INFORMATION SCIENCES **40**, 165–174 (1986)

Fuzzy Entropy and Conditioning

BART KOSKO

VERAC, Incorporated, 9605 Scranton Road, San Diego, California 92121

Communicated by Lotfi Zadeh

ABSTRACT

A new nonprobabilistic entropy measure is introduced in the context of fuzzy sets or messages. Fuzzy units, or *fits*, replace bits in a new framework of fuzzy information theory. An appropriate measure of entropy or fuzziness of messages is shown to be a simple ratio of distances: the distances between the fuzzy message and its nearest and farthest nonfuzzy neighbors. Fuzzy conditioning is examined as the degree of subsethood (submessagehood) of one fuzzy set or message in another. This quantity is shown to behave as a conditional probability in many contexts. It is also shown that the entropy of A is the degree to which $A \cup A^c$ is a subset of $A \cap A^c$, an intuitive relationship that cannot occur in probability theory. This theory of subsethood is then shown to solve one of the major problems with Bayes-theorem learning and its variants—the problem of requiring that the space of alternatives be partitioned into disjoint exhaustive hypotheses. Any fuzzy subsets will do. However, a rough inverse relationship holds between number and fuzziness of partitions and the information gained from experience. All results reduce to fuzzy cardinality.

I. INTRODUCTION

Entropy is uncertainty. It permeates discourse and systems. It connects deeply with information and conditioning (learning). And, in principle, it has nothing to do with probability theory. Below we develop a general entropy measure based on an intuitive ratio of distances. The measure quantifies the fuzziness of discourse and systems and connects naturally with conditioning.

II. FITS VERSUS BITS

We first develop apparatus needed to represent, and to think about, fuzzy sets as fuzzy messages.

Let $X = \{x_1, x_2, \ldots, x_n\}$ be a set. We represent subsets of X as *bit vectors* or *bivalent messages*. For instance, if $X = \{x_1, x_2, x_3\}$, then $X = (1,1,1)$, $\varnothing =$

52 Vanderbilt Ave., New York, NY 10017 0020-0255/86/$03.50

$(0,0,0)$, and the subset $A=\{x_1,x_3\}$ is represented as $A=(1,0,1)$. The power set 2^X of X is the set of all of X's subsets. There are 2^n possible messages defined on X (in 2^X). In the example, $2^X=\{(0,0,0),(1,0,0),(0,1,0),(0,0,1),(1,1,0),(0,1,1),(1,0,1),(1,1,1)\}$.

We represent fuzzy subsets of X as *fit vectors* or *fuzzy messages*. "Fit" seems appropriate because it contracts "fuzzy unit" in the way that "bit" contracts "binary unit" and because a fit value measures the degree to which an element x fits in or belongs to a subset A. We assume the fuzzy sets are unit-interval-valued, i.e., the *membership function* m_A of A is the mapping $m_A: X \to [0,1]$. By $F(2^X)$ we denote the *fuzzy power set* of X, the nonfuzzy set of all its fuzzy subsets. Hence when the range R of the map $m_A: X \to R$ is the unit interval, X has continuum-many fuzzy subsets. If R is a finite set of membership values, say m of them, then there are m^n possible fuzzy messages defined on X [in $F(2^X)$]. If $X=\{x_1,x_2,x_3\}$, two fuzzy messages are $A=(0.9,0,0.8)$ and $A^c=(0.1,1,0.2)$. Look at the "pathological" set $P=(0.5,0.5,\dots,0.5)$. It is the fuzziest set of all, since $P=P\cap P^c=P^c=P\cup P^c$. Here P is the set analogue of the glass that is half empty and half full.

Operations on fuzzy messages are performed pointwise: $A\cap B=(\min(m_A(x_1),\, m_B(x_1)),\dots,\min(m_A(x_n),\, m_B(x_n)))$, $A\cup B=(\max(m_A(x_1),\, m_B(x_1)),\dots,\max(m_A(x_n),\, m_B(x_n)))$, and $A^c=(1-m_A(x_1),\dots,1-m_A(x_n))$. More generally, conjunction and disjunction are replaced by triangular norm and conorm operations respectively [1,2].

Every fuzzy message can be uniquely decomposed into a weighted disjunction of nonfuzzy messages. This remarkable theorem was proved by Zadeh [3], in passing, in the context of fuzzy relations and graphs. The algorithm is simple (and the proof nearly by inspection). Associate a bit vector with every fit value $m_A(x_j)$ such that the ith bit value is 1 if $m_A(x_i)\geqslant m_A(x_j)$, 0 if not; then union or OR the weighted bit vectors together. Suppose $A=(0.2,0,0.6,1,0.6)$. Then $A=0(1,1,1,1,1)\vee 0.2(1,0,1,1,1)\vee 0.6(0,0,1,1,1)\vee 1(0,0,0,1,0)$. Hence $A=X$ with 0 weight, $A=\{x_3,x_4,x_5\}$ with 0.6 weight, and $A=\{x_4\}$ with maximal weight. The higher the weight, the more reliable the bivalent message. A fuzzy message is a weighted superimposition of nonfuzzy messages. All we need now are fuzzy sampling and fuzzy channel capacity theorems!

The *l^p-distance* between fuzzy messages A and B is defined by $l^p(A,B)=(\Sigma_i|m_A(x_i)-m_B(x_i)|^p)^{1/p}$. In particular, l^1 is *fuzzy Hamming distance*. Kaufmann [4] first pointed out that every fuzzy message A has a *nearest nonfuzzy message* $\bar{A}$ (for any l^p metric) given by $m_{\bar{A}}(x_i)=1$ if $m_A(x_i)\geqslant 0.5$, 0 if $m_A(x_i)\leqslant 0.5$. (0.5 values are equidistant; hence $\bar{A}$ is not unique.) $\bar{A}$ is easier found than described. If $A=(0.2,0.6,0,0.9,1,0.6,0.7)$, then $\bar{A}=(0,1,0,1,1,1,1)$. Note that if $P=(0.5,0.5,\dots,0.5)$, then $\bar{P}$ can be *any* nonfuzzy message (element of 2^X). P is equidistant from them all.

It seems to have gone unnoticed that every fuzzy message A has a *farthest nonfuzzy message* $\underline{A}$. $\underline{A}$ is given by $m_{\underline{A}}(x_i) = 0$ if $m_A(x_i) \geqslant 0.5, 1$ if $m_A(x_i) \leqslant 0.5$. Hence $\underline{A} = (\overline{A})^c$. From the above example, $\underline{A} = (1,0,1,0,0,0,0)$. Note again that $\underline{P}$ can also be any nonfuzzy message.

The cardinality, or size, of a fuzzy message is the sum of its fit values. Zadeh [5, 6] calls this sum a *sigma-count*: $\Sigma\text{Count}(A) = \Sigma_i m_A(x_i)$. If $A = (0.9,0,0.8)$, then $\Sigma\text{Count}(A) = 1.7$, and thus fuzzy cardinality can be a real number as well as an integer (see Dubois and Prade [7] for extensions). Note that $\Sigma\text{Count}(A) = l^1(A, \varnothing)$. Hence the fuzzy cardinality of A is its Hamming distance to the null message $\varnothing$. Kosko [8] has shown that the sigma-count is a positive measure that generalizes classical counting measure in combinatorics.

III. FUZZY ENTROPY

De Luca and Termini [9] first axiomatized nonprobabilistic entropy. They obtain probabilistic (Shannon) entropy as a special case in special circumstances. The upshot is that entropy equals fuzziness, and entropy equals information. The De Luca–Termini axioms are intuitive and have stuck in the fuzzy literature; we adopt them here.

Let E be a set-to-point map $E: F(2^X) \to [0,1]$. Hence E is a fuzzy set defined on fuzzy sets. E is an *entropy measure* if it satisfies the four De Luca–Termini axioms:

(DT1) $E(A) = 0$ iff $A \in 2^X$ (A nonfuzzy),
(DT2) $E(A) = 1$ iff $m_A(x_i) = 0.5$ for all i,
(DT3) $E(A) \leqslant E(B)$ if A is less fuzzy than B, i.e., if $m_A(x) \leqslant m_B(x)$ when $m_B(x) \leqslant 0.5$ and $m_A(x) \geqslant m_B(x)$ when $m_B(x) \geqslant 0.5$,
(DT4) $E(A) = E(A^c)$.

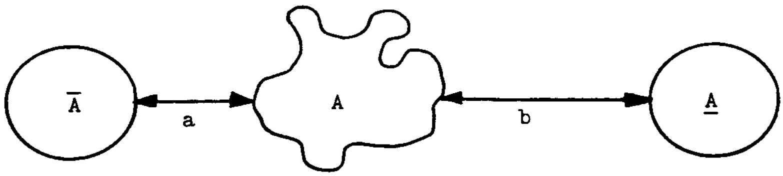
Motivated by the classical Shannon entropy function, De Luca and Termini propose the parametrized entropy measures $E_k(A) = D_k(A) + D_k(A^c)$. Here $k > 0$ and $D_k(A) = -k\Sigma_i m_A(x_i)\log m_A(x_i)$.

Most fuzzy entropy measures focus on the intersection $A \cap A^c$, since $A \cap A^c \neq \varnothing$ iff A is fuzzy (iff $A \cup A^c \neq X$). For instance, Yager [10] proposes

$$Y_p(A) = 1 - \frac{l^p(A, A^c)}{[l^p(A, \varnothing)]^p} \qquad \text{for} \quad p \geqslant 1.$$

Similarly, exploiting the distance from A to $\overline{A}$, A's nearest nonfuzzy message, Kaufmann [4] proposes $K_p(A) = (2/n^{1/p})l^p(A, \overline{A})$ for $p \geqslant 1$, where the normalizing constant stems from the metrical fact that $0 \leqslant l^p(A, B) \leqslant n^{1/p}$.

Let us propose another approach. Consider two distances: between A and $\overline{A}$ and between A and $\underline{A}$, A's farthest nonfuzzy neighbor. Let $a = l^p(A, \overline{A})$ and $b = l^p(A, \underline{A})$:



Note that if A is nonfuzzy, then $\overline{A} = A$ and hence $a = 0$. Now suppose $A = P = (0.5, 0.5, \ldots, 0.5)$. Then A is equidistant to all nonfuzzy sets. Hence $a = b$. Finally, note for A^c that $l^p(A^c, \underline{A}) = a$ and $l^p(A^c, \overline{A}) = b$.

These intuitions suggest taking the *ratio* a/b as the fuzzy entropy:

$$R_p(A) = \frac{l^p(A, \overline{A})}{l^p(A, \underline{A})} \qquad \text{for} \quad p \geqslant 1.$$

The same intuitions lead at once to a verification that R_p satisfies the De Luca–Termini axioms (DT1)–(DT4), giving Theorem 1:

THEOREM 1. *R_p is a fuzzy entropy measure.*

A surprising and intuitively satisfying fact is that R_1 (and in general R_p) is a ratio of sigma-counts involving only A and A^c (the ratio of sigma-counted underlap $A \cap A^c$ to "overlap" $A \cup A^c$). The result is a generalized form of the Tanimoto similarity measure discussed by Kohonen [11].

THEOREM 2.

$$R_1(A) = \frac{\Sigma \mathrm{Count}(A \cap A^c)}{\Sigma \mathrm{Count}(A \cup A^c)}.$$

Proof. Proof by cases yields the equality $|m_A - m_B| = \max(m_A, m_B) - \min(m_A, m_B)$. We use this fact to show $l^1(A, \overline{A}) = \Sigma\mathrm{Count}(A \cap A^c)$; the proof that $l^1(A, \underline{A})$ equals the denominator is symmetric. Suppose $m_A(x) \leqslant m_{\overline{A}}(x) = 1$. Then $m_A \geqslant \frac{1}{2}$ and thus $m_{A^c} \leqslant \frac{1}{2}$. Then $|m_A - m_{\overline{A}}| = 1 - m_A = m_{A^c} = \min(m_A, m_{A^c})$, using the above equality. Now suppose $m_A(x) \geqslant m_{\overline{A}}(x) = 0$. Then $m_A \leqslant \frac{1}{2}$ and thus $m_{A^c} \geqslant \frac{1}{2}$. Then $|m_A - m_{\overline{A}}| = m_A - 0 = \min(m_A, m_{A^c})$.

Hence taking the sum of these absolute values to obtain $l^1(A, \bar{A})$ amounts to finding the sigma-count of $A \cap A^c$. ■

Hence entropy reduces to cardinality (as does everything else!). Two further theorems are useful. First, computing $R_1(A)$ can be simplified by noting that $\Sigma\text{Count}(A \cup A^c) = n - \Sigma\text{Count}(A \cap A^c)$. Second, $R_p(A) \leqslant K_p(A)$, proven with $|m_A - m_{\underline{A}}|^p = |m_{A \cup A^c}|^p$.

IV. FUZZY CONDITIONING AS FUZZY SUBSETHOOD

Conditioning is learning. It is updating, making one thing a function of another, passing from one fuzzy message to another. And this is subsethood. In this section we develop a general method for measuring fuzzy-message conditioning and show its relationship to fuzzy entropy. Formally, we seek and obtain a power-set function $S(A, B)$ that "properly" measures the *degree of subsethood* of A in B.

What characterizes subsethood? When Zadeh [12] formulated fuzzy set theory, he abstracted a relation among membership functions: $A \subset B$ iff $m_A(x) \leqslant m_B(x)$ for all x. This certainly generalizes nonfuzzy subsethood, for instance when $A = (1,0,0,0,1) = \{x_1, x_5\}$ and $B = (1,0,1,0,1) = \{x_1, x_3, x_5\}$, allowing us to state unambiguously that $C = (0.2,0,0.1,0.1,0.4,0.3,0.4,0.2,0.5)$ is a subset (*submessage*) of $D = (0.3,1,1,0.5,0.6,0.3,0.7,0.8,0.5)$. And this is consistent with the truth-tabular definition of bivalent implication wherein a single 1-0 pair falsifies the implication. But what if there is a violation of $m_A \leqslant m_B$ in a single fit-value slot? Should not $C' = (0.2, 0, 0.1, 0.1,0.4,0.3,0.4,0.2,0.6)$ still be a submessage of D to a very high degree? And surely the size $[m_A(x_v) - m_B(x_v) > 0]$ and relative frequency of violations should affect the measure. In short, how does one measure degrees of conditioning?

A rigorous but nonoperational answer was given by Bandler and Kohout [13]. They observed that, with nonfuzzy sets, A is a subset of B iff $A \in 2^B$. So A is a fuzzy subset of B iff $A \in F(2^B)$. Hence we have identified $S(A, B)$: $S(A, B) = m_{F(2^B)}(A)$, the degree of membership of A in B's fuzzy power set. But, again, how do we measure it?

A "good" $S(A, B)$ should measure violations of $m_A \leqslant m_B$. It should measure the violations in magnitude, number, and proportion. A little thought shows that the simplest measure is $\Sigma_i \max(0, m_A(x_i) - m_B(x_i))$, normalized by $\Sigma\text{Count}(A)$. Since this ratio measures violations (supersethood), we negate it and arrive at $S(A, B)$:

$$S(A, B) = 1 - \frac{\Sigma_i \max(0, m_A(x_i) - m_B(x_i))}{\Sigma\text{Count}(A)}.$$

There are three properties of this selection that are compelling, even deep. First, it subsumes Zadeh's notion of subsethood, since if $m_A \leqslant m_B$ for all x, then $\max(0, m_A - m_B) = 0$ for all x and thus $S(A, B) = 1$. Second, it connects naturally with fuzzy logical implication. Ignoring the sum over i and the normalizing cardinality—recalling that A is a subset of B if $x \in A$ implies $x \in B$ for all x—we get

$$\begin{aligned} 1 - \max(0, m_A - m_B) &= 1 - [1 - \min(1, 1 - (m_A - m_B))] \\ &= \min(1, 1 - m_A + m_B) \\ &= t_L(A \to B), \end{aligned}$$

the Łukasiewicz implication operator for continuous-valued logic (truth), where we have used the max-min De Morgan's law in the first equality. In fuzzy logic the Łukasiewicz operator is by far the most popular selection for logical implication; it is often taken as a default implication operator, just as min and max are taken as default conjunction and disjunction operators. Third, this selection for $S(A, B)$ naturally reduces to cardinality and looks like a conditional probability. This surprising result is important enough to state as a theorem.

THEOREM 3.

$$S(A, B) = \frac{\Sigma\text{Count}(A \cap B)}{\Sigma\text{Count}(A)}.$$

Proof. $\Sigma\text{Count}(A) = 0$ iff $A = \varnothing$, and $S(\varnothing, B) = 1$. So we can assume $\Sigma\text{Count}(A) > 0$. Then the theorem follows from the equality

$$\Sigma\text{Count}(A \cap B) = \Sigma\text{Count}(A) - \sum_i \max(0, m_A(x_i) - m_B(x_i)).$$

This equality is obtained by summing over the following equality:

$$\min(m_A(x_i), m_B(x_i)) = m_A(x_i) - \max(0, m_A(x_i) - m_B(x_i)).$$

And this equality is quickly verified by checking the cases $m_A(x_i) \leqslant m_B(x_i)$ and $m_A(x_i) > m_B(x_i)$. ■

For convenience's sake, from now on we will use the ratio in Theorem 3 to define $S(A, B)$.

Another surprise is that fuzzy entropy reduces to fuzzy conditioning. The entropy of the fuzzy message A is the extent to which $A \cup A^c$ is a submessage of $A \cap A^c$, a state of affairs that can only occur in probability theory if $A = X = \varnothing$! We state this intuitive fact as a theorem for $p = 1$ (the general p-dimensional case is immediate).

THEOREM 4.

$$R_1(A) = S(A \cup A^c, A \cap A^c).$$

Proof. The result follows from Theorem 2 and Theorem 3 and the general equality for fuzzy sets $A \cap B = (A \cap B) \cap (A \cup B)$. ■

V. FUZZY PARTITIONING AND LEARNING

Bayes's theorem remains the premier learning theorem. It motivated much of Kalman filtering theory in signal processing and, with the elaborations of Dempster [14, 15] and Shafer [16], it holds sway in artificial-intelligence R&D management of uncertainty in expert systems. (Curiously enough, its much-maligned rival, the perceptron learning theorem [17, 18], has recently motivated a powerful new learning theorem known as the Cohen-Grossberg theorem in artificial neural systems (ANS) [19, 20] that is the current "premier" theorem for Hopfield networks [21, 22] and other adaptive networks and associative memories characterized by a Lyapunov or energy function.) But Bayes's theorem has serious problems, and the present fuzzy theory can help.

Bayes's theorem rests on an exact partitioning of the universe of discourse. The hypothesis space X is carved into n-many disjoint exhaustive hypothesis sets H_i: $X = \bigcup_i^n H_i$, and $H_i \cap H_j = \varnothing$ if $i \neq j$. A piece of evidence E is any subset $E \subset X$. Bayes's theorem then relates the probability $P(H_i|E)$ that hypothesis H_i is true given evidence E to the converse probability $P(E|H_i)$ that the evidence E is observed given H_i is true, the prior probability $P(H_i)$ that H_i is true, and the resultant probability $P(E)$ that the evidence is observed:

$$P(H_i|E) = \frac{P(E|H_i)\,P(H_i)}{P(E)} = \frac{P(E|H_i)\,P(H_i)}{\sum_{j=1}^{n} P(E|H_j)\,P(H_j)}.$$

An immediate problem with Bayes's theorem is finding the prior probabilities $P(H_i)$. If these can be generated, why not just generate the $P(H_i|E)$ as well? The expert-systems analogue of this question is: where do the uncertainty numbers come from? An expert says 30% one day and 60% the next, etc. Growing pessimism over this problem leads many AI researchers (see Doyle [23] and Cohen and Sullivan [24]) to reject quantitative uncertainty modeling altogether. The present theory escapes these objections only insofar as there is no requirement that $\Sigma_i^n \Sigma\text{Count}(H_i) = 1$.

The problem we address is inexact partitioning. The assumption of well-defined disjoint exhaustive hypotheses $H_1, \ldots, H_n$ is unrealistic. It occurs in practice typically only for $n = 2$. It's hard to carve up the universe into nonfuzzy pieces. Alternatives overlap, boundaries blur. In general the hypotheses H_i are fuzzy subsets of X. In general they have no more structure than that. Rather than defuzzify these sets and make a partition out them to fit Bayes's theorem, we can deal with them in general as is by using sigma-counts and degrees of subsethood. In fact Zadeh [5, 6] calls $S(A, B)$ the *relative sigma-count* (defining it directly by the relation of Theorem 3) and treats it in many contexts as a conditional probability, even though $S(A, B) + S(A, B^c) \geqslant 1$. The following theorem, a fuzzy Bayes's theorem of sorts, assumes only that E and the H_i are fuzzy subsets of X, and reduces to Bayes's theorem in the very special case where the subsets are nonfuzzy and the H_i partition X. Roughly speaking, the theorem and corollary say that the fuzzier the partition, the less is learned from experience. Put another way, since the fuzzy overlap between hypotheses increases with the number of hypotheses, you pay for partitioning.

THEOREM 5

$$S(E, H_i) = \frac{S(H_i, E)\,\Sigma\text{Count}(H_i)}{\frac{1}{n}\sum_{j=1}^{n}\left(\Sigma\text{Count}(E \cap H_j) + \Sigma\text{Count}(E \cup H_j) - \Sigma\text{Count}(H_j)\right)}.$$

Proof. Clearly $m_E(x) + m_{H_j}(x) = \min(m_E(x), m_{H_j}(x)) + \max(m_E(x), m_{H_j}(x))$ holds for all j. Now sum over all $x \in X$ (recalling that n is the number of hypotheses, not the cardinality of X). This gives

$$\Sigma\text{Count}(E) + \Sigma\text{Count}(H_j) = \Sigma\text{Count}(E \cap H_j) + \Sigma\text{Count}(E \cup H_j).$$

Now sum both sides over j. Manipulation then gives $\Sigma\text{Count}(E)$ as the expression in the denominator of Theorem 5. The rest proceeds as in the proof of Bayes's theorem: expand $S(E, H_i)$ using Theorem 3, replace the denominator with the above expression, and replace the numerator with $\Sigma\text{Count}(E \cap H_i) = \Sigma\text{Count}(H_i \cap E) = S(H_i, E)\Sigma\text{Count}(H_i)$. ∎

By turning the equality in the above proof into two inequalities, lower and upper bounds can be put on $S(E, H_i)$. The resultant corollary is notationally closer to Bayes's theorem.

COROLLARY.

$$\frac{S(H_i, E)\Sigma\text{Count}(H_i)}{\frac{1}{n}\sum_{j=1}^{n}\Sigma\text{Count}(E \cup H_j)} \leqslant S(E, H_i)$$

$$\leqslant \frac{S(H_i, E)\Sigma\text{Count}(H_i)}{\frac{1}{n}\sum_{j=1}^{n} S(H_j, E)\Sigma\text{Count}(H_j)}.$$

Note that if the denominator of the first term in the Corollary is multiplied and divided by $\Sigma\text{Count}(H_j)$, then the two bounding ratios differ in a natural way—the lower bound is normalized by a sum of unions, the upper bound by a sum of intersections.

As an example, let $X = (x_1, \ldots, x_6)$, $H_1 = (0.4, 0.2, 0.1, 0, 0, 0)$, $H_2 = (0, 0.2, 0.6, 0.3, 0.1, 0)$, $H_3 = (0, 0.1, 0, 0.6, 0.6, 0.5)$, and $E = (0, 0.2, 0.6, 0.2, 0, 0)$. Then Theorem 5 gives $S(E, H_1) = 0.3$, $S(E, H_2) = 1$, and $S(E, H_3) = 0.3$, as can be checked by evaluating the degrees of subsethood directly with Theorem 3 and noting that $m_E \leqslant m_{H_2}$.

A final note concerns the two-hypothesis ($n = 2$) case. Suppose that H_1 and H_2 are H and H^c, as occurs in practice. Then, using the fact that $\Sigma\text{Count}(E) \leqslant \Sigma\text{Count}(E \cap H) + \Sigma\text{Count}(E \cap H^c)$, we pay for the fuzziness involved by getting back an inequality instead of equality in the familiar Bayes relationship:

$$S(E, H) \geqslant \frac{S(H, E)\Sigma\text{Count}(H)}{S(H, E)\Sigma\text{Count}(H) + S(H^c, E)\Sigma\text{Count}(H^c)}.$$

Another generalization uses the triangular norm product instead of min to measure intersection: $m_{A \cap B} = m_A m_B$. (Then the corresponding triangular co-norm to measure union is the "probabilistic" conorm: $m_{A \cup B} = m_A + m_B - m_A m_B$.) We then get back equality in the above relationship despite the fuzzy sets involved, but only if $n < 3$. This is because $m_E m_H + m_E m_{H^c} = m_E m_H + m_E(1 - m_H) = m_E$. This illustrates again the tradeoff between possibilities (partitions) and information, and why theories with fewer and crisper variables are easier to test.

The author wishes to thank David Brown of TRW MEAD, Professor Clark Guest of UC San Diego's EECS Department, and Robert Sasseen of VERAC for several long and fruitful discussions on fuzzy entropy and conditioning.

REFERENCES

1. E. P. Klement, Operations on fuzzy sets and fuzzy numbers related to triangular norms, in *Proceedings of the 11th International Symposium on Multi-Valued Logic*, 1981, pp. 218–225.
2. H. Prade, A computational approach to approximate and plausible reasoning with applications to expert systems, *IEEE Trans. Pattern Anal. and Machine Intelligence* 7(3):260–283 (1985).
3. L. A. Zadeh, Similarity relations and fuzzy orderings, *Inform. Sci.* 3:177–206 (1971).
4. A. Kaufmann, *Introduction to the Theory of Fuzzy Subsets*, Vol. I, Academic, New York, 1975.
5. L. A. Zadeh, A computational approach to fuzzy quantifiers in natural languages, *Comput. Math. Appl.* 9(1):149–184 (1983).
6. L. A. Zadeh, The role of fuzzy logic in the management of uncertainty in expert systems, *Fuzzy Sets and Systems* 11:199–227 (1983).
7. D. Dubois and H. Prade, Fuzzy Cardinality and the modeling of imprecise quantification, *Fuzzy Sets and Systems* 16(3):199–235 (Aug. 1985).
8. B. Kosko, Counting with fuzzy sets, *IEEE Trans. Pattern Anal. and Machine Intelligence*, vol. PAMI-8, no. 4, 556–557, July 1986.
9. A. De Luca and S. Termini, A definition of a nonprobabilistic entropy in the setting of fuzzy sets theory, *Inform. and Control* 20:301–312 (1972).
10. R. R. Yager, On the measure of fuzziness and negation, Part I: Membership in the unit interval, *Internat. J. Gen. Systems* 5:221–229 (1979).
11. T. Kohonen, *Self-Organization and Associative Memory*, Springer, Berlin, 1984.
12. L. A. Zadeh, Fuzzy sets, *Inform. and Control* 8:338–353 (1965).
13. W. Bandler and L. Kohout, Fuzzy power sets and fuzzy implication operators, *Fuzzy Sets and Systems* 4:13–30 (1980).
14. A. P. Dempster, Upper and lower probabilities induced by a multivalued mapping, *Ann. Math. Statist.* 38:325–339 (1967).
15. A. P. Dempster, A generalization of Bayesian inference, *J. Roy. Statist. Soc. Ser. B* 30:205–247 (1968).
16. G. Shafer, *A Mathematical Theory of Evidence*, Princeton U.P., (Princeton, N.J., 1976.
17. F. Rosenblatt, *Principles of Neurodynamics*, Spartan, New York, 1962.
18. M. Minsky and S. Papert, *Perceptrons*, M.I.T. Press, Cambridge, Mass., 1969.
19. M. A. Cohen and S. Grossberg, Absolute stability of global pattern formation and parallel memory storage by competitive neural networks, *IEEE Trans. Systems, Man, and Cybernet.* SMC-13:815–826 (Oct. 1983).
20. R. Hecht-Nielsen, Performance Limits of Optical, Electro-Optical, and Electronic Artificial Neural System Processors, *Proc. SPIE: Hybrid and Optical Systems*, H. Szu, Ed., 1986.
21. J. J. Hopfield, Neural networks and physical systems with emergent collective computational abilities, *Proc. Nat. Acad. Sci. U.S.A.* 79:2554–2558 (1982).
22. Y. S. Abu-Mostafa and J. St. Jacques, Information capacity of the Hopfield model, *IEEE Trans. Inform. Theory* IT-31 (4):461–464 (July 1985).
23. J. Doyle, Methodological simplicity in expert system construction: The case of judgments and reasoned assumptions, *AI Mag.* 3(2):39–43 (1984).
24. P. R. Cohen and M. Sullivan, An endorsement-based plan recognition program, in *Proceedings of the IJCAI-85*, Los Angeles, Aug. 1985, pp. 475–479.

Received 12 May 1986; revised 2 June 1986