Uncertainty Representation Using Fuzzy Measures

Ronald R. Yager, *Fellow, IEEE*

***Abstract*—We introduce the fuzzy measure and discuss its use as a unifying structure for modeling knowledge about an uncertain variable. We show that a large class of well-established types of uncertainty representations can be modeled within this framework. A view of the Dempster–Shafer (D–S) belief structure as an uncertainty representation corresponding to a set of possible fuzzy measures is discussed. A methodology for generating this set of fuzzy measures from a belief structure is described. A measure of entropy associated with a fuzzy measure is introduced and its manifestation for different fuzzy measures is described. The problem of uncertain decision making for the case in which the uncertainty represented by a fuzzy measure is considered. The Choquet integral is introduced as providing a generalization of the expected value to this environment.**

***Index Terms*—Decision making, entropy, fuzzy measure, uncertainty.**

I. Introduction

THE MULTIPLICITY of modalities associated with the kinds of information that can appear in the current generation of intelligent systems requires a wide spectrum of uncertainty representation calculi. Among those commonly used are probability theory, possibility theory, fuzzy sets, rough sets, Dempster–Shafer (D–S), and random set theory. Close connections exist between some of these, as is the case with D–S and random set theory [1], as well as between fuzzy set theory and possibility theory. These calculi rather then being competitive are needed to represent the different types of uncertainties such as randomness, lack of specificity, and imprecision. As a result of this situation, there arises a need for a unified framework in which to model the above-mentioned calculi and any other required type of uncertainty representations. A promising unifying framework for this task is the class of monotonic nonadditive measures called *fuzzy measures* [2]–[4] which have properties very suitable for the representation and management of uncertain information. In the following, we shall discuss some ideas in this direction.

II. Fuzzy Measures and the Representation of Uncertainty

Assume V is a variable which attains its value in the space $X = \{x_1, x_2, \ldots, x_n\}$. In situations in which the exact value of the variable V is unknown, the best we can do is to try to formulate our knowledge about V in a useful mathematical structure. One such structure is a fuzzy measure [2]–[4]. One useful feature of this measure is its ability to represent in a unified manner different types of characterizations of uncertainty. A second desirable feature of the fuzzy measure is its ability to interact with integrals, such as the Choquet [5], [6] and fuzzy integral [7], to provide the necessary tools to aid in decision making under uncertainty.

Formally a fuzzy measure μ on a space X is a mapping from subsets of X into the unit interval μ: $2^X \to [0, 1]$ such that $\mu(X) = 1$, $\mu(\emptyset) = 0$, and if $A \subset B$ then $\mu(A) \le \mu(B)$.

Within the framework of using the fuzzy measures to represent information about an uncertain variable, $\mu(E)$ can be interpreted as a measure associated with our belief that the value of V is contained in the subset E. We shall generically refer to this measure as the *confidence* we have that $V \in E$.

It is clear that the properties of μ are in accordance with this interpretation. Since we are completely confident that V lies in the space X, we have $\mu(X) = 1$. Since we are also certain that the value of V does not lie in the null set, we have $\mu(\emptyset) = 0$. The third condition, called the monotonicity property, is a reflection of the fact that we cannot be more confident that V lies in a smaller set than in a larger one.

Other than satisfying these three requirements, we are free to define the fuzzy measure. Thus, the fuzzy measure provides a framework for expressing a large class of structures useful for expressing our knowledge of uncertain variables. Classes of fuzzy measures which involve a simplification of the process of obtaining the value of $\mu(E)$ are of considerable interest. In some cases, the measure has been given a very specific name. In the following, we shall describe some of these special situations.

The situation in which we know precisely the value of V, $V = x$, is expressed by the measure μ_x in which

$$\mu_x(A) = 1 \qquad \text{for } x \in A$$
$$\mu_x(A) = 0 \qquad \text{for } x \notin A.$$

Murofushi and Sugeno [4] call such a measure a Dirac measure focused at x.

A closely related situation is the measure μ in which

$$\mu(A) = 1 \qquad \text{for } A\text{'s s.t. } B \subseteq A$$
$$\mu(A) = 0 \qquad \text{for } A\text{'s s.t. } B \not\subset A.$$

Here, we are expressing the knowledge that we know the value of V lies in B.

The case in which our knowledge about the unknown variable is of the form of probabilistic uncertainty can be represented by a fuzzy measure μ in which

$$\mu(A \cup B) = \mu(A) + \mu(B) \qquad \text{for } A \cap B = \emptyset.$$

Here, we are assuming additivity for disjoint sets. In this situation, μ is called a probability measure and often $\mu(A)$ is denoted as $\mathrm{Prob}(A)$, indicating the probability that V lies in the subset

Manuscript received May 15, 2001; revised July 20, 2001 and October 15, 2001. This paper was recommended by Associate Editor E. Santos.

The author is with the Machine Intelligence Institute, Iona College, New Rochelle, NY 10801 USA (e-mail: yager@panix.com).

Publisher Item Identifier S 1083-4419(02)00707-0.

A. This additive measure is completely characterized by the n values $p_i = \mu(\{x_i\})$, denoting the probability that V equals x_i. Since $\mu(A) = \sum_{x_i \in A} p_i$, it is required that $\sum_{i=1}^{n} p_i = 1$. Two important special cases of probability distribution are worth pointing out. The first special case is one in which V is known to be equal to some value $x_{\hat{i}}$. In this case, we get $\mu_{x_{\hat{i}}}$, which implies $p_{\hat{i}} = 1$ and $p_j = 0$ for $j \neq \hat{i}$. The second special case occurs when we have no information to distinguish between the x_i, $p_i = p_j$ for all i and j. From the requirement that $\sum_{j=1}^{n} p_j = 1$, we get that $p_j = 1/n$ for all j. We note in this case $\mu(A) = \sum_{x_i \in A} p_i = \mathrm{Card}[A]/n$. We shall denote this measure as μ_N.

Another type of uncertainty is possibilistic uncertainty [8]. Possibilistic uncertainty [9], [10] is characterized by a fuzzy measure in which $\mu(A) = \mathrm{Max}[\mu(A), \mu(B)]$. In this situation, μ is called a possibilistic measure and is often denoted $\mathrm{Poss}(A)$, indicating the possibility that V lies in the subset A. This type of measure is completely characterized by the n values $\alpha_i = \mu(\{x_i\})$ denoting the possibility that V equals x_i. For this measure $\mu(E) = \mathrm{Max}_{x_i \in E} \mu(\{x_i\})$. The mapping π: $X \to [0, 1]$ where $\pi(x_i) = \mu(\{x_i\})$ is called the possibility distribution and plays a role in possibility theory analogous to that played by the probability distribution. Since $\mu(A) = \mathrm{Max}_{x_j \in A}[\alpha_j]$ and $\mu(X) = 1$, it is required that at least one $\alpha_j = 1$. The case in which V is known to be equal to $x_{\hat{i}}$ again reduces to $\mu_{x_{\hat{i}}}$. On the other hand, for a possibility distribution, if we have no information to distinguish between the x_i, $\alpha_i = \alpha_j$, then since $\mathrm{Max}_j[\alpha_j] = 1$, we obtain that $\alpha_j = 1$ for all j in this case of maximal uncertainty. In this case, $\mu(\varnothing) = 0$ and $\mu(A) = 1$ for all $A \neq \varnothing$. We shall denote this measure as μ^*. Possibilistic uncertainty is often generated from fuzzy subsets used to represent linguistic information [11], [12]. For example, if V is the variable corresponding to John's height and our knowledge of John's height is that he is *tall*, then the fuzzy subset *tall* is seen to induce a possibility distribution in which the membership grade of an element becomes its possibility.

Another class of uncertainty measures is necessity measures. A fuzzy measure is called a necessity measure [9] if it satisfies the property that

$$\mu(A \cap B) = \mathrm{Min}[\mu(A), \mu(B)].$$

In this case, we denote $\mu(A)$ as $N(A)$. If we denote $F_i = X - \{x_i\}$, then a necessity measure is completely characterized by the n values $\mu(F_i)$, which we shall denote as β_i. Specifically, we see that in this case of a necessity measure $\mu(A) = \mathrm{Min}_{x_j \in \overline{A}}[\mu(F_j)] = \mathrm{Min}_{x_j \in \overline{A}}[\beta_j]$. Since $\mu(\varnothing) = \mathrm{Min}_{x_j \in X}[\beta_j]$ and the fact that $\mu(\varnothing) = 0$ requires that at least one $\beta_j = 0$. The situation in which V is known to be equal to $x_{\hat{j}}$ is represented by the measure $\mu_{x_{\hat{j}}}$ which is also a necessity measure in which $\beta_{\hat{j}} = 0$ and $\beta_j = 1$ for all $j \neq \hat{j}$. In the case of maximal uncertainty for the necessity measure, when no distinction exists between the x_j, we have $\beta_j = \beta_i$ for all j and i. Since at least one $\beta_j = 0$, this requires $\beta_j = 0$ for all j. In this case, our measure is $\mu(X) = 1$ and $\mu(A) = 0$ for all $A \neq X$. We shall denote this as μ_*.

Another class of fuzzy measures are those in which our certitude that the value of a variable lies in a subset is based upon the number of elements in the subset [13]. We shall refer to these as cardinality-based measures. For the cardinality-based measure we define

$$\mu(E) = D_{\mathrm{Card}(E)}.$$

Here, the measure is completely defined by the collection of n values D_j. Because of the monotonicity required of fuzzy measures $D_j \geq D_i$ for $j > i$. Furthermore, it is required that $D_0 = 0$ and $D_n = 1$. These cardinality measures can be defined in terms of a function f: $[0, 1] \to [0, 1]$ called a binary unit-interval monotonic (BUM) function having the properties $f(0) = 0$, $f(1) = 1$ and $f(x) \geq f(y)$ if $x > y$. Using this type of function, we define $\mu(E) = f(\mathrm{Card}(E)/n)$.

These cardinality-based measures can also be expressed in terms of a set of weights w_j, $j = 1$ to n, where $w_j \in [0, 1]$ and $\sum_{j=1}^{n} w_j = 1$. Using these weights $\mu(E) = \sum_{j=1}^{\mathrm{Card}(E)} w_j$. These weights are related to the D_j by $w_j = D_j - D_{j-1}$.

One important special case of cardinality-based measures occurs when $D_j = j/n$; in this case $\mu(E) = \mathrm{Card}(E)/n$. It is noted that this measure is the same as the one used to represent lack of information in the case of probabilistic uncertainty μ_N. Another special case is where $D_0 = 0$ and $D_j = 1$ for all other j. It is noted that this measure is the same as the one used to represent lack of information in the case of possibilistic uncertainty μ^*. Another special case of cardinality-based measure is one having $D_n = 1$ and $D_j = 0$ for all other j; this measure is the same as the one used to represent lack of information in the case of the necessity measure μ_*.

Complete uncertainty and complete certainty about the value of a variable are notable special cases with respect to our knowledge about a variable. One characterization of complete uncertainty about the value of a variable is the fuzzy measure μ_* in which $\mu_*(X) = 1$ and $\mu_*(E) = 0$ for $E \neq X$. Here, we are expressing our complete uncertainty in a very conservative way; we are only confident the value is in X. Another characterization of complete uncertainty is the fuzzy measure μ^* in which $\mu^*(\varnothing) = 0$ and $\mu^*(E) = 1$ for all $E \neq \varnothing$. Here, we are expressing our complete uncertainty in a very optimistic way; we are completely confident it lies in every set. As we have noted, these measures are special cases of cardinality-based measures. More generally, the cardinality-based measures provide a family of measures which enables us to represent situations in which we are unable to make any distinction between the possible outcomes. In this regard, it is interesting to note that the Dirac measure, which used to represent the knowledge that V is equal to a particular value, while a special case of probability, possibility, and necessity, is not a cardinality-based measure.

We now consider an extension of the cardinality-based measure. One example is a measure μ characterized by a BUM function f and a collection of n values $q_i \in [0, 1]$. For this measure, we define $\mu(A) = f(\sum_{j s.t. x_j \in A} q_j / \sum_{j=1}^{n} q_j)$. For this measure $\mu(\{x_i\}) = f(q_i / \sum_{j=1}^{n} q_j)$. For this measure, if $q_j = q_i$ for all j and j then $\mu(A) = f(\mathrm{Card}(A)/n)$, which is a pure cardinality-based measure. If f is linear $f(r) = r$, then this becomes a probability measure with $p_i = q_i / \sum_{j=1}^{n} q_j$. Thus, we see that this class provides a mixture of cardinality-based and probability measures. These types of measures have found consider use in the representation of many applications in risk analysis, economics, and cognitive science [14]–[16].

III. Features of Fuzzy Measures

We shall now discuss some general features that can be associated with fuzzy measures used to represent uncertain information.

Definition: Let μ_1 and μ_2 be any two fuzzy measures. If $\mu_1(A) \geq \mu_2(A)$ for all A, we shall denote this as $\mu_1 \geq \mu_2$. In the context of using the fuzzy measure to represent uncertainty if $\mu_1 \geq \mu_2$, we can say μ_1 displays more certitude or μ_1 is a more generous measure than μ_2. Alternatively, we can say μ_1 displays more confidence.

While in general the $\geq$ relationship over the fuzzy measures is not complete, it can be easily shown that for any μ, $\mu^* \leq \mu \geq \mu_*$.

If μ and $\hat{\mu}$ are two possibility measures with distributions α_j and $\hat{\alpha}_j$, then if $\alpha_j \geq \hat{\alpha}_j$ for all j we have $\mu \geq \hat{\mu}$. On the other hand, if μ and $\hat{\mu}$ are two necessity measures with $\beta_j \leq \hat{\beta}_j$ for all j, then $\mu \geq \hat{\mu}$.

Let μ_1 and μ_2 be two cardinality measures generated by the BUM functions f_1 and f_2, respectively. If $f_1(r) \geq f_2(r)$ for all r, then $\mu_1 \geq \mu_2$. This follows directly from the fact that

$$\mu_1(A) = f_1\left(\frac{\mathrm{Card}(A)}{n}\right) \geq f_2\left(\frac{\mathrm{Card}(A)}{n}\right).$$

If μ_1 and μ_2 are any two probability measures, then **neither** $\mu_1 \geq \mu_2$ nor $\mu_2 \geq \mu_1$ is valid; they are incomparable. We see this as follows: if for some subset A we have $\mu_1(A) > \mu_2(A)$, then for probabilistic measure $\mu_1(\overline{A}) = 1 - \mu_1(A)$ and $\mu_2(\overline{A}) = 1 - \mu_2(A)$, we have $\mu_1(\overline{A}) < \mu_2(\overline{A})$.

We now turn to the concept of an impossible outcome.

Definition: Assume μ is a fuzzy measure, we shall call x_j an **impossible** outcome if for all A

$$\mu(A \cup \{x_j\}) = \mu(A).$$

It is clear that for an impossible outcome $\mu(\{x_j\}) = 0$. This follows since $\{x_j\} = \varnothing \cup \{x_j\}$ then $\mu(\{x_j\}) = \mu(\varnothing) = 0$. It should be emphasized that, in general, the condition $\mu(\{x_j\}) = 0$ **does not** imply that x_j is impossible. However, for some types of measures, the condition $\mu(\{x_j\}) = 0$ does insure impossibility.

For the probability measure, which is simply additive, the condition $\mu(\{x_i\}) = 0$ assures us that $\mu(A \cup \{x_j\}) - \mu(A) = 0$ and hence x_i is impossible. The results are similar for a possibility distribution. In this case, since $\mu(A \cup \{x_i\}) = \mathrm{Max}[\mu(A), \mu(\{x_i\})]$, then if $\mu(\{x_i\}) = 0$ we get $\mu(A \cup \{x_i\}) = \mu(A)$.

In the case of a necessity measure, if $\mu(F_i) = 1$ where $F_i = X - x_i$ then x_i is impossible.

On the other hand, for μ_*, even though $\mu_*(\{x_i\}) = 0$ we have $\mu_*\{F_i \cup \{x_i\}) = \mu(X) = 1$ while $\mu(F_i) = 0$, hence x_i is not impossible. More generally, we note that if μ is a cardinality-based measure, then it has no impossible elements.

We now consider the idea of duality.

Definition: Let μ be a fuzzy measure. We define the dual of μ, $\hat{\mu}$, as the measure

$$\hat{\mu}(A) = 1 - \mu(\hat{A}).$$

It can be clearly seen that if μ is a fuzzy measure, $\hat{\mu}$ is also a fuzzy measure, since

$$\hat{\mu}(\varnothing) = 1 - \mu(\overline{\varnothing}) = 1 - \mu(X) = 0$$
$$\hat{\mu}(X) = 1 - \mu(\overline{X}) = 1 - \mu(\varnothing) = 1.$$

Finally, if $B \subset A$ then $\overline{B} \supset \overline{A}$ and hence $\hat{\mu}(A) = 1 - \mu(\overline{A}) \geq 1 - \mu(\overline{B}) \geq \mu(B)$.

As Dubois and Prade [9] point out, a duality exists between possibility and necessity measures. Assume μ is a possibility measure and let $\hat{\mu}$ be its dual then

$$\begin{aligned}\hat{\mu}(A \cap B) &= 1 - \mu(\overline{A \cap B}) = 1 - \mu(\overline{A} \cup \overline{B})] \\ &= 1 - \mathrm{Max}[\mu(\overline{A}), \mu(\overline{B})] \\ &= \mathrm{Min}[1 - \mu(\overline{A}), 1 - \mu(\overline{B})] = \mathrm{Min}[\hat{\mu}(A), \hat{\mu}(B)].\end{aligned}$$

Thus, we see that the dual of a possibility measure is a necessity and vice versa. More specifically, if μ is a possibility measure with $\mu(\{x_i)\} = \alpha_i$ then its dual is the necessity measure in which $\hat{\mu}(F_i) = 1 - \mu(\{x_i\}) = 1 - \alpha_i$. Thus, we see that possibility measures uniquely define dual pairs. Furthermore, if Π is a possibility measure and if N is its associated necessity measure, its dual, then $\Pi(A) \geq N(A)$ for A, Π is a more confident measure than N.

If μ is a probability measure, then it is a self dual, $\hat{\mu} = \mu$. We see this as follows:

$$\hat{\mu}(A) = 1 - \mu\left(\overline{A}\right) = 1 - \left(\sum_{j \in \overline{A}} p_j\right) = \sum_{j \in A} p_j = \mu(A).$$

Consider now the duality of a cardinality-based measure. Assume μ is a cardinality-based measure, then $\hat{\mu}(E) = 1 - \mu(\overline{E}) = 1 - D_{|\overline{E}|} = 1 - D_{n-|E|}$. Thus, if μ is a cardinality-based measure with weights D_j, then its dual $\hat{\mu}$ is also a cardinality-based measure with weights $\hat{D}_j = 1 - D_{n-j}$. In the case where μ is defined by a BUM function f, we have

$$\hat{\mu}(A) = 1 - f\left(\frac{n - |A|}{n}\right) = 1 - f\left(1 - \frac{|A|}{n}\right).$$

With $\hat{f}(x) = 1 - f(1 - x)$, it is the dual then of f, the $\hat{\mu}$ is defined by $\hat{f}$.

IV. Fuzzy Measures and Dempster–Shafer Belief Structures

A fuzzy measure μ with respect to a variable V provides a description of our knowledge about the variable. When using a fuzzy measure, although there exists some uncertainty with respect to the actual value of the variable, we assume there exists no uncertainty with respect to our knowledge of the description of the uncertainty. An example of this kind of additional uncertainty would be a situation in which we only knew the

$\mu(A) \in [a, b]$. In this situation, we only have partial information with respect to the underlying fuzzy measure.

The D–S belief structure [17], [18] provides a framework for the representation of knowledge about the value of an uncertain variable which can be used when there exists some uncertainty regarding our knowledge of the underlying fuzzy measure [19].

Assume V is a variable taking its value in the space X. A D–S belief structure is defined in terms of a mapping m: $2^X \to [0, 1]$ such that

1) $m(\varnothing) = 0$;
2) $\sum_{B \subseteq X} m(B) = 1$.

Using the terminology of Shafer [17], we call the subsets B_j for which $m(B_j) > 0$ the focal elements, hence condition 2) above becomes $\sum_{j=1}^{q} m(B_j) = 1$, where q is the number of focal elements.

A prototypical model of the D–S belief structure is the random set structure [1]. Here, we perform a random experiment in which our outcome, instead of being an element in X, is a subset of X. In this random experiment, $m(B_j)$ is the probability that the outcome is the set B_j. In this framework, the additional uncertainty stems from the fact that we do not know how the actual element is selected for the set chosen.

The use of a D–S belief structure to describe our knowledge about a variable can be seen as essentially providing *partial information* about the fuzzy measure associated with the variable. Thus, given a D–S belief structure, there are many possible fuzzy measures that can satisfy the constraints it imposes. The two most well-known fuzzy measures associated with a belief structure are the plausibility and belief measures [20]. The plausibility measure is defined as

$$Pl(A) = \sum_{i,\, B_i \cap A \neq \varnothing} m(B_i)$$

and the belief measure is defined as

$$Bel(A) = \sum_{i,\, B_i \subset A} m(B_i).$$

As we shall subsequently see, these two measures provide extreme approaches to selecting a fuzzy measure associated with a D–S structure. In [19], the reader is provided a methodology for generating a collection of fuzzy measures that can be associated with a belief structure. In the following, we describe this methodology.

Let m be a D–S belief structure with focal elements $B_1, B_2, \ldots, B_q$. For each focal element B_j, let W_j be a weighting vector of dimension $|B_j|$ whose components $w_j(i)$ satisfy the conditions $w_j(i) \in [0, 1]$ and $\sum_{i=1}^{|B_j|} w_j(i) = 1$. In this framework, we shall call these weighting vectors the allocation vectors. In [19], the reader is shown that a set measure defined by

$$\mu(E) = \sum_{j=1}^{q} \left(m(B_j) \sum_{i=1}^{|B_j \cap E|} w_j(i) \right)$$

is a fuzzy measure completing the belief structure m. In the preceding, if $|B_i \cap E| = 0$, then we take the sum as 0. Thus, by selecting a collection $W = \langle W_1, W_2, \ldots, W_q \rangle$ of allocation vectors, we define a unique fuzzy measure associated with a belief structure.

A number of special cases are worth pointing out. If all of the W_j are such that $w_j(1) = 1$, the top element in the vector is 1, and the resulting fuzzy measure is the plausibility measure. If all of the W_j are selected such that $w_j(|B_j|) = 1$, the bottom element in the vector is 1; this results in the belief measure. It can be easily shown that these two measures are the extreme; that is, if μ is a fuzzy measure generated from a collection of allocation vectors, $W = \langle W_1, \ldots, W_q \rangle$, then for all $A \subset X$, $Pl(A) \geq \mu(A) \geq Bel(A)$, thus $Pl \geq \mu \geq Bel$.

As we have just indicated by selecting each of the W_j to satisfy the conditions described, we can obtain a fuzzy measure compatible with the given D–S belief structure. While we can individually select each of the associated allocation vectors, it is often more expedient to select all of the vectors in some consistent way, as in the case of plausibility and belief. One way of selecting the allocation vectors in a consistent manner is by using a BUM function f. We recall a BUM function is a mapping f: $[0, 1] \to [0, 1]$ such that $f(0) = 0$, $f(1) = 1$, and $f(x) \geq f(y)$ if $x \geq y$. Using a BUM function, we can globally define all of the OWA allocation vectors W_j as follows. For each B_j we define each W_j such that

$$w_j(i) = f\left(\frac{i}{|B_j|}\right) - f\left(\frac{i-1}{|B_j|}\right).$$

An interesting special case occurs when $f(x) = x$; it is a linear function. Here, we clearly see that the W_j are determined such that $w_j(i) = 1/|B_j|$; the weights in each allocated vector are uniformly distributed. In this case, the resulting fuzzy measure is such that

$$\mu(E) = \sum_{j=1}^{q} m(B_j) \cdot \frac{|B_j \cap E|}{|B_j|}.$$

Here, we see that $\mu(E)$ gets a contribution from B_j proportion to the number of elements in their intersection. We note in this case that $\mu(\{x_k\}) = \sum_{j \text{s.t} \, x_k \in B_j} (m(b_j)/|B_j|)$ and that $\mu(E) = \sum_{k \text{s.t.} \, x_k \in E} \mu(\{x_k\})$, thus the measure generated from this linear f is a probability measure. More specifically, it is the pignistic distribution used by Smets [21].

Two other special cases of using a BUM function are worth noting. If f is such that $f(0) = 0$ and $f(x) = 1$ for $x \neq 1$, then the associated W_j are such that the $w_j(1) = 1$ for all j; this is of course the plausibility measure. On the other hand, if f is such that $f(1) = 1$ and $f(x) = 0$ for all $x \neq 0$, we obtain the belief measure. Another interesting special case is one in which $f(x) = 0$ for $x < 1/2$ and $f(x) = 1$ for $x \geq 1/2$; this can be seen as generating a median-type vector W_j; that is, $w_j(i) = 1$ for the middle element and all of the others are 0. A slight generalization of this is the case with $f(x) = 0$ for $x < \alpha$ and $f(x) \geq 1$ for $x \geq \alpha$. Using this, we see that $\mu(E)$ gets the full contribution of B_j, $m(B_j)$, if at least the α portion of the elements in B_j are also in E.

Let f_1 and f_2 be two BUM functions such that $f_1(x) \geq f_2(x)$ for all x. Let μ_1 and μ_2 be the fuzzy measures obtained by using f_1 and f_2, respectively, to define the allocation vectors; then it can be easily shown that $\mu_1 \geq \mu_2$.

Let us now consider the type of fuzzy measures resulting from some special cases of belief structures. Consider the so-called Bayesian belief structure. Here, each focal element is a singleton $B_j = \{x_j\}$. Let W_j be any collection of allocation vectors and μ be the fuzzy measure generated by these vectors $\mu(E) = \sum_{j=1}^{q}(m(B_j)\sum_{i=1}^{|B_j \cap E|} w_j(i))$. However, since each focal is a singleton, the $|B_j| = 1$, then each W_j consists of only one component with value equals one, $w_j(1) = 1$ for all j. In this case, we see that $\sum_{i=1}^{|B_j \cap E|} w_j(i) = 1$ if $B_j \cap E \neq \emptyset$, if $x_j \in E$ and $\sum_{i=1}^{|B_j \cap E|} w_j(i) = 0$ if $B_j \cap E = \emptyset$, if $x_j \notin E$. Thus, here we always get

$$\mu(E) = \sum_{j \text{s.t.} x_j \in E} m(B_j)$$

which is a probability measure. Thus, this Bayesian D–S structure is only compatible with this probability measure. The obvious reason for this is that the Bayesian D–S belief is a true probability distribution and therefore precisely specifies the associated fuzzy measure.

In order to get some further intuition on the effect of selecting different allocation vectors, we shall consider the case of the D–S belief structure in which $B_1 = X$ and $m(X) = 1$. This essentially corresponds to a case in which we have no information about the value of the variable other than that it lies in X. In this case, we just need to choose one weighting vector W whose cardinality is $|X| = n$. Using this, we get

$$\mu(E) = \sum_{j=1}^{|X \cap E|} w(j) = \sum_{j=1}^{|E|} w(j).$$

Here, $\mu(E)$ is the sum of the first $|E|$ elements in the vector W. If W is determined by a function f, $w_j = f(j/n) - f((j-1)/n)$, then we see that $\mu(E) = f(|E|/n)$.

V. Calculating the Amount of Information in a Fuzzy Measure

Many applications involving probabilistic uncertainty make use of the Shannon entropy. Often these applications involve the use of the Shannon entropy as a criteria function for comparing uncertainty reduction methodologies. In this section, we shall suggest a measure of information associated with a fuzzy measure in the spirit of the Shannon entropy. In order to accomplish this, we shall use a concept introduced into the framework of fuzzy measures by Murofishi [22] and used by Grabisch [23].

Let μ be a fuzzy measure of $X = \{x_1, x_2, \ldots, x_n\}$. For any x_i in X its Shapley index [24] V_i is defined by

$$V_i = \sum_{k=0}^{n-1} \gamma_k \sum_{\substack{K \subset F_i \\ |K|=k}} \mu(K \cup \{x_i\}) - \mu(K)$$

where $|K|$ is the cardinality of the set K, $F_i = X - \{x_i\}$, and $\gamma_k = ((n-k-1)!k!)/n!$. It can be shown that it is always the case that $V_i \in [0, 1]$ and $\sum_{i=1}^{n} V_i = 1$. The Shapley index V_i can be seen as indicating the average increment in confidence obtained by adding x_i to a subset that does not contain it. We use this Shapley index to help indicate the amount of information in a fuzzy measure when it is being used to represent our knowledge about an uncertain variable.

In [25], Yager suggested the use of the Shapley index in the calculation of the amount of information contained in a fuzzy measure associated with a variable. Let μ be a fuzzy measure on the space $X = \{x_1, x_2, \ldots, x_n\}$ representing our knowledge of a variable. Let V_i be the Shapley index of x_i, then define $H(\mu) = -\sum_{i=1}^{n} V_i \ln V_i$; we call this the Shapley entropy associated with μ

It is well established that this formulation of entropy provides a measure of uncertainty [26] associated with an uncertain variable. We recall that in a situation where $V_i \in [0, 1]$ and $\sum_{i=1}^{n} V_i = 1$(as is the case here), $H(\mu)$ attains its maximal value of $\ln(n)$ when $V_i = 1/n$ for all i and its minimal value of 0, when $V_i = 1$ for some i.

One interesting capability provided by the Shapley entropy is the ability to use the principle of maximal entropy. Assume we have a variable about which all our available information allows us to conclude is that one of q fuzzy measures, μ_i for $i = 1$ to q, is the appropriate fuzzy measure. Here, there is some uncertainty as to which is the appropriate fuzzy measure associated with the variable. We are indifferent about these fuzzy measures. The principle of maximal entropy suggests that we select as the measure the one that maximizes the entropy. That is, we select μ_j^* where $H(\mu_j^*) = \text{Max}_j[H(\mu_j)]$. The basic idea here is a conservative one; we choose the fuzzy measure which has the most uncertainty and in doing so we are introducing the least unjustified information.

VI. Some Cases of Entropy

Let us look at this entropy for some particular cases of fuzzy measures to see that it performs appropriately. Consider first the special case of complete certainty. This corresponds to a measure in which $\mu(E) = 1$ if $x_1 \in E$ and $\mu(E) = 0$ if $x_1 \notin E$. In this case, we see that if $x_1 \notin E$, then $\mu(E \cup \{x_1\}) - \mu(E) = 1$. On the other hand, for any $x_j \neq x_1$ and $E\mu(E \cup \{x_j\}) - \mu(E) = 0$. From this we see that for $j \neq 1$, $V_j = 0$ and $v_1 = 1$. Thus, in this case we get $H(\mu) = 0$, the case of minimal uncertainty.

Consider now the case of a probability measure. If p_j denotes the probability of x_j, then $\mu(E) = \sum_{x_j \in E} p_j$. In this case, if $X_i \notin E$ then $\mu(E \cup \{x_i\}) - \mu(E) = p_i$. Based upon this, we get that $V_i = p_i \sum_{k=0}^{n-1} \gamma_k \sum_{\substack{E \subset F_i \\ |E|=k}} 1$. Using some algebraic manipulations, it can be shown that $\sum_{k=0}^{n-1} \gamma_k \sum_{\substack{E \subset F_i \\ |E|=k}} 1 = 1$. Thus, for the probabilistic measure, we get $V_i = p_i$ and hence $H(\mu) \sum_{j=1}^{n} V_j \ln V_j = -\sum_{j=1}^{n} p_j \ln p_j$. Thus, appropriately, our proposed measure reduces to the classical Shannon entropy.

Consider the case of a cardinality based on fuzzy measure $\mu(E) = \sum_{j=1}^{|E|} w_j$. Here, for all E and x_i not in E, $\mu(E \cup \{x_i\}) - \mu(E) = w_{|E|+1}$. From this, we get

$$V_i = \sum_{k=0}^{n-1} \gamma_k \sum_{\substack{K \subset F_i \\ |K|=k}} w_{k+1} = \sum_{k=0}^{n-1} \gamma_k \frac{(n-1)!}{(n-1-k)!k!} w_{k+1}$$
$$= \frac{1}{n} \sum_{k=0}^{n-1} w_{k+1} = \frac{1}{n}.$$

This corresponds to the case of maximal entropy $H(\mu) = \ln(n)$. The implication of this result is very significant for our uncertainty framework. Essentially, what this is saying is that every cardinality-based fuzzy measure is a manifestation of a maximal uncertainty situation. Thus, we see that with this class of cardinality-based fuzzy measures, we have a whole family of potential representations of complete uncertainty. We have already seen some manifestation of this situation. We have already indicated three examples of cardinality measures: μ^*, μ_*, and the case where $D_j = j/n$. In the probabilistic framework, we saw that the case where $p_j = 1/n$ gives us the maximal uncertainty. We note that this is also an example of a cardinality-based measure where $D_j = j/n$

A class of fuzzy measures can be seen to be a collection of fuzzy measures sharing some properties but distinguished by the value of some parameters. A prototypical example of these are the probability measures. Here, while each member of this class obeys the simple additivity property, they are distinguished by the parameters corresponding to the probabilities of the singleton sets. One benefit of using families is that, once having decided upon a family, the process of selecting the actual fuzzy measure becomes simply a process of determining (learning) the values of the associated parameters. We observe that any useful class of fuzzy measures should be able to represent the states of complete uncertainty and complete knowledge about a variable. Thus, it would appear that any useful class should have a cardinality-based measure as one of its members to represent complete uncertainty and μ_x to represent complete certainty.

We now turn to the expression of the Shapley value and the associated entropy for the measures based upon a D–S structure m with focal elements B_j. We recall that the measures of plausibility Pl and belief Bel are two fuzzy measures consistent with this belief structure. In [25], Yager showed that these two fuzzy measures have the same Shapley value.

Theorem: If μ is a plausibility or belief measure obtained from a D–S belief structure with basic assignment function m and focal elements B_j, then the Shapley value for each x_i is

$$V_i = \sum_{j \text{ s.t. } x_i \in B_j} \frac{m(B_j)}{\mathrm{Card}(B_j)}.$$

We can alternatively express this as $V_i = \sum_{j=1}^{q} (m(B_j)/\mathrm{Card}(B_j))\, B_j(x_i)$ where $B_j(x_i)$ is the membership of x_i in B_j. We can see this as the Shapley value for a D–S belief structure.

Let us now look at the entropy associated with the index $H(m) = -\sum_{i=1}^{n} V_i \ln(V_i)$

$$H(m) = -\sum_{j=1}^{q} \frac{m(B_j)}{card(B_j)} \sum_{i=1}^{n} B_j(x_i) \cdot \ln\left(\sum_{K=1}^{q} \frac{m(B_K)}{card(B_K)} \bullet (B_K)(x_i)\right).$$

We shall now obtain the Shapley index for the case of possibilistic uncertainty. We recall that for possibilistic we have a fuzzy measure μ such that $\mu(E) = \mathrm{Max}_{x_j \in E}[\mu(\{x_j\})$. Let us denote $\alpha_j = \mu(\{x_j\})$. We shall use a result from [27] which shows any possibility measure can be represented as the plausibility measure of some D–S structure which has consonant focal elements. A D–S belief structure is called a consonant belief structure if the focal elements are nested $B_1 \supset B_2 \supset B_3 \supset B_4 \supset \cdots \supset B_q$. Let us examine the relationship between a given possibility measure and the associated consonant belief structure. Assume μ is a possibility measure, where $\mu(\{x_i\}) = Poss[\{x_i\}] = \alpha_i$. Without loss of generality, we shall assume the x_i have been indexed such that $\alpha_i \geq \alpha_j$ if $i > j$. In this case, $\alpha_n = 1$. Consider now a D–S belief structure where $B_j = \{x_i | i = j \text{ to } n\}$. We see that $B_j \subset B_k$ if $k < j$ and therefore this is a consonant belief structure. In this belief structure, let the basic assignment function m be such that $m(B_j) = \alpha_j - \alpha_{j-1}$ for $j = 1$ to n where by definition we let $\alpha_0 = 0$. For this D–S structure, we see that

$$Pl(\{x_j\}) = \sum_{i=1}^{n} m(B_i)B_i(x_j) = \sum_{i=1}^{j} m(B_i) = \sum_{i=1}^{j} \alpha_i - \alpha_{i-1} = \alpha_j.$$

Thus, we see that this plausibility measure is the same as the original possibility measure. From this representation of the possibility measure in terms of a plausibility measure derived D–S belief structure we can, using the results of the preceding section, obtain the Shapley index of the possibility measure $V_i = \sum_{\substack{j \text{ s.t} \\ x_i \in B_j}} (m(B_j)/\mathrm{Card}(B_j))$. Using the particular values of m and B_j, we get

$$V_i = \sum_{j=1}^{i} \frac{m(B_j)}{n+1-j} = \sum_{j=1}^{i} \frac{\alpha_j - \alpha_{j-1}}{n+1-j}.$$

It is interesting to note that the value V_i obtained here corresponds to the probability obtained for x_i by one of the methods used to convert a possibility distribution to a probability distribution.

VII. Decision Making With Fuzzy Measures

One important situation in which we must be able to use an uncertainty representation is the task of decision making under uncertainty. In decision making under uncertainty, the choice of an action rather resulting in a unique outcome has an associated collection of possible outcomes. The basic issues involved in decision making under uncertainty can best be understood using the decision matrix shown in Fig. 1.

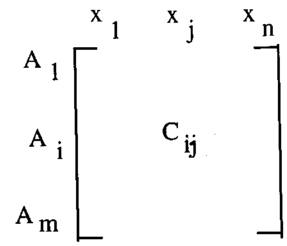

Fig. 1. Decision matrix.

In this matrix, the A_i correspond to a collection of alternatives available to the decision maker. The elements x_j are possible values associated with some relevant variable V. Finally, C_{ij} is the payoff to the decision maker if he selects alternative A_i and V assumes the value x_j. The decision maker must select one alternative with the objective of receiving the best payoff. In decision making under uncertainty, the value of V is unknown before the decision maker must select an alternative action. The decision maker's knowledge about the value of V is expressed by some uncertainty representation.

In uncertain decision making, each course of action is a multi-dimensional object consisting of the set of possible payoffs that can result from the selection of this action. The difficulty in making decisions in the face of uncertainty is rooted in the difficulty humans have in comparing multi-dimensional objects. One approach to selecting between alternatives having uncertain payoffs is to associate with each alternative a single value, called a representative value, and then order the alternatives with respect to these scalar values. To do this, we must have a function to map each alternative into its single representative value. We call this a valuation function and denote it as Val. One can associate with such a representative value generating function some basic properties. The first property is idempotency; if all the payoffs for a give alternative are the same, this should be its representative value. A second, closely related property is boundedness; the representative value should lie between the largest and smallest outcomes for the alternative. Monotonicity should be satisfied; if one of the payoffs for an alternative increases, the representative value for that alternative should not decrease. Since these representative values are to be used to choose the best alternative, it is desirable that the calculation of the representative value be made in such a way as to satisfy Arrow's condition of indifference to irrelevant alternatives [29]; the removal of an alternative that is not the best should not change the optimal choice. This effectively requires that the calculation of the representative value of an alternative not use any payoff information about other alternatives. The means that the calculation of the representative value of an alternative just be a function of payoffs for that alternative and the underlying uncertainty representation; that implies that the valuation function be a pointwise operator.

The expected value provides such a valuation function in the case in which the uncertainty associated with V is represented by a probability distribution. In the more general case in which the knowledge about V is carried by a fuzzy measure μ, we need an analogous tool to obtain a representative value for the individual alternatives.

A very promising approach to providing a generalization of the expected value to uncertainties represented by a fuzzy measure is to use the Choquet integral [30]. Let us consider an alternative A_i and we shall describe the calculation of $Val(A_i)$ using the Choquet integral. Let V be a variable which can assume one of the values $x_j \in X$. Here, our knowledge about the value of this variable is represented by a fuzzy measure μ. Additionally, let C_{ij} be the payoff that occurs if we select outcome A_i and V assumes the value x_j. With loss of generality, we shall assume the indexing of the C_{ij} has been done so that $C_{i1} \geq C_{i2} \geq \cdots \geq C_{in}$. Using the Choquet integral [31] to calculate the valuation of A_i we obtain $Val(A_i) = \sum_{j=1}^{n} (\mu(H_j) - \mu(H_{j-1}))C_{ij}$. In the preceding $H_j = \{x_1, x_2, \ldots, x_j\}$, the set of outcomes with the j largest payoffs. The intuition behind using this formulation as a generalization of the expected value becomes clear if we consider the probabilistic situation and begin with the expected value $E(A_i) = \sum_{j=1}^{n} p_j C_{ij}$. Letting $\mathrm{Sum}(j) = \sum_{i=1}^{j} p_i$, the cumulative distribution function, and seeing that $p_j = (\mathrm{Sum}(j) - \mathrm{Sum}(j-1))$ we can rewrite $E(A_i) = \sum_{j=1}^{n} (\mathrm{Sum}(j) - \mathrm{Sum}(j-1))C_{ij}$. Noting under our indexing, $C_{ik} \geq C_{il}$ if $k < l$, that $\mathrm{Sum}(j)$ is the probability that we receive at least C_{ij} and realizing that in the case of when our uncertainty is represented by the measure μ, the value $\mu(H_j)$ indicates our belief that we receive at least C_{ij}, then by replacing $\mathrm{Sum}(j)$ with $\mu(H_j)$ in the expected value formulation, we obtain the Choquet formula.

A more general expression of the Choquet integral which does not rely on the special indexing, $C_{ik} \geq C_{il}$ if $k < l$ is as follows. We let $c\text{-index}(k)$ indicate the index of the kth largest C_{ij}. We let $H_k = \{x_{c\text{-index}(1)}, x_{c\text{-index}(2)}, \ldots, x_{c\text{-index}(k)}\}$, the states of nature with the k highest payoffs. Using this, we have $Val(A_i) = \sum_{k=1}^{n} (\mu(H_k) - \mu(H_{k-1})C_{ic\text{-index}(k)}$.

VIII. Conclusion

We introduced the fuzzy measure as a unifying structure for modeling knowledge about an uncertain variable. We indicated that many well-established types of uncertainty representation can be modeled in this framework. A measure of entropy associated with a fuzzy measure was introduced and its manifestation for different fuzzy measures was described. The problem of uncertain decision making for the case in which the uncertainty was represented by a fuzzy measure was considered. The Choquet integral was introduced as providing a generalization of the expected value to this environment.

References

[1] I. R. Goodman and H. T. Nguyen, *Uncertainty Models for Knowledge-Based Systems*. Amsterdam, The Netherlands: North-Holland, 1985.

[2] M. Sugeno, "Fuzzy measures and fuzzy integrals: A survey," in *Fuzzy Automata and Decision Process*, M. M. Gupta, G. N. Saridis, and B. R. Gaines, Eds. Asterdam, The Netherlands: North-Holland, 1977, pp. 89–102.

[3] M. Grabisch, T. Murofushi, and M. Sugeno, *Fuzzy Measures and Integrals*. Heidelberg, Germany: Physica-Verlag, 1999.

[4] T. Murofushi and M. Sugeno, "Fuzzy measures and fuzzy integrals," in *Fuzzy Measures and Integrals*, M. Grabisch, T. Murofushi, and M. Sugeno, Eds. Heidelberg, Germany: Physica-Verlag, 1999, pp. 3–41.

[5] G. Choquet, "Theory of capacities," *Ann. de l'Institut Fourier*, vol. 5, 1953.

[6] D. Denneberg, *Non-Additive Measure and Integral*. Norwell, MA: Kluwer, 1994.

[7] M. Sugeno, "Theory of fuzzy integrals and its application," Doctoral dissertation, Tokyo Institute of Technology, Tokyo, Japan, 1974.

[8] L. A. Zadeh, "Fuzzy sets as a basis for a theory of possibility," *Fuzzy Sets Syst.*, vol. 1, pp. 3–28, 1978.

[9] D. Dubois and H. Prade, *Possibility Theory: An Approach to Computerized Processing of Uncertainty*. New York: Plenum, 1988.

[10] G. De Cooman, D. Ruan, and E. E. Kerre, *Foundations and Applications of Possibility Theory*. Singapore: World Scientific, 1995.

[11] L. A. Zadeh, "Test-score semantics as a basis for a computational approach to the representation of meaning," *Literary Linguist. Comput.*, vol. 1, pp. 24–35, 1986.

[12] ——, "Fuzzy logic = computing with words," *IEEE Trans. Fuzzy Syst.*, vol. 4, pp. 103–111, 1996.

[13] R. R. Yager, "Applications and extensions of OWA aggregations," *Int. J. Man–Mach. Stud.*, vol. 37, pp. 103–132, 1992.

[14] D. Denneberg, "Nonadditive measure and integral, basic concepts and their role for applications," in *Fuzzy Measures and Integrals*, M. Grabisch, T. Murofushi, and M. Sugeno, Eds. Heidelberg, Germany: Physica-Verlag, 1999, pp. 42–69.
[15] A. Tversky and D. Kahneman, "Advances in prospect theory: Cumulative representation of uncertainty," *J. Risk and Uncertainty*, vol. 5, pp. 297–323, 1992.
[16] Y. Tsukamoto, "A fuzzy measure model of uncertainty averter," in *Proc. 1st Asian Conf. Fuzzy Sets and Systems*, Singapore, 1993.
[17] G. Shafer, *A Mathematical Theory of Evidence*. Princeton, NJ: Princeton Univ. Press, 1976.
[18] R. R. Yager, J. Kacprzyk, and M. Fedrizzi, *Advances in the Dempster–Shafer Theory of Evidence*. New York: Wiley, 1994.
[19] R. R. Yager, "A class of fuzzy measures generated from a Dempster–Shafer belief structure," *Int. J. Intell. Syst.*, vol. 14, pp. 1239–1247, 1999.
[20] P. Smets, "Belief functions," in *Non-Standard Logics for Automated Reasoning*, P. Smets, E. H. Mamdani, D. Dubois, and H. Prade, Eds. New York: Academic, 1988, pp. 253–277.
[21] P. Smets and R. Kennes, "The transferable belief model," *Artif. Intell.*, vol. 66, pp. 191–234, 1994.
[22] T. Murofushi, "A technique for reading fuzzy measures (i): The Shapley value with respect to a fuzzy measure" (in Japanese), in *Proc. 2nd Fuzzy Workshop*, Nagaoka, Japan, 1992, pp. 39–48.
[23] M. Grabisch, "Alternative representations of discrete fuzzy measures for decision making," *Int. J. Uncertainty, Fuzziness, and Knowledge-Based Systems*, vol. 5, pp. 587–607, 1997.
[24] L. S. Shapley, "A value for n-person games," in *Contributions to Game Theory*, H. W. Kuhn and A. W. Tucker, Eds. Princeton, NJ: Princeton Univ. Press, 1953, pp. 307–317.
[25] R. R. Yager, "On the entropy of fuzzy measures," *IEEE Trans. Fuzzy Syst.*, vol. 8, pp. 453–461, 2000.
[26] C. L. Shannon, "The mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 623–656, 1948.
[27] D. Dubois and H. Prade, "On several representations of an uncertain body of evidence," in *Fuzzy Information and Decision Processes*, M. M. Gupta and E. Sanchez, Eds. Amsterdam, The Netherlands: North-Holland, 1982, pp. 309–322.
[28] G. J. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic: Theory and Applications*. Englewood Cliffs, NJ: Prentice-Hall, 1995.
[29] K. J. Arrow, *Social Choice and Individual Values*. New York: Wiley, 1951.
[30] T. Murofushi and M. Sugeno, "An interpretation of fuzzy measures and the Choquet integral as an integral; with respect to a fuzzy measure," *Fuzzy Sets Syst.*, vol. 29, pp. 201–227, 1989.
[31] M. Grabisch, "Fuzzy measures and integrals: A survey of applications and recent issues," in *Fuzzy Information Engineering: A Guided Tour of Applications*, D. Dubois, H. Prade, and R. R. Yager, Eds. New York: Wiley, 1997, pp. 507–529.

Ronald R. Yager (S'66–M'83–SM'93–F'97) received the undergraduate degree from the City College of New York, and the Ph.D. degree from the Polytechnic Institute of New York, Brooklyn.

He has served at the National Science Foundation as Program Director in the Information Sciences program. He was a NASA/Stanford Visiting Fellow as well as a Research Associate at the University of California, Berkeley. He has served as a Lecturer at NATO Advanced Study Institutes. Currently, he is Director of the Machine Intelligence Institute and Professor of Information and Decision Technologies at Iona College, New Rochelle, NY. He is Editor-in-Chief of the *International Journal of Intelligent Systems*. He serves on the editorial board of a number of journals including *Neural Networks, Data Mining and Knowledge Discovery, Fuzzy Sets and Systems, Journal of Approximate Reasoning,* and *International Journal of General Systems*. He is one of the co-founders of the Conference on Information Processing and the Management of Uncertainty. He has published over 500 articles and 15 books. In addition to his pioneering work in the area of fuzzy logic, he has made fundamental contributions in decision making under uncertainty and the fusion of information. His current research interests include e-commerce, data mining, information retrieval, aggregation theory, and the development of technologies for a more intelligent Internet.

Dr. Yager is a Fellow of the IEEE, the Fuzzy Systems Association, and the New York Academy of Sciences. He serves on the editorial board of IEEE TRANSACTIONS ON FUZZY SYSTEMS and IEEE INTELLIGENT SYSTEMS.