measure</u> of F with respect to the possibility distribution Π_X is defined by

$$\mathrm{Poss}\{X \text{ is } F | X \text{ is } G\} = \sup_u (\mu_F(u) \wedge \mu_G(u)) \ . \tag{2.16}$$

It should be noted that the left-hand member of (2.16) is a set function whereas $\Pi_{(Y|X)}$ is a fuzzy relation defined by (2.12).

The foregoing discussion provides us with the necessary background for defining some of the basic concepts relating to information granularity. We begin with the concept of a <u>fuzzy granule</u>.

<u>Definition</u>. Let X be a variable taking values in U and let G be a fuzzy subset of U. (Usually, but not universally, $U = R^n$ and G is a convex fuzzy subset of U.) A <u>fuzzy granule</u>, g, in U is induced (or characterized) by a proposition of the form

$$g \triangleq X \text{ is } G \text{ is } \lambda \tag{2.17}$$

---

[2] There are a number of alternative ways in which $\Pi_{(Y|X)}$ may be defined in terms of F and G [17], [18], [19]. Here we use a definition which is consistent with the relation between the extended concepts of upper and lower probabilities as described in Sec. 3.

where λ is a fuzzy probability which is characterized by a possibility distribution over the unit interval. For example, if $U = R^1$, we may have

$$g \triangleq X \text{ is small is not very likely} \tag{2.18}$$

where the denotation of <u>small</u> is a fuzzy subset SMALL of R^1 which is characterized by its membership function μ_{SMALL}, and the fuzzy probability <u>not</u> <u>very</u> <u>likely</u> is characterized by the possibility distribution function

$$\pi(v) = 1 - \mu^2_{LIKELY}(v) \ , \quad v \in [0,1] \tag{2.19}$$

in which μ_{LIKELY} is the membership function of the denotation of <u>likely</u> and v is a numerical probability in the interval [0,1].

If the proposition $p \triangleq X$ is G is interpreted as a fuzzy event [20], then (2.17) may be interpreted as the proposition

$$\text{Prob}\{X \text{ is } G\} \text{ is } \lambda$$

which by (2.3) translates into

$$\Pi_{\text{Prob}\{X \text{ is } G\}} = \lambda \ . \tag{2.20}$$

Now, the probability of the fuzzy event $p \triangleq X$ is G is given by [20]

$$\text{Prob}\{X \text{ is } G\} = \int_U p_X(u)\mu_G(u)du \tag{2.21}$$

where $p_X(u)$ is the probability density associated with X. Thus, the translation of (2.17) may be expressed as

$$g \triangleq X \text{ is } G \text{ is } \lambda \longrightarrow \pi(p_X) = \mu_\lambda\left(\int_U p_X(u)\mu_G(u)du\right) \tag{2.22}$$

which signifies that g induces a possibility distribution of the probability distribution of X, with the possibility of the probability density p_X given by the right-hand member of (2.22). For example, in the case of (2.18), we have

$$X \text{ is small is not very likely} \longrightarrow \pi(p_X) = 1 - \mu^2_{LIKELY}\left(\int_U p_X(u)\mu_{SMALL}(u)du\right). \tag{2.23}$$

As a special case of (2.17), a fuzzy granule may be characterized by a proposition of the form

$$g \triangleq X \text{ is } G \tag{2.24}$$

which is not probability-qualified. To differentiate between the general case (2.17) and the special case (2.24), fuzzy granules which are characterized by propositions of the form (2.17) will be referred to as <u>πp-granules</u> (signifying that they correspond to possibility distributions of probability distributions), while those corresponding to (2.24) will be described more simply as <u>π-granules</u>.

A concept which we shall need in our analysis of bodies of evidence is that of a <u>conditioned</u> π-granule. More specifically, if X and Y are variables taking values in U and V, respectively, then a <u>conditioned</u> <u>π-granule</u> in V is characterized by a conditional proposition of the form

$$g \triangleq \text{If } X = u \text{ then } Y \text{ is } G \tag{2.25}$$

where G is a fuzzy subset of V which is dependent on u. From this definition it follows at once that the possibility distribution induced by g is defined by the possibility distribution function

$$\pi_{(Y|X)}(v|u) \triangleq \mathrm{Poss}\{Y = v | X = u\} = \mu_G(v) . \qquad (2.26)$$

An important point which arises in the characterization of fuzzy granules is that the same fuzzy granule may be induced by distinct propositions, in which case the propositions in question are said to be semantically equivalent [16]. A particular and yet useful case of semantic equivalence relates to the effect of negation in (2.17) and may be expressed as ($\leftrightarrow$ denotes semantic equivalence)

$$X \text{ is } G \text{ is } \lambda \leftrightarrow X \text{ is not } G \text{ is ant } \lambda \qquad (2.27)$$

where ant λ denotes the antonym of λ which is defined by

$$\mu_{\text{ant }\lambda}(v) = \mu_\lambda(1-v) \ , \quad v \in [0,1] . \qquad (2.28)$$

Thus, the membership function of ant λ is the mirror image of that of λ with respect to the midpoint of the interval [0,1].

To verify (2.27) it is sufficient to demonstrate that the propositions in question induce the same fuzzy granule. To this end, we note that

$$\begin{aligned} X \text{ is not } G \text{ is ant } \lambda \rightarrow \pi(p_X) &= \mu_{\text{ant }\lambda}\left(\int_U p_X(u)(1-\mu_G(u))du\right) \qquad (2.29) \\ &= \mu_{\text{ant }\lambda}\left(1 - \int_U p_X(u)\mu_G(u)du\right) \\ &= \mu_\lambda\left(\int_U p_X(u)\mu_G(u)du\right) \end{aligned}$$

which upon comparison with (2.22) establishes the semantic equivalence expressed by (2.27).

In effect, (2.27) indicates that replacing G with its negation may be compensated by replacing λ with its antonym. A simple example of an application of this rule is provided by the semantic equivalence

$$X \text{ is small is likely} \leftrightarrow X \text{ is not small is unlikely} \qquad (2.30)$$

in which unlikely is interpreted as the antonym of likely.

A concept that is related to and is somewhat weaker than that of semantic equivalence is the concept of semantic entailment [16]. More specifically, if g_1 and g_2 are two propositions such that the fuzzy granule induced by g_1 is contained in the fuzzy granule induced by g_2, then g_2 is semantically entailed by g_1 or, equivalently, g_1 semantically entails g_2. To establish the relation of containment it is sufficient to show that

$$\pi_1(p_X) \leq \pi_2(p_X) \ , \quad \text{for all } p_X \qquad (2.31)$$

where π_1 and π_2 are the possibilities corresponding to g_1 and g_2, respectively.

As an illustration, it can readily be established that ($\mapsto$ denotes semantic entailment)

$$X \text{ is } G \text{ is } \lambda \mapsto X \text{ is very } G \text{ is } {}^2\lambda \qquad (2.32)$$

or, more concretely,

$$X \text{ is small is likely} \mapsto X \text{ is very small is } {}^2\text{likely} \qquad (2.33)$$

where the left-square of λ is defined by

$$\mu_{2_\lambda}(v) = \mu_\lambda(\sqrt{v}) \ , \quad v \in [0,1]$$

and μ_λ is assumed to be monotone nondecreasing. Intuitively, (2.32) signifies that an intensification of G through the use of the modifier <u>very</u> may be compensated by a dilation (broadening) of the fuzzy probability λ.

To establish (2.32), we note that

$$X \text{ is } G \text{ is } \lambda \rightarrow \pi_1(p_X) = \mu_\lambda\left(\int_U p_X(u)\mu_G(u)du\right) \tag{2.34}$$

$$X \text{ is very } G \text{ is } {}^2\lambda \rightarrow \pi_2(p_X) = \mu_{2_\lambda}\left(\int_U p_X(u)\mu_G^2(u)du\right) \tag{2.35}$$
$$= \mu_\lambda\left(\sqrt{\int_U p_X(u)\mu_G^2(u)du}\right) .$$

Now, by Schwarz's inequality

$$\sqrt{\int_U p_X(u)\mu_G^2(u)du} \geq \int_U p_X(u)\mu_G(u)du \tag{2.36}$$

and since μ_λ is monotone nondecreasing, we have

$$\pi_1(p_X) \leq \pi_2(p_X)$$

which is what we wanted to demonstrate.

3. Analysis of Granular Evidence

As was stated in the introduction, a <u>body of evidence</u> or, simply, <u>evidence</u>, E, may be regarded as a collection of propositions

$$E = \{g_1,\ldots,g_N\} . \tag{3.1}$$

In particular, evidence is <u>granular</u> if its constituent propositions are characterizations of fuzzy granules.

For the purpose of our analysis it is necessary to differentiate between two types of evidence which will be referred to as <u>evidence of the first kind</u> and <u>evidence of the second kind</u>.

Evidence of the first kind is a collection of fuzzy πp-granules of the form

$$g_i \triangleq Y \text{ is } G_i \text{ is } \lambda_i \ , \quad i = 1,\ldots,N \tag{3.2}$$

where Y is a variable taking values in V, $G_1,\ldots,G_N$ are fuzzy subsets of V and $\lambda_1,\ldots,\lambda_N$ are fuzzy probabilities.

Evidence of the second kind is a probability distribution of conditioned π-granules of the form

$$g_i \triangleq Y \text{ is } G_i . \tag{3.3}$$

Thus, if X is taken to be a variable which ranges over the index set $\{1,\ldots,N\}$, then we assume to know (a) the probability distribution $P_X = \{p_1,\ldots,p_N\}$, where

$$p_i \triangleq \text{Prob}\{X = i\} \ , \quad i = 1,\ldots,N \tag{3.4}$$

and (b) the conditional possibility distribution $\Pi_{(Y|X)}$, where

$$\Pi_{(Y|X=i)} = G_i \ , \quad i = 1,\ldots,N . \tag{3.5}$$

In short, we may express evidence of the second kind in a symbolic form as

$$E = \{P_X, \Pi_{(Y|X)}\}$$

which signifies that the evidence consists of P_X and $\Pi_{(Y|X)}$, rather than P_X and $P_{(Y|X)}$ (conditional probability distribution of Y given X), which is what is usually assumed to be known in the traditional probabilistic approaches to the analysis of evidence. Viewed in this perspective, the type of evidence considered in the theories of Dempster and Shafer is evidence of the second kind in which the G_i are crisp sets and the probabilities $p_1,\ldots,p_n$ are known numerically.

In the case of evidence of the first kind, our main concern is with obtaining an answer to the following question. Given E, find the probability, λ, or, more specifically, the possibility distribution of the probability λ, that Y is Q, where Q is an arbitrary fuzzy subset of V.

In principle, the answer to this question may be obtained as follows.

First, in conformity with (2.20), we interpret each of the constituent propositions in E,

$$g_i \triangleq Y \text{ is } G_i \text{ is } \lambda_i \ , \quad i = 1,\ldots,N \tag{3.6}$$

as the assignment of the fuzzy probability λ_i to the fuzzy event $q_i \triangleq Y$ is G_i. Thus, if $p(\cdot)$ is the probability density associated with Y, then in virtue of (2.22) we have

$$\pi_i(p) = \mu_{\lambda_i}\left(\int_V p(v)\mu_{G_i}(v)dv\right) \tag{3.7}$$

where $\pi_i(p)$ is the possibility of p given g_i, and μ_{λ_i} and μ_{G_i} are the membership functions of λ_i and G_i, respectively.

Since the evidence $E = \{g_1,\ldots,g_N\}$ may be regarded as the conjunction of the propositions $g_1,\ldots,g_N$, the possibility of $p(\cdot)$ given E may be expressed as

$$\pi(p) = \pi_1(p) \wedge \cdots \wedge \pi_N(p) \tag{3.8}$$

where $\wedge \triangleq$ min. Now, for a p whose possibility is expressed by (3.8), the probability of the fuzzy event $q \triangleq X$ is Q is given by

$$\rho(p) = \int_V p(v)\mu_Q(v)dv \ . \tag{3.9}$$

Consequently, the desired possibility distribution of $\rho(p)$ may be expressed in a symbolic form as the fuzzy set [21]

$$\lambda = \int_{[0,1]} \pi(p)/\rho(p) \tag{3.10}$$

in which the integral sign denotes the union of singletons $\pi(p)/\rho(p)$.

In more explicit terms, (3.10) implies that if ρ is a point in the interval [0,1], then $\mu_\lambda(\rho)$, the grade of membership of ρ in λ or, equivalently, the possibility of ρ given λ, is the solution of the variational problem

$$\mu_\lambda(\rho) = \text{Max}_p\left(\pi_1(p) \wedge \cdots \wedge \pi_N(p)\right) \tag{3.11}$$

subject to the constraint

$$\rho = \int_V p(v)\mu_Q(v)dv \ . \tag{3.12}$$

In practice, the solution of problems of this type would, in general, require both discretization and approximation, with the aim of reducing (3.11) to a computationally feasible problem in nonlinear programming. In the longer run, however, a more effective solution would be a "fuzzy hardware" implementation which would yield directly a linguistic approximation to λ from the specification of q and E.

It should be noted that if we were concerned with a special case of evidence of the first kind in which the probabilities λ_i are numerical rather than fuzzy, then we could use as an alternative to the technique described above the maximum entropy principle of Jaynes [22] or its more recent extensions [23]-[26]. In application to the problem in question, this method would first yield a probability density $p(\cdot)$ which is a maximum entropy fit to the evidence E, and then, through the use of (3.12), would produce a numerical value for λ.

A serious objection that can be raised against the use of the maximum entropy principle is that, by constructing a unique $p(\cdot)$ from the incomplete information in E, it leads to artificially precise results which do not reflect the intrinsic imprecision of the evidence and hence cannot be treated with the same degree of confidence as the factual data which form a part of the database. By contrast, the method based on the use of possibility distributions leads to conclusions whose imprecision reflects the imprecision of the evidence from which they are derived and hence are just as credible as the evidence itself.

Turning to the analysis of evidence of the second kind, it should be noted that, although there is a superficial resemblance between the first and second kinds of evidence, there is also a basic difference which stems from the fact that the fuzzy granules in the latter are π-granules which are conditioned on a random variable. In effect, what this implies is that evidence of the first kind is conjunctive in nature, as is manifested by (3.8). By contrast, evidence of the second kind is disjunctive, in the sense that the collection of propositions in E should be interpreted as the disjunctive statement: g_1 with probability λ_1 or g_2 with probability λ_2 or ... or g_N with probability λ_N.

As was stated earlier, evidence of the second kind may be expressed in the equivalent form

$$E = \{P_X, \Pi_{(Y|X)}\}$$

where X is a random variable which ranges over the index set $U = \{1,\dots,N\}$ and is associated with a probability distribution $P_X = \{p_1,\dots,p_N\}$; and $\Pi_{(Y|X)}$ is the conditional possibility distribution of Y given X, where Y is a variable ranging over V and the distribution function of $\Pi_{(Y|X)}$ is defined by

$$\pi_{(Y|X)}(v|i) \triangleq \text{Poss}\{Y=v|X=i\}\ , \quad i \in U,\ v \in V\ . \tag{3.13}$$

For a given value of X, $X=i$, the conditional possibility distribution $\Pi_{(Y|X)}$ defines a fuzzy subset of V which for consistency with (3.2) is denoted by G_i. Thus,

$$\Pi_{(Y|X=i)} = G_i\ , \quad i = 1,\dots,N \tag{3.14}$$

and more generally

$$\Pi_{(Y|X)} = G_X\ . \tag{3.15}$$

As was pointed out earlier, the theories of Dempster and Shafer deal with a special case of evidence of the second kind in which the G_i and Q are crisp sets and the probabilities $p_1,\dots,p_N$ are numerical. In this special case, the event $q \triangleq Y \in Q$ may be associated with two probabilities: the lower probability λ_* which is defined--in our notation--as

$$\lambda_* \triangleq \text{Prob}\{\Pi_{(Y|X)} \subset Q\} \tag{3.16}$$

and the upper probability λ^* which is defined as[3]

$$\lambda^* \triangleq \text{Prob}\{\Pi_{(Y|X)} \cap Q \neq \theta\} \quad (\theta \triangleq \text{empty set}) . \tag{3.17}$$

The concepts of upper and lower probabilities do not apply to the case where the G_i and Q are fuzzy sets. For this case, we shall define two more general concepts which are related to the modal concepts of necessity and possibility and which reduce to λ_* and λ^* when the G_i and Q are crisp.

For our purposes, it will be convenient to use the expressions sup F and inf F as abbreviations defined by[4]

$$\sup F \triangleq \sup_v \mu_F(v) , \quad v \in V \tag{3.18}$$

$$\inf F \triangleq \inf_v \mu_F(v) , \quad v \in V \tag{3.19}$$

where F is a fuzzy subset of V. Thus, using this notation, the expression for the conditional possibility measure of Q given X may be written as (see (2.16))

$$\begin{aligned} \text{Poss}\{Y \text{ is } Q|X\} &= \text{Poss}\{Y \text{ is } Q|Y \text{ is } G_X\} \\ &= \sup(Q \cap G_X) \end{aligned} \tag{3.20}$$

Since X is a random variable, we can define the expectation of Poss{Y is Q|X} with respect to X. On denoting this expectation by $E\Pi(Q)$, we have

$$\begin{aligned} E\Pi(Q) &\triangleq E_X \,\text{Poss}\{Y \text{ is } Q|X\} \\ &= \sum_i p_i \sup(Q \cap G_i) \end{aligned} \tag{3.21}$$

We shall adopt the expected possibility, $E\Pi(Q)$, as a generalization of the concept of upper probability. Dually, the concept of lower probability may be generalized as follows.

First, we define the conditional certainty (or necessity) of the proposition $q \triangleq$ Y is Q given X by

$$\text{Cert}\{Y \text{ is } Q|X\} \triangleq 1 - \text{Poss}\{Y \text{ is not } Q|X\} . \tag{3.22}$$

Next, in view of the identities

$$\begin{aligned} 1 - \sup(F \cap G) &= \inf((F \cap G)') \\ &= \inf(F' \cup G') \\ &= \inf(G \Rightarrow F') \end{aligned} \tag{3.23}$$

where the implication $\Rightarrow$ is defined by (see (2.13))

$$G \Rightarrow F' \triangleq G' \cup F' , \tag{3.24}$$

we can rewrite the right-hand member of (3.22) as

---

[3]It should be noted that we are not normalizing the definitions of λ_* and λ^*--as is done in the papers by Dempster and Shafer--by dividing the right-hand members of (3.16) and (3.17) by the probability that $\Pi_{(Y|X)}$ is not an empty set. As is pointed out in [27], the normalization in question leads to counterintuitive results.

[4]The definitions in question bear a close relation to the definitions of universal and existential quantifiers in L_{Aleph_1} logic [28].

$$\text{Cert}\{Y \text{ is } Q|X\} = \inf(G_X \Rightarrow Q) \ . \tag{3.25}$$

Finally, on taking the expectation of both sides of (3.22) and (3.25), we have

$$\begin{aligned} EC(Q) &\triangleq E_X \, \text{Cert}\{Y \text{ is } Q|X\} \\ &= \sum_i p_i \inf(G_i \Rightarrow Q) \\ &= 1 - E\Pi(Q') \end{aligned} \tag{3.26}$$

As defined by (3.26), the expression $EC(Q)$, which represents the <u>expected certainty</u> of the conditional event (Y is $Q|X$), may be regarded as a generalization of the concept of lower probability.

The set functions $E\Pi(Q)$ and $EC(Q)$ may be interpreted as fuzzy measures. However, in general, these measures are neither normed nor additive. Instead, $E\Pi(Q)$ and $EC(Q)$ are, respectively, superadditive and subadditive in the sense that, for any fuzzy subsets Q_1 and Q_2 of V, we have

$$EC(Q_1 \cup Q_2) \geq EC(Q_1) + EC(Q_2) - EC(Q_1 \cap Q_2) \tag{3.27}$$

and

$$E\Pi(Q_1 \cup Q_2) \leq E\Pi(Q_1) + E\Pi(Q_2) - E\Pi(Q_1 \cap Q_2) \ . \tag{3.28}$$

It should be noted that these inequalities generalize the superadditive and subadditive properties of the measures of belief and plausibility in Shafer's theory.

The inequalities in question are easy to establish. Taking (3.28), for example, we have

$$\begin{aligned} E\Pi(Q_1 \cup Q_2) &= \textstyle\sum_i p_i \sup_v \Big[(\mu_{Q_1}(v) \vee \mu_{Q_2}(v)) \wedge \mu_{G_i}(v) \Big] \\ &= \sum_i p_i \sup_v (\mu_{Q_1}(v) \wedge \mu_{G_i}(v) \vee \mu_{Q_2}(v) \wedge \mu_{G_i}(v)) \\ &= \sum_i p_i \Big[\sup_v (\mu_{Q_1}(v) \wedge \mu_{G_i}(v)) \vee \sup_v (\mu_{Q_2}(v) \wedge \mu_{G_i}(v)) \Big] \end{aligned} \tag{3.29}$$

Now, using the identity ($a,b \triangleq$ real numbers)

$$a \vee b = a + b - a \wedge b \tag{3.30}$$

the right-hand member of (3.29) may be rewritten as

$$\begin{aligned} E\Pi(Q_1 \cup Q_2) = \sum_i p_i \Big[&\sup_v (\mu_{Q_1}(v) \wedge \mu_{G_i}(v)) + \sup_v (\mu_{Q_2}(v) \wedge \mu_{G_i}(v)) \\ &- \big[\sup_v (\mu_{Q_1}(v) \wedge \mu_{G_i}(v)) \wedge \sup_v (\mu_{Q_2}(v) \wedge \mu_{G_i}(v)) \big] \Big] \end{aligned} \tag{3.31}$$

Furthermore, from the min-max inequality

$$\sup_v f(v) \wedge \sup_v g(v) \geq \sup_v (f(v) \wedge g(v)) \tag{3.32}$$

it follows that

$$\begin{aligned} &\sup_v (\mu_{Q_1}(v) \wedge \mu_{G_i}(v)) \wedge \sup_v (\mu_{Q_2}(v) \wedge \mu_{G_i}(v)) \\ &\geq \sup_v (\mu_{Q_1}(v) \wedge \mu_{Q_2}(v) \wedge \mu_{G_i}(v)) \end{aligned} \tag{3.33}$$

and hence that

$$E\Pi(Q_1 \cup Q_2) \leq \sum_i p_i \sup_v(\mu_{Q_1}(v) \wedge \mu_{G_i}(v)) + \sum_i p_i \sup_v(\mu_{Q_2}(v) \wedge \mu_{G_i}(v)) - \sum_i p_i \sup_v(\mu_{Q_1}(v) \wedge \mu_{Q_2}(v) \wedge \mu_{G_i}(v)) \quad . \tag{3.34}$$

Finally, on making use of (3.21) and the definition of $Q_1 \cap Q_2$, we obtain the inequality

$$E\Pi(Q_1 \cup Q_2) \leq E\Pi(Q_1) + E\Pi(Q_2) - E\Pi(Q_1 \cap Q_2) \tag{3.35}$$

which is what we set out to establish.

The superadditive property of $EC(Q)$ has a simple intuitive explanation. Specifically, because of data granularity, if Q_1 and Q_2 are roughly of the same size as the granules $G_1,\dots,G_N$, then $EC(Q_1)$ and $EC(Q_2)$ are likely to be small, while $E(Q_1 \cup Q_2)$ may be larger because the size of $Q_1 \cup Q_2$ is likely to be larger than that of $G_1,\dots,G_N$. For the same reason, with the increase in the relative size of Q_1 and Q_2, the effect of granularity is likely to diminish, with $EC(Q)$ tending to become additive in the limit.

In the foregoing analysis, the probabilities $p_1,\dots,p_N$ were assumed to be numerical. This, however, is not an essential restriction, and through the use of the extension principle [21], the concepts of expected possibility and expected certainty can readily be generalized, at least in principle, to the case where the probabilities in question are fuzzy or linguistic. Taking the expression for $E\Pi(Q)$, for example,

$$E\Pi(Q) = \sum_i p_i \sup(Q \cap G_i) \tag{3.36}$$

and assuming that the p_i are characterized by their respective possibility distribution functions $\pi_1,\dots,\pi_N$, the determination of the possibility distribution function of $E\Pi(Q)$ may be reduced to the solution of the following variational problem

$$\pi(z) \triangleq \mathrm{Max}_{v_1,\dots,v_N} \; \pi_1(v_1) \wedge \cdots \wedge \pi_N(v_N) \tag{3.37}$$

subject to

$$z = v_1 \sup(Q \cap G_1) + \cdots + v_N \sup(Q \cap G_N)$$

$$v_1 + \cdots + v_N = 1$$

which upon solution yields the possibility, $\pi(z)$, of a numerical value, z, of $E\Pi(Q)$. Then, a linguistic approximation to the possibility distribution would yield an approximate value for $\Pi_{E\Pi(Q)}$ expressed as, say, <u>not very high</u>.

As was alluded to already, a basic issue in the analysis of evidence relates to the manner in which two or more distinct bodies of evidence may be combined. In the case of evidence of the second kind, for example, let us assume for simplicity that we have two bodies of evidence of the form

$$E = \{E_1, E_2\} \tag{3.38}$$

in which

$$E_1 = \{P_{X_1}, \Pi_{(Y|X_1)}\} \tag{3.39}$$

$$E_2 = \{P_{X_2}, \Pi_{(Y|X_2)}\} \tag{3.40}$$

where Y takes values in V; while X_1 and X_2 range over the index sets $U_1 = \{1,\dots,N_1\}$ and $U_2 = \{1,\dots,N_2\}$, and are associated with the joint probability distribution $P_{(X_1,X_2)}$ which is characterized by

$$p_{ij} \triangleq \mathrm{Prob}\{X_1 = i, \; X_2 = j\} \quad . \tag{3.41}$$

For the case under consideration, the expression for the expected possibility of the fuzzy event $q \triangleq Y$ is Q given E_1 and E_2 becomes

$$E\Pi(Q) = E_{(X_1,X_2)} \text{Poss}\{Y \text{ is } Q|(X_1,X_2)\} \tag{3.42}$$
$$= \sum_{i,j} p_{ij} \sup(Q \cap G_i \cap H_j)$$

where

$$\Pi_{(Y|X_1=i)} \triangleq G_i \tag{3.43}$$

and

$$\Pi_{(Y|X_2=j)} \triangleq H_j \ . \tag{3.44}$$

The rule of combination of evidence developed by Dempster [7] applies to the special case of (3.42) in which the sets G_i and H_j are crisp and X_1 and X_2 are independent. In this case, from the knowledge of $E\Pi(Q)$ (or $EC(Q)$) for each of the constituent bodies of evidence and $Q \subset V$, we can determine the probability distributions of X_1 and X_2 and then use (3.42) to obtain $E\Pi(Q)$ for the combined evidence. Although simple in principle, the computations involved in this process tend to be rather cumbersome. Furthermore, as is pointed out in [27], there are some questions regarding the validity of the normalization employed by Dempster when

$$G_i \cap H_j = \theta \tag{3.45}$$

for some i, j, and the probability of the event "Y is θ" is positive.

4. Concluding Remarks

Because of its substantial relevance to decision analysis and model validation, analysis of evidence is likely to become an important area of research in the years ahead.

It is a fact of life that much of the evidence on which human decisions are based is both fuzzy and granular. The concepts and techniques outlined in this paper are aimed at providing a basis for a better understanding of how such evidence may be analyzed in systematic terms.

Clearly, the mathematical problems arising from the granularity and fuzziness of evidence are far from simple. It may well be the case that their full solution must await the development of new types of computing devices which are capable of performing fuzzy computations in a way that takes advantage of the relatively low standards of precision which the results of such computations are expected to meet.

References and Related Papers

1. A.N. Tikhonov and V. Ya. Arsenin, Methods of Solution of Ill-Posed Problems, Nauka, Moscow, 1974.

2. S. Gutman, "Uncertain dynamical systems--a Lyapounov min-max approach," IEEE Trans. on Automatic Control, AC-24, 437-443, 1979.

3. F. Schlaepfer and F. Schweppe, "Continuous-time state estimation under distrubances bounded by convex sets," IEEE Trans. on Automatic Control, AC-17, 197-205, 1972.

4. D.R. Ries and M.R. Stonebraker, "Locking granularity revisited," ERL Memorandum M78/71, Electronics Research Laboratory, University of California, Berkeley, 1978.

5. R.E. Moore, Interval Analysis, Prentice-Hall, Englewood Cliffs, N.J., 1966.

6. G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, 1976.

7. A.P. Dempster, "Upper and lower probabilities induced by a multivalued mapping," Ann. Math. Statist. 38, 325-329, 1967.

8. P. Smets, "Un modele mathematico-statistique simulant le processus du diagnostic medicale," Free University of Brussels, 1978.

9. L.J. Cohen, The Implications of Induction, Methuen, London, 1970.

10. G.L.S. Shackle, Decision, Order and Time in Human Affairs, Cambridge University Press, Cambridge, 1961.

11. P. Suppes and M. Zanotti, "On using random relations to generate upper and lower probabilities," Synthese 36, 427-440, 1977.

12. H.T. Nguyen, "On random sets and belief functions," J. Math. Anal. Appl. 65, 531-542, 1978.

13. L.A. Zadeh, "Fuzzy sets as a basis for a theory of possibility," Fuzzy Sets and Systems 1, 3-28, 1978.

14. H.T. Nguyen, "On conditional possibility distributions," Fuzzy Sets and Systems 1, 299-309, 1978.

15. E. Hisdal, "Conditional possibilities: independence and noninteraction," Fuzzy Sets and Systems 1, 283-297, 1978.

16. L.A. Zadeh, "PRUF--a meaning representation language for natural languages," Int. J. Man-Machine Studies 10, 395-460, 1978.

17. M. Mizumoto, S. Fukame and K. Tanaka, "Fuzzy reasoning methods by Zadeh and Mamdani, and improved methods," Proc. Third Workshop on Fuzzy Reasoning, Queen Mary College, London, 1978.

18. B.S. Sembi and E.H. Mamdani, "On the nature of implication in fuzzy logic," Proc. 9th International Symposium on Multiple-Valued Logic, Bath, England, 143-151, 1979.

19. W. Bandler and L. Kohout, "Application of fuzzy logics to computer protection structures," Proc. 9th International Symposium on Multiple-Valued Logic, Bath, England, 200-207, 1979.

20. L.A. Zadeh, "Probability measures of fuzzy events," J. Math. Anal. Appl. 23, 421-427, 1968.

21. L.A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning, Part I," Information Sciences 8, 199-249, 1975; Part II, Information Sciences 8, 301-357, 1975; Part III, Information Sciences 9, 43-80, 1975.

22. E.T. Jaynes, "Information theory and statistical mechanics," Parts I and II, Physical Review 106, 620-630; 108, 171-190, 1957.

23. S. Kullback, Information Theory and Statistics, John Wiley, New York, 1959.

24. M. Tribus, Rational Descriptions, Decisions and Designs, Pergamon Press, New York, 1969.

25. J.E. Shore and R.W. Johnson, "Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy," NRL Memorandum Report 3-898, Naval Research Laboratory, Washington, D.C., 1978.

26. P.M. Williams, "Bayesian conditionalization and the principle of minimum information," School of Mathematical and Physical Sciences, The University of Sussex, England, 1978.

27. L.A. Zadeh, "On the validity of Dempster's rule of combination of evidence," Memorandum M79/24, Electronics Research Laboratory, University of California, Berkeley, 1979.

28. N. Rescher, Many-Valued Logic, McGraw-Hill, New York, 1969.

29. J. Kampé de Feriet and B. Forte, "Information et probabilité," Comptes Rendus Acad. Sci. A-265, 110-114, 142-146, 350-353, 1967.

30. M. Sugeno, "Theory of fuzzy integrals and its applications," Tokyo Institute of Technology, 1974.

31. T. Terano and M. Sugeno, "Conditional fuzzy measures and their applications," in Fuzzy Sets and Their Application to Cognitive and Decision Processes (L.A. Zadeh, K.S. Fu, K. Tanaka and M. Shimura, eds.), 151-170, 1975.

32. E. Sanchez, "On possibility-qualification in natural languages," Memorandum M77/28, Electronics Research Laboratory, University of California, Berkeley, 1977.

33. P.M. Williams, "On a new theory of epistemic probability (review of G. Shafer: A Mathematical Theory of Evidence)," Brit. J. for the Philosophy of Science 29, 375-387, 1978.

34. P. Suppes, "The measurement of belief," J. Roy. Statist. Soc. B 36, 160-175, 1974.

35. K.M. Colby, "Simulations of belief systems," in Computer Models of Thought and Language (R.C. Schank and K.M. Colby, eds.), W. Freeman, San Francisco, 1973.

36. B.C. Bruce, "Belief systems and language understanding," N.I.H. Report CBM-TR-41, Rutgers University, 1975.

37. R.P. Abelson, "The structure of belief systems," in Computer Models of Thought and Language (R.C. Schank and K.M. Colby, eds.), W. Freeman, San Francisco, 1973.

38. R.D. Rosenkrantz, Inference, Method and Decision, D. Reidel, Dordrecht, 1977.

39. N. Rescher, Plausible Reasoning, Van Gorcum, Amsterdam, 1976.

40. A. Tversky and D. Kahneman, "Judgment under uncertainty: heuristics and biases," Science 185, 1124-1131, 1974.

41. G. Banon, "Distinctions between several types of fuzzy measures," Proc. Int. Colloquium on the Theory and Applications of Fuzzy Sets, University of Marseille, Marseille, 1978.

42. I.J. Good, "Subjective probability as the measure of a non-measurable set," in Logic, Methodology and Philosophy of Science: Proceedings of the 1960 International Congress (E. Nagel, P. Suppes and A. Tarski, eds.), Stanford University Press, 1962.