December 9, 2002

Probability Theory and Fuzzy Logic

Lotfi A. Zadeh*

How does fuzzy logic relate to probability theory? Questions in this vein were raised almost immediately after the publication of my first paper on fuzzy sets (1965). Among the questioners, some have described fuzzy logic as merely a disguised version of probability theory. Some have asserted that anything that can be done with fuzzy logic can be done equally well or better with probability-based techniques. And some, including myself, have argued that probability theory and fuzzy logic are complementary rather than competitive.

In what follows, I advance a view that, to my knowledge, has not been advanced before. More specifically, my contention is that probability theory should be based on fuzzy logic, rather than on classical, Aristotelian, bivalent logic that served and is continuing to serve as its foundation since the times of Bernoulli. The core of my argument is that standard probability theory, call it PT, has fundamental limitations that stem from the bivalent logic on which it is based. Today, my argument may sound like a voice in the wilderness. Nevertheless, I have no doubt that eventually it will gain wide acceptance. What should be underscored, however, is that saying that PT has fundamental limitations does not mean that it has not achieved brilliant successes. What it does mean is that alongside the brilliant successes there are many basic problems which are beyond PT's reach; many questions which PT cannot answer; and many instances of counterintuitive results.

Among the negative consequences of basing PT on bivalent logic there are three stand out in importance. First, the conceptual structure of PT is brittle in the sense that

* Professor in the Graduate School and Director, Berkeley initiative in Soft Computing (BISC), Computer Science Division and the Electronics Research Laboratory, Department of EECS, University of California, Berkeley, CA 94720-1776; Telephone: 510-642-4959; Fax: 510-642-1712; E-Mail: zadeh@cs.berkeley.edu . Research supported in part by ONR N00014-00-1-0621, ONR Contract N00014-99-C-0298, NASA Contract NCC2-1006, NASA Grant NAC2-117, ONR Grant N00014-96-1-0556, ONR Grant FDN0014991035, ARO Grant DAAH 04-961-0341 and the BISC Program of UC Berkeley.

in PT bivalent-logic-based definitions of basic concepts are lacking in robustness. More specifically, if C is a bivalent concept and u is an object in the domain of C, then it is either true that u is an instance of C or it is not true, with no partiality of truth allowed. The problem is that bivalence of C leads to contradictions which were brought to light in the ancient Greek sorites paradox. As an illustration, consider the concept of independence of events. By definition, A and B are independent events if and only if P(A,B)= P(A) P(B). Now, assume that this equality holds to within epsilon. As epsilon increases, at what point will A and B cease to be independent? What this question makes clear is that independence is not a bivalent concept; it is a matter of degree. The same applies to the concepts of stationarity, normality and almost all other concepts in PT. What we see is that bivalence of the conceptual structure of PT is in fundamental conflict with reality—a reality in which almost everything has shades of gray. It is this reality that is the point of departure in fuzzy logic.

The second problem relates to what may be called "the dilemma of it is possible but not probable." An instance of this dilemma is the following. Suppose that 99% of professors have a Ph.D. degree, and that Robert is a professor. What is the probability that Robert has a Ph.D. degree? PT's answer is: between 0 and 1. More generally, if A and B are events such that the intersection of A and B is a proper subset of B, and the Lebesgue measure of the intersection is arbitrarily close to that of B, then all that can be said about the conditional probability of B given A is that it is between 0 and 1.

Last, but most important, PT is lacking in capability to operate on perception-based information. Such information has the form of propositions drawn form a natural language--propositions which describe one or more perceptions. For example, "Eva is young;" "Usually Robert returns from work at about 6 pm;" and, "It is very unlikely that there will be a significant increase in the price of oil in the near future."

The inability of PT to operate on perception-based information is a serious limitation because perceptions have a position of centrality in human cognition.
Thus, human have a remarkable capability to perform a wide variety of physical and mental tasks without any measurements and any computations. Everyday examples of such tasks are parking a car, driving in city traffic, playing tennis and summarizing a story.

Basically, a natural language is a system for describing perceptions. Perceptions an intrinsically imprecise, reflecting the bounded ability of sensory organs, and ultimately the brain, to resolve detail and store information. More specifically, perceptions are f-granular in the sense that (a) the boundaries of perceived classes are fuzzy; and (b) the perceived values of attributes are granular, with a granule being a clump of values drawn together by indistinguishability, similarity, proximity or functionality.

F-granularity of perceptions induces f-granularity of their descriptions in a natural language—and hence, f-granularity of perception-based information. Here we come to a basic point, namely, f-granularity of propositions drawn from a natural language puts them will beyond the reach of existing predicate-logic-based techniques of meaning representation.

The implication of this point is that bivalent-logic-based methods of natural language processing do not have the capability to deal with perception-based information. This is the basis of the conclusion that PT do not have the capability to operate on perception-based information.

As an illustration, consider a perception-based version of the balls-in-box problem. A box contains about twenty balls of various sizes. Most are large. There are several times as many large balls as small balls. What is the number of small balls? What is the probability that a ball drawn at random is neither large nor small?

To enable PT to deal with problems of this kind, it is necessary to restructure probability theory by replacing bivalent logic on which it is based with fuzzy logic. The rationale for this replacement is that fuzzy logic is, in essence, the logic of perceptions, while bivalent logic is this logic of measurements.

The result of restructuring of PT is what may be called perception-based probability theory, PTp. Conceptually, mathematically and computationally, PTp is more general and more complex than PT. The basics of PTp are described in my recent paper," Toward a Perception-based Theory of Probabilistic Reasoning with Imprecise Probabilities," Journal of Statistical Planning and Inference, Elsevier Science, Vol. 105, 233-264, 2002. A key idea in PTp is that subjective probabilities are perceptions of likelihood, and as such are intrinsically imprecise.

Basically, PTp is the result of a three-stage generalization of PT. The first stage involves f-generalization and leads to PT^+, a probability theory in which probabilities, events and relations are, or are allowed to be, fuzzy-set-valued. For example, the probability of an event may be described as "very high," and a relation between X and Y may be defined as "X is much larger than Y."

The second stage involves f.g-generalization and leads to a probability theory labeled PT^{++}. In PT^{++}, probabilities, events and relations are, or are allowed to be, f-granular. For example, if Y is a function of X, Y=f(X), then f may be described as a collection of fuzzy if-then rules of the form: if X is A_i then Y is B_i, i=1,…,n, in which the A_i and B_i are fuzzy sets in the domains of X and Y, respectively.

The third stage involves nl-generalization and leads to perception-based probability theory PTp. It is PTp that has the capability to operate on perception-based information.

A concept which plays a key role in nl-generalization is that of Precisiated Natural Language (PNL). Basically, PNL consists of those propositions in a natural language, NL, which are precisiable through translation into a precisiation language. In the case of PNL, the precisiation language is the Generalized Constraint Language, GCL.

A key idea which underlies PNL is that the meaning of a proposition, p, in NL may be represented as a generalized constraint of the form X isr R, where X is the constrained variable, R is the constraining relation; and r is an indexing variable whose value defines the way in which R constrains X. In general, X, R and r are implicit in p. Thus, in PNL, representation of the meaning of p involves explicitation of X, R and r. As a very simple example, the meaning of "Eva is young," may be represented as "Age(Eva) is young," in which X=Age(Eva), R=young, and the constraint is possibilistic in the sense that "young" defines the fuzzy set of possible values of Age(Eva). The principal types of constrains are possibilistic (r=blank); probabilistic (r=p); veristic (r=v); and usuality (r=u). Thus, X isu R means that usually (X is R).

Within PTp, PNL plays an essential role in (a) representation of imprecise probabilities and probability distributions; (b) definition of basic concepts such as independence and stationarity; and (c) deduction from perception-based information.

An important example is what may be called bimodal information, that is, information which involves a mixture of possibilistic and probabilistic constraints. More specifically, if X is a random variable, its bimodal probability distribution may be represented as

$$X \text{ isp } (P_1|A_1 + P_2|A_2 + \ldots + P_n|A_n),$$

where $A_1,\ldots,A_n$ are fuzzy subsets of the domain of X, and $P_1,\ldots,P_n$ are fuzzy probabilities of $A_1,\ldots,A_n$, respectively. For example, if X=Age(Robert) and the domain of X is the interval [0,100], then

X=low|young + high|middle-aged + low|old

means that the probability that Robert is young is low; probability that Robert is middle-aged is high; and probability that Robert is old is low.

In PTp, the default assumption is that probability distributions are bimodal. A typical question is: What is the expected value of a bimodal distribution? In general, the expected value of a bimodal distribution is fuzzy-set-valued.

In conclusion, standard probability theory, PT, is subsumed by perception-based probability theory, PTp. An essential difference between PT and PTp is that in PT only likelihood is a matter of degree. In PTp, everything—and especially truth and possibility—is, or is allowed to be, a matter of degree. Conceptually, mathematically and computationally, PTp is more complex than PT. In this instance, as in many others, complexity is the price of constructing a probability theory which has a close rapport with the pervasive imprecision, uncertainty and ill-definedness of the real world.