TR/12/84 November 1984

Fuzzy Sets, Probability *& the* Nature of Uncertainty

by

R.I. Phelps

Introduction

This paper considers problems associated with the use of probabilistic and fuzzy methods to deal with uncertainty in decision support systems. A distinction is drawn between the statistical approach to inference and the use of probability and fuzzy methods within a structured, knowledge based approach. The application of these contrasting methods within a structured approach is considered and drawbacks of the two methods noted. The nature of uncertainty is then investigated and a distinction is made between two types of uncertainty: uncertainty in frequency of occurrence and uncertainty in similarity judgements. These two types of uncertainty are examined and it is concluded that probability is the natural formalism for the former type and that fuzzy methods fit more naturally with the latter type. Consideration of similarity judgements points toward the research that is necessary for a useful fuzzy theory.

Structured and unstructured processes.

In contrasting the probabilistic and fuzzy approaches an argument often used against probability is the inadequacy of statistical methods in providing decision support systems for complex situations [1]. For the sake of a concrete exposition, the discussion will be set in the field of medical diagnosis systems. This is a field in which both fuzzy and probabilistic, knowledge-based and statistical approaches have been tried 2. Data is available in the form of patient's symptoms and a decision is to be made about the disease causing these symptoms. The pure statistical approach is to find a correlation between symptoms and diseases without constructing any intervening model of the process linking the two. In this sense it is an unstructured process. Past data or subjective beliefs are used to estimate $p(\underline{S}|D)$, the probability that a particular set of symptoms $\underline{S}$ would arise from a particular disease D, for all possible combinations of $\underline{s}$ and D. The great practical drawback is that the number of possible combinations of symptoms is generally enormous so that sufficient data is not available (and even if it were the task of estimating and. storing all the $p(\underline{S}|D)$ would be prohibitive). In effect this approach is a complete enumeration of all possible paths from symptoms to diseases.

Because the data requirements are fatal to the application of this approach in its pure form, the general response has been to assume independence of the various symptoms conditional on the disease:

$$p(S_1, S_2, \ldots S_n|D) = p(S_1 | D) \ldots p(S_n | D)$$

Using Bayes Theorem and .taking logs of the odds on D, $p(D)/p(\overline{D})$, we get:

$$\log \frac{P(D \mid S_1 \ldots S_n)}{P(\overline{D} \mid S_1 \ldots S_n)} = \log \frac{P(D)}{P(\overline{D})} + \sum_i \log \frac{P(S_i \mid D)}{P(S_i \mid \overline{D})}$$

giving a simple addit ive score where the effect of each symptom is to add its ' we ight of evidence' to the prior odds on D. With this assumption only the $P(S_i|D)$ need to be estimated separately (not the joint $P(S_i \ldots S_n|D)$) which enormously reduces the data requirements and makes the computation feasible. This in essence constitutes the statistical approach to inference.

The knowledge based approach to inference is by contrast essentially structured. A model is constructed which aims to capture what an expert knows about his field. It may include interactions between symptoms and how diseases and symptoms are l inked through intermediate parts of the system, e.g. by firstly deciding whether there is a liver disease before going on to reach a final diagnosis. The inference procedure is often expressed by production rules [3], typically of the form:

IF symptom A and symptom B and ... THEN hypothesis H with certainty C.

Thus symptoms are considered as interacting groups rather than independent contributors to the diagnosis, and there is not necessarily a direct link from symptoms to disease but instead intermediate hypotheses are considered. Data for these structured systems is needed to give values to the certainty factors, but these need to be estimated only for the particular (S,H) symptoms-hypothesis sets occuring in the expert derived inference rules, not for all possible sets, and so the computational task is greatly reduced.

The general failure of statistical methods to provide useful and acceptable decision support systems in such complex areas, due largely to the data requirements and independence assumptions outlined above, has attached a certain degree of guilt by association to the use of probabilities in expert systems. Statistical models use probability, but probability can equally well be used in a knowledge-based system. Fuzzy and probabilistic approaches within such a system are now considered.

Structured systems - probabilistic and fuzzy approaches.

If probabilities are to be used for the uncertainty factors within a knowledge based-system, then as previously mentioned, the data requirements are reduced to ascertaining probabilities for the (S.,H) sets occurring in the rule-base. For a large system this will still be computationally prohibitive and so, as before, the assumption of independence of the symptoms must be made

$$p(S_1,\ldots S_n| H) = p(S_1 |lH) \ldots p(S_n| H)$$

Since the rules often group together specifically dependent sets of symptoms this leads to a prima facie contradiction between the inference structure and the uncertainty measure. Apart from this dependency problem another difficulty arises in the absence of data to accurately estimate a probability. A subjective estimate needs to be made and arguments against subjective prior distributions are well known in Statistics [4]. Although the use of probability brings these problems, it does have the advantage of an axiomatic behavioural foundation [5] It is therefore possible to have confidence that the operations of probability theory are the most 'rational' operations possible if the behavioural axioms are accepted as desirable for rat ional thought, which they generally are. At the very least, a probabilistic system wi1l be attempting to approximate these most rational operations as closely as possible.

Fuzzy sets were intended to be used in dealing with cooplex systems such as these, where statistical methods cannot easily be applied. A fuzzy set A consists of a universe X and a mapping

$$\mu_A : X \rightarrow |0,1|.$$

$\mu_A(x), x \in x$, is the degree of membership of x in A

$$\mu_{\bar{A}}(x) = 1 - \mu_A(x)$$

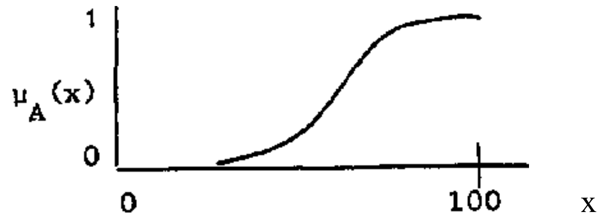
As operat ions to combine fuzzy sets Zadeh |6| suggested:

$$\mu_{AUB}(x) = \text{Max}\{_A(x), \mu_B(x)\}$$
$$\mu_{A \cap B}(x) = \text{Min}\{\mu_A(x), \mu_{A(x)}\}$$

however, these operators are not the only ones possible. The possible operators depend on the set - theoretic axioms we would like U & ∩ to satisfy, and ideas on these differ. The notion of membership is quite wide and not restricted to measuring the likelihood of occurrences. Thus membership functions can be constructed for concepts such as 'hot':

A = 'hot'

X = [0,100] C



Linguistic 'hedges' can also be dealt with by simple expedients, e.g. for B = 'very hot' we might use $\mu_B(x) = \mu_A^2(x)$.

By considering μ(x) as a 'possibility' of x occurring, membership functions can be used specifically to measure likelihood, forming a direct replacement for subjective probability and intended for cases where objective frequency data is not available. The possibility of x is π_A. (x). This can be thought of as the possibility of x occurring out of the set of all possible outcomes represented by A

$$\Pi_A(x) = \mu_A(x)$$

The operations postulated by Zadeh [7] are similar to those for fuzzy sets :

$$\Pi_{A,B}(x \cup y) - \text{Max}\{\Pi_A(X), \Pi_B(y)\}$$

and if the variables are non-interactive, i.e. the possible outcomes in A do not depend on the outcome occurring in B and vice-versa, then

$$\Pi_{A\,B}(x \cap y) = \text{Min}\{\pi_A(x), \pi_B(y)\}$$

In application this noninteractive condition is generally assumed to hold. It is weaker than probabilistic independence which implies it [8]. Conditional possibilities can also be defined — they have the property that they are usually identical with the joint possibilities. Although there is a theoretical distinction between membership functions and possibilities the fact that possibility is defined as a membership function and that the standard operations are the same in both cases means that there is little distinction in the propogation of uncertainty measures in practice.

There are a number of problems with the fuzzy approach. Firstly there is the question of how membership function values are obtained. This is essentially a subjective choice on the part of the modeller, and corresponds to the problem of choosing a subjective prior distribution in probability. Secondly, the choice of fuzzy operators is also subjective |9|. Max and min operators tend to be used as the default operators, but as they often do not give results considered sensible, modellers do not hesitate to replace them with other (ad-hoc) operations which give better results in their particular application. Thirdly, the problem of dependent variables noted for probabilities is equally present under the fuzzy approach although hidden by the definition of the operators. Under the non interaction assumption the same operator is applied to variables which are dependent and independent in probabilistic terms. This is equivalent to assuming independence in the probability framework. Fourthly, the application to inference rules is not straightforward. For a rule of form:

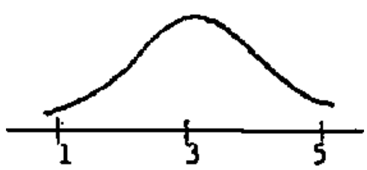
If S_1 and S_2 then H with certainty 0.4 ,

where the possibility values of S_1 and S_2 have been established at 0.7 and 0.5, use of the Min operator gives a possibility value of 0,5 for (S_1 and S_2) but there remains the problem of how to combine this with the rule certainty of 0.4 in order to end up with the possibility value of H Again, approaches to this tend to be ad-hoc. Fifthly, there is no behavioural axiomatic derivation for fuzzy sets and so the implications of the operations as a model for human behaviour are not clear. Certain set-theoretic conditions and restricted operator spaces can be shown tp lead to the use of Max and Min operators, but although these conditions may look pleasing from a mathematical viewpoint it is not clear that they constitute desirable behavioural rules. Fung & Fu |10| consider combining the fuzzy preferences of group members to obtain a fuzzy preference for the whole group. They set down rather restrictive axioms of combination which lead to the use of Max and Min operators, but these behavioural axioms are not obviously desirable ones, even in this special case of group decision making.

Frequency and Similarity

Uncertainty as studied in probability and statistics is a measure of how likely an event is to occur. It is essentially the study of frequencies of events. A frequency view of probability acknowledges this explicitly and a subjectivist "degree of belief" view is founded on behavioural principles such as avoiding loss in a repeated sequence of gambles. Although there may be no directly applicable data, a subjective belief in the probability of an event will be based on considerations of the frequency of other similar events. The mathematical axioms of probability are designed to make sense when interpreted in this frequential manner.

This is not the only type of uncertainty, however, as is shown by considering some problems. Consider a distribution (or membership function) for remaining life after detection of a serious disease.



Remaining life (Yrs).

This may have been obtained from a person with only a vague belief that it will be somewhere in the 1-5 year region and more likely around 3 years, or from a person who has derived this curve from a study of much data about the lives of this sort of patient. Clearly these two persons' curves should not be treated as holding the same information. Yet if a frequency distribution is the only way of encoding uncertainty then there is no way to represent the difference in certainties expressed by the two distributions. This difference exists because the second person has derived his judgement entirely from frequency information whereas the first person has had to supplement any frequency information he had with frequency information from other similar examples. e.g. similar diseases.

As another example, if we have data on past predictions of time left to live made by two consultants then we might wish to assess our degree o confidence in each consultant. In order to do this it is necessary to examine their predictions and the actual outcomes• If neither consultant is precisely accurate then it does not make sense to talk of the frequencies of their being right, and the assessment must be made on the basis of the similarity between two patterns: the pattern of predictions and the pattern of outcomes.

These examples indicate that there is a second kind of uncertainty which is an uncertainty not of how likely an event is but of how similar it is to others- This notion of similarity is not frequency-based in th "likely" sense. It is a question of categorization and thus relates to the idea of a 'degree of membership' For example, is a psychiatrist's couch more similar to a chair or a bed? There is uncertainty here which cannot be quantified in frequency terms. A figure on a 0-1 scale may be used to quantify this type of uncertainty but it is not a frequency measurement, rather an assessment of the degree of similarity between two structural descriptions or patterns. This type of uncertainty is present in linguistic statements involving 'hedges' e.g. 'it is quite hot' where 'quite' is indicative of degree of similarity. Even when uncertainty could be assessed purely from frequencies, it is unlikely that humans often do so - it seems intuitively more likely that consultants match incomplete patterns of symptoms with patterns typical of diseases rather than performing complex frequency calculations in their heads.

So in making uncertainty judgements humans are generally doing two things (i) making a judgement as to how similar aspects of one event are to aspects of other events (ii) using these other events to make a frequency judgement. Both probability and fuzzy set approaches attempt to treat these two different types of uncertainty in the same way. Probabilities are treated by the same rules irrespective of the input from frequency and similarity considerations. In the fuzzy approach there is a theoretical distinction between possibilities and membership functions, but since a possibility is a particular membership value and the same Max, Min operations are used in both cases, this distinction disappears in practice .

Uncertainty of frequency.

For this type of uncertainty, the previous exposition has indicated a sounder axiomatic development and fewer problems in application for the probability formalism relative to the fuzzy approach. The main difficulty in application is that of dependent variables, also present in the fuzzy case. To overcome this several methods have been proposed, including fitting "weighting factors" $\{a_i\}$ to the equation:

$$\log \frac{P(H \mid \underline{S})}{P(H \mid \underline{S})} = a_0 + \sum_i ai \log \frac{P(S_i \mid H)}{P(S_i \mid H)}$$

for a set of values of $\underline{S}$. The a are intended to improve the fit by 'dampening down' the weights of evidence; however any linear model is necessarily unable to account for strong interactions between two or more variables. Another suggest ion is the use of "Lancaster models" |11 |. There is no Sth order interaction in the Lancaster sense among n variables x if

$$\pi_{i \in A} \left(F_i^* - F_i\right) = 0 \quad , \quad \text{for all } A \subseteq \{1,\ldots,n\} \ ,|A| = S+1$$

where F is a dummy, replaced after multiplication by $F_i^* \ldots F_j^* F_r \ldots F_t \rightarrow$ $F_{i \ldots j}$, $F_r \ldots F_t$ and $F_{i \ldots j}$ is the joint distribution of $(x_i \ldots x_j)$. No 1st order interaction means complete independence. If there is no Sth order interaction then $F_{1 \ldots n}$ can be constructed from knowing $F_{i \ldots j}$ for all $|\{i,\ldots,j\}| = S$. Thus if we assume no 2nd order interaction we need only know the joint distribution for all pairs of variables in order to construct the complete joint distribution for all n variables. In effect, no 2nd order interaction allows dependency only within pairs of variables and not between these pairs. Problems of this method are that interaction may occur between 3 or more variables and that the number of pairs, n(n+l), grows rapidly with n . Other possibilities are to reduce the number of variables $(S_1,\ldots,S_n)$ to a manageable size by using a data reduction technique such as factor analysis, principal components or multidimensional scaling. However, if explanatory power is desirable in the system then the opacity of this approach is a drawback even if the data reduction proves possible.

Uncertainty of similarity.

For this type of uncertainty the idea of events having degrees of membership in a shared class is clearly relevant to measuring similarity, pointing toward the use of fuzzy sets. However, there is still the question of whether the operations of fuzzy sets are useful in measuring similarity of concepts. It seems not. Osherson & Smith [12] point out that goldfish is more typical of the concept 'pet fish' (and will therefore have a higher membership value) than it is of 'pet' or of 'fish'. But use of the Min rule of corabinat ion gives the opposite result .

There is a need for a study of how humans compare concepts - matching patterns or their features against one another. How should such similarities be measured and combined and how should they interact with probabilities? A preliminary discussion of this point is given in Cohen and Murphy [13]. This is a question central to artificial intelligence since it is bound up with knowledge representation and the formation of concepts: measures of similarity will differ depending on the way in which knowledge is structured into patterns and held and conversely the aggregation of similar instances into concepts presupposes a measure of similarity.

The idea that an instance has a degree of membership in a class or concept is strongly linked to the psychological theory of prototypes [14]. Thus the potential use of fuzzy sets in knowledge based systems is likely to be with prototypically structured knowledge bases. The present move toward frame-based representations is a move in this direction.

Conclusions

Uncertainty can be divided into two types: frequency and similarity. These two types require different treatment and rather than being rivals, the probability and fuzzy approaches can be seen as complementary, with probability having the greater relevance to frequency measurement and fuzzy sets the greater relevance to similarity measurement . Although the concept of degree of membership is correct for similarities, the operations of combination currently used are unsatisfactory. It is suggested that fuzzy set theory should develop in the direction of providing a useful theory of similarity, based on human judgements, and that such development should proceed hand-in-hand with developments in knowledge representation formalisms.

REFERENCES.

1. A .R.Feinstein The haze of Bayes, the aerial palaces of decision analys is and the computerized oui ja board , Clinical pharmacology and therepeutics, 21, 482-496.

2. D.J.Spiegelhalter & R.P. Knill-Jones. Statistical & Knowledge based approaches to clinical decision making. J.R.S.S. Series A 1984.

3. N.J.Nilsson. Principles of A.I. Tioga 1980.

4 . V . Barnett . Comparative Statistical inference. Wiley 1982.

5. L. .J. Savage. Foundations of Statistics, Wiley 1954 .

6. L.A.Zadeh. Fuzzy Sets. Information & Control, 8. 1965.

7. C.A.Zadeh. Fuzzy Sets on a bas is for a theory of possibility. Fuzzy sets and systems 1, 1978,

8. E. Hisdal . Conditional possibilities, independence and non-interaction. Fuzzy sets and systems 1, 1978.

9. R. Jain. In Fuzzy Sets. Ed, P.F.Wang & S.K.Chang. Plenum Press, 1980.

10. L.W.Fung & K.S.Fu. An axiomatic approach to rational decision making in a fuzzy environment. In fuzzy sets and their applications to cognitive and decision processes. Ed. L.A.Zadeh et al. Academic Press 1975.

11. R. Zentgraf . A note on Lancaster's definition of higher-order interactions. Bioraetrika, 62, pp 375-8. 1975.

12. D.N.Osherson & E.E.Smith. On the adequacy of prototype theory as a theory of concepts. Cognition. 9pp. 35-58 . 1981 .

13. B. Cohen & G.L. Murphy. Models of concepts. Cognitive Science 8. pp. 27-58. 1984.

14. E. Rosch & B.B. Lloyd, Eds. Cognition and Categorization. Erlbaum 1978.