

Generalizing the Dempster–Shafer Theory to Fuzzy Sets

JOHN YEN, MEMBER, IEEE

Abstract—With the desire to manage imprecise and vague information in evidential reasoning, several attempts have been made to generalize the Dempster–Shafer (D–S) theory to deal with fuzzy sets. However, the important principle of the D–S theory, that the belief and plausibility functions are treated as lower and upper probabilities, is no longer preserved in these generalizations. A generalization of the D–S theory in which this principle is maintained is described. It is shown that computing the degree of belief in a hypothesis in the D–S theory can be formulated as an optimization problem. The extended belief function is thus obtained by generalizing the objective function and the constraints of the optimization problem. To combine bodies of evidence that may contain vague information, Dempster’s rule is extended by 1) combining generalized compatibility relations based on the possibility theory, and 2) normalizing combination results to account for partially conflicting evidence. Our generalization not only extends the application of the D–S theory but also illustrates a way that probability theory and fuzzy set theory can be integrated in a sound manner in order to deal with different kinds of uncertain information in intelligent systems.

I. INTRODUCTION

EVIDENTIAL REASONING, which is the task of inferring the likelihood of some hypotheses by collecting and combining relevant evidence for or against these hypotheses, is central to many computer systems that help users in decisionmaking, diagnosis, pattern recognition, and speech understanding. The problem of evidential reasoning is complicated by information being conveyed by a piece of evidence is often not only uncertain, but also imprecise, incomplete, and vague. For example, a sensor’s output may indicate that a flying object is about 50 miles from Los Angeles and that it belongs to a general class of missiles. But the sensor gives no further information about the specific type of the missile. Therefore, an evidential reasoning mechanism that can cope with all these different kinds of uncertainties in a sound manner is highly desirable.

Previous work on evidential reasoning has been largely based on three theoretical frameworks: the Bayesian

probability theory, the Dempster–Shafer (D–S) theory of evidence, and the fuzzy set theory. These frameworks differ in their strengths and weaknesses. The Bayesian probability theory has a well-developed decision-making theory, but it requires precise probability judgments. Hence, it is weak in representing and managing imprecise information. To cope with this weakness, a Bayesian approach often needs to transform a piece of imprecise evidence into a precise one by using additional assumptions [1]. The D–S theory is based on probability theory, yet it allows probability judgments to capture the imprecise nature of the evidence. As a result, degrees of likelihood are measured by probability intervals, as opposed to point probabilities in the Bayesian approaches. One of the weaknesses of the D–S theory is that its decision theory is still a research topic [2]. The fuzzy set theory focuses on the issue of representing and managing vague information such as “the temperature is *high*” or “the missile is *about* 50 miles from Los Angeles.” One of its strengths is its *possibility theory* as a foundation for dealing with imprecise data. Although the fuzzy set theory is still somewhat controversial at this point, it has been used successfully to solve many complex real-world problems. For example, Hitachi has used fuzzy control to develop an automatic train operation system for Sendai’s municipal subway [3].

In this paper, we describe an approach that addresses the issue of managing imprecise and vague information in evidential reasoning by combining the D–S theory with the fuzzy set theory. Although several researchers have extended the D–S theory to deal with vague information [4]–[7], their extensions have not been able to preserve an important principle in the D–S theory: that the belief and the plausibility measures are lower and upper probabilities. Viewing this, we generalize the D–S theory in a way that preserves this principle. We achieve this by first generalizing the fundamental constructs of the theory and then deriving other extensions to the theory from these generalizations. The primitive constructs that have been generalized are 1) the compatibility relation, which relates the evidence to the hypotheses, and 2) the objective function and the constraints of the optimization problem, which compute the belief and the plausibility functions. From these generalized basic components, we derive the belief function, the plausibility function, and the rule of

Manuscript received November 11, 1988; revised October 17, 1989.

This article is partially based on the author’s Ph.D. thesis at the University of California, Berkeley, which was supported by National Science Foundation Grant DCR-8513139. Part of the research described in the paper was done at USC/Information Sciences Institute and was supported by the Defense Advanced Research Projects Agency under Contract No. MDA903-86-C-0178.

The author is with the Department of Computer Science, Texas A&M University, College Station, TX 77843.

IEEE Log Number 8934160.

combination for the generalized theory of evidence. Finally, we discuss the relationship between Shafer's consonant support functions and the possibility distributions based on our generalized framework.

II. THE PROBLEM

The problem we want to solve in this paper can be described as follows. Suppose X and Y are two variables that take their possible values from two spaces, S and T , respectively. The space S is an evidence space that consists of a set of mutually exclusive and exhaustive evidential elements. The space T is a hypothesis space that is formed by a set of mutually exclusive and exhaustive hypotheses. A body of evidence for the hypothesis space T is constituted by (1) a set of rules that associate evidential elements to hypotheses in the form of

$$\text{IF } X = s_i \text{ THEN } Y \text{ is } A_i$$

where s_i is an evidential element and A_i is a fuzzy subset of T , and (2) a probability distribution of the evidence space S . Our objective is to answer questions like "What is the likelihood that Y is B given a collection of bodies of evidence?" where B is a fuzzy subset of T .

To illustrate this, let us consider a computer system that infers the age of a person based on various information about the person. Such a system may contain two bodies of evidence, one regarding the boldness of the person, the other about whether he/she likes punk rock. The rules for these two bodies of evidence are listed below.

IF the person is bold, THEN his age is *NOT YOUNG*.
 IF the person is not bold, THEN his age is *UNKNOWN*.
 IF the person likes punk rock, THEN his age is *YOUNG*.
 IF the person does not like punk rock, THEN his age is *UNKNOWN*.

where *not young* and *young* are fuzzy subsets of the interval $[0, 100]$. Suppose the system is given the following probability judgments about a person named John:

$$\begin{aligned} P(\text{bold}) &= 0.8, & P(\text{not bold}) &= 0.2, \\ P(\text{likes punk}) &= 0.4, & P(\text{does not like punk}) &= 0.6. \end{aligned}$$

The system is asked to determine how likely it is that John is a *middle-aged* person.

The important characteristic about the problem being considered here is that it contains both probabilistic information and vague information (e.g., *young*, *middle-aged*). The Dempster-Shafer theory has been shown to solve a special case of this problem where A_i and B are crisp sets [4]. Hence, we will briefly describe the basics of the D-S theory before we discuss previous work and our approach in generalizing the theory.

III. BASICS OF THE DEMPSTER-SHAFFER THEORY

The Dempster-Shafer theory originated from the concept of lower and upper probability induced by a multivalued mapping [8]. Glenn Shafer further extended the theory in his book [9].

A multivalued mapping from space S to space T associates each element in S with a set of elements in T , i.e., $\Gamma: S \rightarrow 2^T$. The image of an element s in S under the mapping is called the *granule* of s , denoted as $G(s)$. The multivalued mapping can also be viewed as a *compatibility relation* between the spaces S and T . A compatibility relation C between S and T characterizes the possibilistic relationship between their elements. An element s of S is compatible with an element t of T if it is possible that s is an answer to S and t is an answer to T at the same time [10] and the *granule* of s is the set of all elements in T that are compatible with s .

$$G(s) = \{t | t \in T, sCt\}.$$

Given a probability distribution of space S and a compatibility relation between S and T , a basic probability assignment (BPA) of space T , denoted by $m: 2^T \rightarrow [0, 1]$, is induced:¹

$$m(A) = \frac{\sum_{G(s_i)=A} p(s_i)}{1 - \sum_{G(s_i)=\phi} p(s_i)} \quad (1)$$

where the subset A is also called a *focal element*.

The probability distribution of space T , which is referred to as *the frame of discernment*, is constrained by the basic probability assignment, but in general, it is not uniquely determined by the BPA. The belief measure and the plausibility measure of a set B are, respectively, the lower probability and the upper probability of the set subject to those constraints. These two quantities are obtained from the BPA as follows:

$$\text{bel}(B) = \sum_{A \subset B} m(A) \quad (2)$$

$$\text{Pls}(B) = \sum_{A \cap B \neq \phi} m(A). \quad (3)$$

Hence, the belief interval $[\text{bel}(B), \text{Pls}(B)]$ is the range of B 's probability.

An important advantage of the D-S theory is its ability to express degrees of ignorance. In the theory, the commitment of belief to a subset does not force the remaining belief to be committed to its complement, i.e., $\text{bel}(B) + \text{bel}(B^c) \leq 1$. The amount of belief committed to neither B nor B 's complement is the degree of ignorance. Consequently, the theory provides a framework within which disbelief can be distinguished from a lack of evidence for belief.

If m_1 and m_2 are two BPA's induced by two independent evidential sources, the combined BPA is calculated

¹If we assume that Γ does not map any element of the space E to the empty set, the denominator (i.e., the normalization factor) in 1) becomes one.

according to Dempster's rule of combination:

$$m_1 \oplus m_2(C) = \frac{\sum_{A_i \cap B_j = C} m_1(A_i) m_2(B_j)}{1 - \sum_{A_i \cap B_j = \phi} m_1(A_i) m_2(B_j)}. \quad (4)$$

The basic combining steps that result in Dempster's rule are discussed in Section V-F.

IV. PREVIOUS WORK

Zadeh was the first to generalize the Dempster-Shafer theory to fuzzy sets, based on his work on the concept of information granularity and the theory of possibility [4], [11]. A possibility distribution, denoted by Π , is a fuzzy restriction that acts as an elastic constraint on the values of a variable [12], [13]. Zadeh first generalized the granule of a D-S compatibility relation to a conditional possibility distribution. Then he defined the *expected certainty*, denoted by $EC(B)$, and the *expected possibility*, denoted by $E\Pi(B)$, as a generalization of D-S belief and plausibility functions:

$$E\Pi(B) = \sum_i m(A_i) \sup(B \cap A_i)$$

$$EC(B) = \sum_i m(A_i) \inf(A_i \Rightarrow B) = 1 - E\Pi(B^c)$$

where A_i denotes fuzzy focal elements induced from conditional possibility distributions, $\sup(B \cap A_i)$ measures the degree that B intersects with A_i , and $\inf(A_i \Rightarrow B)$ measures the degree to which A_i is included in B . It is easy to verify that the expected possibility and the expected certainty reduce to the D-S belief and plausibility measures when all A_i and B are crisp sets.

Following Zadeh's work, Ishizuka, Yager, and Ogawa have extended the D-S theory to fuzzy sets in slightly different ways [5]–[7]. They all extend D-S's belief function by defining a measure of inclusion $I(A \subset B)$, the degree to which set A is included in set B , and by using the following formula, similar to Zadeh's expected certainty $EC(B)$.

$$\text{bel}(B) = \sum_{A_i} I(A \subset B) m(A_i)$$

Their definitions of the measures of inclusion are listed as follows.

Ishizuka:

$$I_I(A \subset B) = \frac{\min_x [1, 1 + (\mu_B(x) - \mu_A(x))]}{\max_x \mu_A(x)}. \quad (5)$$

Yager:

$$I_Y(A \subset B) = \min_x [\mu_{\bar{A}}(x) \vee \mu_B(x)]. \quad (6)$$

Ogawa:

$$I_O(A \subset B) = \frac{\sum_i \min[\mu_A(x_i), \mu_B(x_i)]}{\sum_i \mu_B(x_i)}. \quad (7)$$

Based on Zadeh's expected certainty, Ishizuka and Yager arrive at different inclusion measures by using different implication operators in fuzzy set theory. Ogawa uses relative sigma count, which is analogous to conditional probability in spirit, to compute the degree of inclusion.

In order to combine two mass distributions with fuzzy focal elements, Ishizuka extended Dempster's rule by taking into account the degree of intersection of two sets, $J(A, B)$.

$$m_1 \oplus m_2(C) = \frac{\sum_{A_i \cap B_j = C} J(A_i, B_j) m_1(A_i) m_2(B_j)}{1 - \sum_{i,j} (1 - J(A_i, B_j)) m_1(A_i) m_2(B_j)} \quad (8)$$

where

$$J(A, B) = \frac{\max_x [\mu_{A \cap B}(x)]}{\min [\max_x \mu_A(x), \max_x \mu_B(x)]}.$$

There are four problems with these extensions. First, the belief functions sometimes are not sensitive to significant changes in focal elements because degrees of inclusion are determined by certain "critical" points due to the use of "min" and "max" operators. Second, the definitions of "fuzzy intersection operator" and "fuzzy inclusion operator" are not unique. Consequently, it is difficult to choose the most appropriate definition for a given application. Third, although expected possibility and expected certainty (or, equivalently, expected necessity) degenerate to Dempster's lower and upper probabilities in the case of crisp sets, it is not clear that this is a "necessary" extension. Fourth, the generalized formula for combining evidence is not well justified.

V. OUR APPROACH

Instead of directly modifying the formulas in the D-S theory, we generalize the primitive constructs of the theory and derive other extensions to the theory from these generalizations. We first generalize the compatibility relation in the D-S theory to a joint possibility distribution. Then, we formulate the linear programming problems that compute the belief measures and the plausibility measures. By extending the objective function and the constraints of the optimization problem, we obtain the formula for computing belief function in the generalized framework. We also extend Dempster's rule of combination by generalizing its steps in 1) combining the compatibility relations and 2) normalizing the combination result to account for the partial conflict between pieces of evidence. Finally, we achieve the commutativity of the extended Dempster rule by postponing its normalization step.

A. Generalizing the Compatibility Relation to a Possibility Distribution

In the Dempster–Shafer theory, the compatibility relation is limited to black-and-white answers. For example, given the question of whether s and t could be answers to S and T respectively, the compatibility relation may record only that the given situation is completely possible (i.e., (s, t) is in the relation C) or completely impossible (i.e., (s, t) is not in C). In general, however, the possibility that both s and t are answers to S and T is a matter of degree. To cope with this, we generalize Shafer's compatibility relation to a fuzzy relation that records joint possibility distribution of the spaces S and T .

Definition 1: A generalized compatibility relation between the spaces S and T is a fuzzy relation $C: 2^{S \times T} \rightarrow [0, 1]$ that represents the joint possibility distribution of the two spaces, i.e.,

$$C(s, t) = \Pi_{X, Y}(s, t)$$

where X and Y are variables that take values from the space S and the space T , respectively.

Shafer's compatibility relation is a special case of our fuzzy relation in which possibility measures are indicated by either zeros or ones.

In fuzzy set theory, if the relationship of two variables X and Y is characterized by a fuzzy relation R and the value of variable X is A , the value of variable Y can be induced using the *composition operation*, which is defined as:

$$\mu_{A \circ R}(y) = \max_x \{ \min [\mu_A(x), \mu_R(x, y)] \}$$

So, we use the composition rule to generalize the definition of *granule*.

Definition 2: Given a generalized compatibility relation $C: 2^{S \times T} \rightarrow [0, 1]$, the granule of an element s of S , denoted as $G(s)$, is defined to be the composition of the singleton $\{s\}$ and C , which turns out to be the possibility distribution conditioned on s , i.e.,

$$G(s) = \{s\} \circ C = \Pi_{(Y|X=s)}$$

Hence, we generalize granules to conditional possibility distributions just as Zadeh did; however, our approach is more general than Zadeh's approach because we go one step further to generalize the compatibility relation to a joint possibility distribution. As we will see in Section V-F, the generalized compatibility relation is important for justifying our generalization of Dempster's rule.

Given a probability distribution of the space S and a joint possibility distribution between space S and space T such that the granules of S 's elements are normal fuzzy subsets,² a basic probability assignment (BPA) m to T is induced using equation 1. Adopting the terminology of the D–S theory, we call a fuzzy subset of T with nonzero

basic probability a *fuzzy focal element*. A fuzzy basic probability assignment (BPA) is a BPA that has at least one fuzzy focal element.

B. The Optimization Problem for Computing the Belief Function

As a basis for the following discussions, this section formulates the linear programming problems implicitly solved by the belief function. This serves as a foundation upon which we can generalize various basic components of the optimization problem (e.g., the objective function, the constraints) that correspond to basic concepts underlying the belief function.

$\text{Pls}(B)$ and $\text{bel}(B)$ are the upper and lower probabilities of a set B under the constraints imposed by a basic probability assignment. Therefore, the belief function can be obtained by solving the following optimization problem:

$$\text{LP1}—\min \sum_{x_i \in B} \sum_j m(x_i: A_j)$$

subject to the following constraints:

$$m(x_i: A_j) \geq 0, \quad i = 1, \dots, n; j = 1, \dots, 1 \quad (9)$$

$$m(x_i: A_j) = 0, \quad \forall x_i \notin A_j \quad (10)$$

$$\sum_i m(x_i: A_j) = m(A_j) \quad j = 1, \dots, 1. \quad (11)$$

The variable $m(x_i: A_j)$ denotes the probability mass allocated to x_i from the basic probability of a focal element A_j . The objective function simply computes the total probability of the set B where the inner summation gives the probability of an element x_i . The inequality constraint, specified by (9), states the nonnegativity of probability masses. Equation (10) prohibits the basic probability of a focal from being assigned to any elements outside the focal. Equation (11) expresses that all the probability mass assigned by a focal should add up to its basic probability. It follows, from (9) and (11), that the upper bound on $m(x_i: A_j)$ is $m(A_j)$.

Since the distributions of focals' masses do not interact with one another, they can be optimized individually to reach a global optimal solution. Hence, we partition the linear programming problem LP1 into subproblems, each one of which concerns the allocation of the mass of a focal element. The optimal value of the original problem LP1 is the sum of the subproblems' solutions. A subproblem for LP1 is formulated as follows:

$$\text{LP1}_j—\min \sum_{x_i \in B} m(x_i: A_j)$$

subject to the following constraints:

$$m(x_i: A_j) \geq 0$$

$$m(x_i: A_j) = 0 \quad x_i \notin A_j$$

$$\sum_i m(x_i: A_j) = m(A_j).$$

²A fuzzy subset A is normal if $\sup_x \mu_A(x) = 1$. The assumption that all focal elements are normal is further discussed in Section V-F-2).

The linear programming problem for computing the plausibility of set B differs only in the direction of optimization. It is formulated as LP2 as follows:

$$\text{LP2—max } \sum_{x_i \in B} \sum_j m(x_i: A_j) \text{ subject to (9)–(11).}$$

Like LP1, the linear programming problem LP2 can be partitioned into l subproblems, each of which finds an optimal distribution of a focal's mass to make a maximum contribution to the belief in B .

The optimal solutions of the minimization subproblem LP1 $_j$ and the maximization subproblem LP2 $_j$ are denoted as $m_*(B: A_j)$ and $m^*(B: A_j)$ respectively. Adding the optimal solutions of subproblems, we get B 's belief measure and plausibility measure as shown below.

$$\text{bel}(B) = \sum_{A_j \subseteq T} m_*(B: A_j). \quad (12)$$

$$\text{Pls}(B) = \sum_{A_j \subseteq T} m^*(B: A_j). \quad (13)$$

It is easy to show that the optimal solutions of the subproblems are the following:

$$m_*(B: A_j) = \begin{cases} m(A_j) & \text{if } A_j \subset B \\ 0 & \text{otherwise} \end{cases} \quad (14)$$

$$m^*(B: A_j) = \begin{cases} m(A_j) & \text{if } A_j \cap B \neq \phi \\ 0 & \text{otherwise} \end{cases} \quad (15)$$

Equations (2) and (3), the formulas for calculating D–S belief and plausibility, thus follow directly from (12)–(15).

C. Generalizing Objective Functions

Philippe Smets has shown that the belief measure of a fuzzy set B , given a nonfuzzy basic probability assignment, can be obtained by computing the lower bound on the expected value of B 's membership function [14]. Here we show that the same result can be obtained by modifying the objective functions of the optimization problems, discussed in Section V-B, to account for the membership degree of the fuzzy set B .

The objective function of LP1 and LP2 computes the probability of a crisp set B . If B is a fuzzy subset of the frame of discernment, its probability is defined as

$$P(B) = \sum_{x_i} P(x_i) \times \mu_B(x_i)$$

in fuzzy set theory. We can thus generalize the objective function to

$$\sum_{x_i} \sum_j m(x_i: A_j) \times \mu_B(x_i).$$

Based on this generalization of the objective functions and the following theorem, we get the belief function of fuzzy sets for a nonfuzzy basic probability assignment.

Theorem 3: Suppose A is a nonfuzzy focal element. The maximum and minimum probability masses that can be allocated to a fuzzy set B from A are

$$m_*(B: A) = m(A) \times \inf_{x \in A} \mu_B(x) \quad (16)$$

$$m^*(B: A) = m(A) \times \sup_{x \in A} \mu_B(x). \quad (17)$$

Proof: $m^*(B: A)$ is the optimal solution to the following linear programming problem:

$$\min \sum_x m(x: A) \times \mu_B(x)$$

subject to the following constraints:

$$m(x: A) \geq 0$$

$$m(x: A) = 0 \quad \forall x \notin A$$

$$\sum_x m(x: A) = m(A).$$

An optimal solution of this simple linear programming problem can be obtained by assigning all the mass of A to an element of A that has the lowest membership degree in B . Thus, we have $m_*(B: A) = m(A) \times \inf_{x \in A} \mu_B(x)$. Equation (17) can be proved in a similar way.

From (12), (13), (16), and (17), we obtain the following formula for computing the belief and plausibility of fuzzy sets from a crisp basic probability assignment:

$$\text{bel}(B) = \sum_{A_j \subseteq T} m(A_j) \times \inf_{x \in A_j} \mu_B(x)$$

$$\text{Pls}(B) = \sum_{A_j \subseteq T} m(A_j) \times \sup_{x \in A_j} \mu_B(x).$$

Thus, we have shown that Smets' generalization of the D–S belief function is a result of generalizing the objective function of the optimization problem that the belief function is solving.

D. Representing the Probabilistic Constraints of Fuzzy Focal Elements Through Decomposition

To deal with fuzzy focal elements, we decompose them into nonfuzzy focal elements whose probabilistic constraints have been discussed in Section V-B. A fuzzy focal element has two components: a fuzzy subset of the frame of discernment and the probability mass assigned to the subset. In this section, we first describe how a fuzzy set can be decomposed into nonfuzzy sets. Then we define the decomposition of a fuzzy focal element.

An α -level set of A , a fuzzy subset of T , is a crisp set denoted by A_α that comprises all elements of T whose grade of membership in A is greater than or equal to α :

$$A_\alpha = \{x | \mu_A(x) \geq \alpha\}$$

A fuzzy set A may be decomposed into its level-sets through the *resolution identity* [15]:

$$A = \sum_{\alpha} \alpha A_\alpha$$

where the summation denotes the set union operation and αA_α denotes a fuzzy set with a two-valued membership function defined by

$$\begin{aligned}\mu_{\alpha A_\alpha}(x) &= \alpha \quad \text{for } x \in A_\alpha \\ \mu_{\alpha A_\alpha}(x) &= 0 \quad \text{elsewhere.}\end{aligned}$$

The importance of resolution identity is best described by Zadeh [15]: "The resolution identity provides a convenient way of generalizing various concepts associated with nonfuzzy sets to fuzzy sets" [15]. In fact, this is the underlying basis for many of the definitions of fuzzy set operations [16], [17].

In order to decompose a fuzzy focal element, we also need to decompose the focal's basic probability and distribute it among the focal's level-sets. Obviously, the decomposition has to satisfy two conditions:

- 1) The decomposed basic probabilities must add up to the basic probability assigned to the fuzzy focal.

$$\sum_{\alpha} m(A_\alpha) = m(A)$$

- 2) The decomposed basic probabilities must not be negative.

$$m(A_\alpha) \geq 0.$$

Using Dubois and Prade's observation on the relationship between possibility distribution (i.e., membership function of a fuzzy focal) and nonfuzzy consonant focals [18], we reach a decomposition of the fuzzy focal's basic probability that satisfies the two conditions stated above.

Dubois and Prade have shown that if a BPA is a set of nested focal elements, $A_1 \supset A_2 \cdots \supset A_n$, they can be related to the possibility distribution induced, denoted as $\text{Poss}(x)$, as follows:³

$$m(A_i) = \pi_i - \pi_{i-1} \quad (18)$$

where $\pi_i = \inf_{x \in A_i} \text{poss}(x)$, $\pi_0 = 0$, and $\pi_n = 1$. This result can be directly applied to decompose a fuzzy focal element whose basic probability value is one (i.e., $m(A) = 1$) because the α -level sets of A form a set of nested focal elements. Since $\inf_{x \in A_{\alpha_i}} \text{poss}(x) = \alpha_i$, the π_i in (18) becomes the alpha value α_i of the level sets. Thus, we get

$$m(A_{\alpha_i}) = \alpha_i - \alpha_{i-1}. \quad (19)$$

We extend this idea to decompose fuzzy focal elements with arbitrary probability mass (i.e., $0 \leq m(A) \leq 1$) by multiplying the focal's mass with the right-hand side of (19). Formally, the decomposition of a fuzzy element is defined as follows.

Definition 4: The decomposition of a fuzzy focal element A is a collection of nonfuzzy subsets such that 1) they are A 's α -level sets that form a resolution identity,

³We have paraphrased Dubois and Prade's results for the convenience of our discussion.

and 2) their basic probabilities are

$$m(A_{\alpha_i}) = (\alpha_i - \alpha_{i-1}) \times m(A) \quad i = 1, 2, \dots, n \quad (20)$$

where $\alpha_0 = 0$ and $\alpha_n = 1$.

When the focal element is a crisp set, its decomposition is the focal itself because the decomposition contains only one level set, which corresponds to the membership degree "one." The relationship between the decomposition of a fuzzy focal element and Shafer's consonant focals is discussed further in Section VI-A.

The probabilistic constraint of a fuzzy focal is defined to be that of its decomposition, which is a set of nonfuzzy focals. Since we already know how to deal with nonfuzzy focals, decomposing a fuzzy focal into nonfuzzy ones allows us to calculate the belief functions that are constrained by the fuzzy focals.

Definition 5: The probability mass that a fuzzy focal A contributes to the belief (and plausibility) of a fuzzy subset B is the total contribution of A 's decomposition to B 's belief (and plausibility), i.e.,

$$m^*(B; A) = \sum_{\alpha} m^*(B; A_\alpha) \quad (21)$$

$$m_*(B; A) = \sum_{\alpha} m_*(B; A_\alpha). \quad (22)$$

E. Computing the Belief Function

Based on generalizing the objective function and expressing the probabilistic constraints of fuzzy focal elements through their decompositions, we are able to derive the following formula for computing the belief function and the plausibility function.

$$\text{bel}(B) = \sum_A m(A) \sum_{\alpha_i} [\alpha_i - \alpha_{i-1}] \times \inf_{x \in A_{\alpha_i}} \mu_B(x) \quad (23)$$

$$\text{Pls}(B) = \sum_A m(A) \sum_{\alpha_i} [\alpha_i - \alpha_{i-1}] \times \sup_{x \in A_{\alpha_i}} \mu_B(x). \quad (24)$$

It is also trivial to show that the derived formulas preserve the following important property of the D-S theory: The belief of a (fuzzy) set is the difference of one and the plausibility of the set's complement.

1) An Example: The following example illustrates how one applies the formula described in Section V-E for computing the belief function. Suppose the frame of discernment is the set of integers between 1 and 10. A fuzzy basic probability assignment consists of two focal elements A and C :

$$\begin{aligned}A &= \{0.25/1, 0.5/2, 0.75/3, 1/4, 1/5, \\ &\quad 0.75/6, 0.5/7, 0.25/8\} \\ C &= \{0.5/5, 1.6, 0.8/7, 0.4/8\}\end{aligned}$$

where each member of the list is in the form of $\mu_A(x_i)/x_i$. We are interested in the degree of belief and the degree of plausibility of the fuzzy subset B :

$$B = \{0.5/2, 1/3, 1/4, 1/5, 0.9/6, 0.6/7, 0.3/8\}.$$

The decomposition of fuzzy focal A consists of four nonfuzzy focals:

$$\begin{aligned} A_{0.25} &= \{1, 2, \dots, 8\} \text{ with mass } 0.25 \times m(A) \\ A_{0.5} &= \{2, 3, \dots, 7\} \text{ with mass } 0.25 \times m(A) \\ A_{0.75} &= \{3, 4, \dots, 6\} \text{ with mass } 0.25 \times m(A) \\ A_1 &= \{4, 5\} \text{ with mass } 0.25 \times m(A) \end{aligned}$$

and the decomposition of fuzzy focal C also consists of four nonfuzzy focals:

$$\begin{aligned} C_{0.4} &= \{5, 6, 7, 8\} \text{ with mass } 0.4 \times m(C) \\ C_{0.5} &= \{5, 6, 7\} \text{ with mass } 0.1 \times m(C) \\ C_{0.8} &= \{6, 7\} \text{ with mass } 0.3 \times m(C) \\ C_1 &= \{6\} \text{ with mass } 0.2 \times m(C) \end{aligned}$$

Let us denote $\inf_{x \in A_i} \mu_B(x)$ as $f_{B,A}(\alpha_i)$. So, we have

$$\begin{aligned} m_*(B: A) &= m(A) \times [0.25 \times f_{B,A}(0.25) + 0.25 \times f_{B,A}(0.5) \\ &\quad + 0.25 \times f_{B,A}(0.75) + 0.25 \times f_{B,A}(1)] \\ &= m(A) \times [0.25 \times 0 + 0.25 \times 0.5 + 0.25 \times 0.9 + 0.25 \times 1] \\ &= 0.6 \times m(A) \end{aligned}$$

$$\begin{aligned} m_*(B: C) &= m(C) \times [0.4 \times f_{B,C}(0.4) + 0.1 \times f_{B,C}(0.5) \\ &\quad + 0.3 \times f_{B,C}(0.8) + 0.2 \times f_{B,C}(1)] \\ &= m(C) \times [0.4 \times 0.3 + 0.1 \times 0.6 + 0.3 \times 0.6 + 0.2 \times 0.9] \\ &= 0.54 \times m(C) \end{aligned}$$

Thus, we have

$$\text{bel}(B) = 0.6 m(A) + 0.54 m(C)$$

Similarly, we can calculate the plausibility of B :

$$\text{Pls}(B) = m(A) + 0.86 m(C)$$

2) *A Comparison with Alternative Approaches*: In this section, we will use the example discussed in Section V-E-1) to compare our approach with the alternative fuzzy evidential reasoning methods discussed in Section IV. The degrees of belief in the fuzzy set B computed using these methods are listed as follows:

$$\text{Ishizuka: } \text{bel}(B) = 0.75 m(A) + 0.8 m(C).$$

$$\text{Yager: } \text{bel}(B) = 0.5 m(A) + 0.6 m(C).$$

$$\text{Ogawa: } \text{bel}(B) = 0.8962 m(A) + 0.434 m(C).$$

We will compare how these results are changed in response to a change of fuzzy focal element. More specifically, we change the membership function of the fuzzy focal element A in three different ways. First, we increase the gradient of $\mu_A(x)$ for $1 \leq x \leq 3$ while keeping $\mu_A(2)$ unchanged. The modified focal element, denoted as A' , is

$$A' = \{0.166/1, 0.5/2, 0.833/3, 1/4, 1/5, 0.75/6, 0.5/7, 0.25/8\}.$$

TABLE I
THE CONTRIBUTION TO $\text{bel}(B)$ FROM THE FOCAL ELEMENT A
AND ITS VARIATIONS

Focal Elements	Yager	Ishizuk	Ogawa	Yen
A	0.5	0.75	0.8962	0.6
A'	0.5	0.834	0.9119	0.6252
A''	0.5	0.75	0.9434	0.5
A'''	0.5	1	0.8962	0.675

TABLE II
CHANGES TO $\text{bel}(B)$ DUE TO CHANGES IN THE FOCAL ELEMENT A

Changes of Focal Element A	Yager	Ishizuk	Ogawa	Yen
$A \rightarrow A'$	unchanged	increased	increased	increased
$A \rightarrow A''$	unchanged	unchanged	increased	decreased
$A \rightarrow A'''$	unchanged	increased	unchanged	increased

Second, we modify A into A'' by increasing the gradient of $\mu_A(x)$ for $1 \leq x \leq 3$ while preserving the membership value $\mu_A(1)$:

$$A'' = \{0.25/1, 0.75/2, 1/3, 1/4, 1/5, 0.75/6, 0.5/7, 0.25/8\}.$$

Finally, we get A''' by decreasing the membership value $\mu_A(1)$ while maintaining the membership values of other points:

$$A''' = \{0/1, 0.5/2, 0.75/3, 1/4, 1/5, 0.75/6, 0.5/7, 0.25/8\}.$$

Since only the focal element A has been changed, we can analyze the impact to the belief function by comparing the contributions of the focal element A and its variations to the degree of belief in B . Table I lists the portion of each modified focal's mass that contributes to B 's belief measure (i.e., the ratio $m_*(B: A)/m(A)$) for each fuzzy evidential reasoning method.

Table II shows how $\text{bel}(B)$ computed by different methods change as the focal element A changes in three ways. As shown in the table, Yager's method is insensitive to any of the three changes in the focal's membership function; Ishizuka's method is insensitive to a change from A to A'' ; and Ogawa's approach is insensitive to a change from A to A''' . Our approach is sensitive to all three kinds of changes in the focal's membership function.

This comparison indicates that previous approaches to generalizing the Dempster-Shafer model to fuzzy sets are not always responsive to a change of the focal element. In general, Ishizuka's belief function and Yager's belief function are insensitive to a focal element's change unless it results in a change of the "critical point," a point whose membership value is the minimal value in Equation (5) and Equation (6) for computing the inclusion measure, i.e.,

$$\mu_A(x_I) = \min_x [1, 1 + (\mu_B(x) - \mu_A(x))]$$

$$\mu_A(x_Y) = \min_x [\mu_{\bar{A}}(x) \vee \mu_B(x)]$$

where x_I and x_Y denote the critical points for Ishizuka's inclusion measure and Yager's inclusion measure respectively. In our example, the critical points for Yager's inclusion measure $I_Y(A \subset B)$ and Ishizuka's inclusion measure $I_I(A \subset B)$ are $x_Y = 2$ and $x_I = 1$ respectively. As the focal element A changes to A' , the critical point for Yager's inclusion measure remains the same. As a result, Yager's belief measure of the fuzzy subset B remains unchanged. Similarly, a change from A to A'' does not change the critical point x_I . Hence, Ishizuka's belief measure of B remains the same in this case.

Ogawa's belief measure of a fuzzy subset B is not responsive to a change in the focal element's membership function unless the intersection between the focal and the fuzzy subset B is different. Since the intersection $A \cap B$ is the same as $A'' \cap B$, Ogawa's belief measure of B remains unchanged when the focal A changes to A'' .

A surprising result of this comparison is that a change from A to A'' increases Ogawa's belief measure, but decreases ours. This can be explained as follows. Ogawa's measure of inclusion is based on the sigma count of $A \cap B$ relative to the sigma count of B . Since the intersection of A and B , Ogawa's measure of inclusion increases as the focal change from A to A'' . However, our belief measure in B decreases because the level set of A'' at membership degree 0.75 contributes less to the belief measure $\text{bel}(B)$ than A 's level-set at 0.75 does (i.e., $f_{B,A''}(0.75) = 0.5$ is less than $f_{B,A}(0.75) = 0.9$) while the contributions of all other level sets remain the same.

In summary, the comparison above indicates that our method for computing the belief function of fuzzy sets is more responsive to any change to a focal element's membership function than previous approaches are. Moreover, a change in our belief measure can always be explained in terms of a change in the underlying probabilistic constraints imposed by the focal elements.

F. Generalizing Dempster's Rule of Combination

Dempster's rule combines the effects of two independent evidential sources, denoted as R and S , on the probability distribution of a hypothesis space, denoted as T . The rule can be viewed as a result of three steps.

- 1) *Combine the compatibility relations.* A combined compatibility relation between the product space $R \times S$ and T can be constructed from the compatibility relation between R and T and the one between S and T using the following principle:

$$rCt \text{ and } sCt \Rightarrow [r, s]Ct$$

where r , s , t , and $[r, s]$ denote elements of R , S , T , and $R \times S$ respectively. As a result, the granule of $[r, s]$ under the combined multivalued mapping is the intersection of the granule of r and the granule of s , i.e.,

$$G([r, s]) = G(r) \cap G(s). \quad (25)$$

This explains why focal elements of different evidential sources are intersected in Dempster's rule.

- 2) *Compute joint probability distributions of the combined evidential source.* Since R and S are assumed to be independent, the joint probability distribution of the space $R \times S$ can be computed from the probability distribution of each individual space:

$$P([r, s]) = P(r) \times P(s)$$

- 3) *Normalize the combined basic probability assignment.* Having obtained the probability distribution of $R \times S$ and the compatibility relation between $R \times S$ and T from the two previous steps, Dempster's rule follows directly from (1), which includes a normalization process to discard probability mass assigned to the empty set.

Two generalizations must be made to Dempster's rule before it can be used to combine fuzzy BPA's in our generalized framework: 1) the first step above has to be extended to allow the combination of fuzzy compatibility relations; and 2) the normalization step needs to consider subnormal fuzzy focal elements that result from combining fuzzy compatibility relations.

1) *Combination of Fuzzy Compatibility Relations:* By employing the noninteractiveness assumption in possibility theory, we generalize equation (25) in order to perform fuzzy intersection to obtain granules of the combined compatibility relation. A compatibility relation in our generalized D-S framework, as discussed in Section V-A, is a joint possibility distribution. Thus, we have

$$C(r, t) = \prod_{X, Z} (r, t) \quad \text{and} \quad C(s, t) = \prod_{Y, Z} (s, t) \quad (26)$$

where X , Y , and Z are variables that take values from the spaces R , S , and T , respectively. Let W be a variable that takes values from the space $R \times S$. The combined fuzzy compatibility relation can be expressed as

$$C([r, s], t) = \prod_{W, Z} ([r, s], t) = \prod_{X, Y, Z} (r, s, t).$$

Marginal possibility distributions $\Pi_{X, Z}$ and $\Pi_{Y, Z}$ are the projection of joint possibility distribution on Y and X respectively, [12] i.e.,

$$\prod_{Y, Z} (s, t) = \max_r \prod_{X, Y, Z} (r, s, t)$$

$$\prod_{X, Z} (r, t) = \max_s \prod_{X, Y, Z} (r, s, t).$$

Hence, the joint possibility distribution is bounded by the marginal possibility distributions:

$$\prod_{X, Y, Z} (r, s, t) \leq \prod_{Y, Z} (s, t) \wedge \prod_{X, Z} (r, t)$$

where \wedge denotes the minimum operator. By employing the assumption that the variables Y , Z and X , Z are noninteractive, a concept analogous to the independence of random variables, we obtain the following joint possibility distribution:

$$\prod_{X, Y, Z} (r, s, t) = \prod_{Y, Z} (s, t) \wedge \prod_{X, Z} (r, t).$$

Thus, the combined fuzzy compatibility relation is

$$C([r, s], t) = C(r, t) \wedge C(s, t). \quad (27)$$

For a fixed pair of r and s , applying equation (27) to all possible elements in T gives us the following relationship between conditional possibility distributions:

$$\prod_{(Z|W=[r,s])} = \prod_{(Z|X=r)} \cap \prod_{(Z|Y=s)}$$

where \cap denotes the fuzzy intersection operator. Equivalently, the granule of the pair $[r, s]$ under the combined compatibility relation defined in (27) is the fuzzy intersection of $G(r)$ and $G(s)$:

$$G([r, s]) = G(r) \cap G(s)$$

2) *Normalizing Subnormal Fuzzy Focal Elements:* An important assumption of our work is that all focal elements are normal. We avoid subnormal fuzzy focal elements because they assign probability mass to the empty set. For example, suppose A is a fuzzy subset of the frame of discernment $\{x_0, x_1, x_2, x_3, x_4\}$, characterized by the membership function

$$A = \{0/x_0, 0.1/x_1, 0.2/x_2, 0.1/x_3, 0/x_4\}.$$

Let the basic probability value of the set A be “ a ”. The decomposition of this focal element A is:

$$A_{0,1} = \{x_1, x_2, x_3\} \text{ with mass } 0.1 \times a$$

$$A_{0,2} = \{x_2\} \text{ with mass } 0.1 \times a$$

$$A_1 = \phi \text{ with mass } 0.8 \times a$$

In general, the probability mass assigned to the empty set by a subnormal fuzzy focal A is the basic probability assigned to the decomposed focal of A that is constructed from A 's α -level set at the degree of membership one:

$$\left[1 - \max_x \mu_A(x)\right] \times m(A).$$

Although we have assumed that the focal elements of fuzzy BPA's are all normal, the intersections of focals may be subnormal. Hence, the combination of fuzzy BPA's should deal with the normalization of subnormal fuzzy focal elements. To do this, we need to normalize the two components of a fuzzy focal element: the focal itself, which is a subnormal fuzzy set, and the probability mass assigned to the focal.

It is straightforward to normalize the focal. Suppose A is a subnormal fuzzy set characterized by the membership function $\mu_A(x)$. A 's normalized set, denoted as \bar{A} , is characterized by the following membership function.

$$\mu_{\bar{A}}(x) = \frac{\mu_A(x)}{\max_x \mu_A(x)} = k \times \mu_A(x)$$

where k is the normalization factor

$$k = 1 / \max_x \mu_A(x).$$

The criterion for normalizing the probability mass of a subnormal focal is that the probabilistic constraints imposed by the subnormal focal should be preserved after the normalization. Since we use the decomposition of a

focal to represent its probabilistic constraint, this means that the probability mass assigned to a decomposed focal should not be changed by the normalization process. Since the α_i cut of the subnormal focal becomes the $k\alpha_i$ cut of the normalized focal, the probability mass assigned to them should be the same:

$$m(A_{\alpha_i}) = m(\bar{A}_{k\alpha_i}). \quad (28)$$

From this condition, we can derive the relationship between $m(\bar{A})$ and $m(A)$ as follows. The left-hand side of (28) can be rewritten as

$$m(A_{\alpha_i}) = m(A)(\alpha_i - \alpha_{i-1}).$$

The right-hand side of (28) can be rewritten as

$$m(\bar{A}_{k\alpha_i}) = m(\bar{A})(k\alpha_i - k\alpha_{i-1}) = k m(\bar{A})(\alpha_i - \alpha_{i-1}).$$

It follows from the three equations above that the mass of the normalized focal is reduced by a factor reciprocal to the ratio by which its membership function is scaled up:

$$m(\bar{A}) = m(A) / k.$$

The remaining mass $(1 - 1/k)m(A)$ is the amount assigned to the empty set by the subnormal fuzzy focal and, hence, should be part of the normalization factor in the generalized Dempster's rule.

We summarize our approach to normalize a subnormal focal element into three steps:

- a) Scale up the membership function so that its peak (i.e., highest membership degree) is one.
- b) Reduce the basic probability using a ratio reciprocal to the scaling factor of the first step.
- c) Assign the basic probability lost during the second step to the empty set.

3) *A Generalized Rule of Combination:* Commutativity is an important requirement for any evidence combination rule, because it is highly desirable to have the effect of the aggregated evidence independent of the order of combination. It is well known that Dempster's rule is commutative [9, p. 62]. Our normalization step discussed in Section V-F-2 is not commutative because it modifies the membership functions of the focal elements' subnormal intersections. To solve this problem, we first show that the normalization process in Dempster's rule can be postponed without changing the combination result. Then, we describe our generalized combining rule where the normalization process is postponed to achieve commutativity.

Normalization in Dempster's rule does not have to apply after each combining operation. It can be postponed to a later point without changing the result. More specifically, several BPA's in the D-S theory can be combined without normalization, and the normalized combined bpa can be obtained by applying the normalization process to the unnormalized combined BPA at the end. In the following discussion, we use the symbol \otimes to denote Dempster's rule without normalization (i.e., the denominator in (4) is one), the letter “N” to denote the

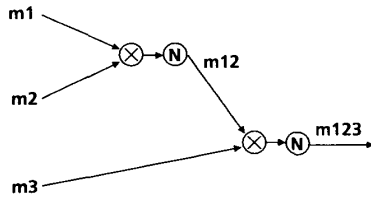


Fig. 1. Combination of evidence with immediate normalization.

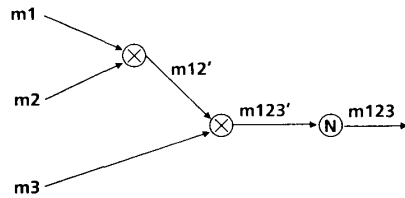


Fig. 2. Combination of evidence with postponed normalization.

normalization process, and the primed letter m' to denote the unnormalized BPA. Fig. 1 and Fig. 2 show two ways to apply Dempster's rule: combine BPA's with immediate normalization, or combine BPA's with postponed normalization. To show that they obtain the same result, we consider three BPA's of a frame of discernment: m_1 , m_2 , and m_3 . We want to show that applying normalization after the three BPA's are combined without normalization yields the same result as using Dempster's rule in the conventional way to combine them, i.e.,

$$(m_1 \oplus m_2) \oplus m_3 = N[(m_1 \otimes m_2) \otimes m_3] \quad (29)$$

We first expand the result of combining the first two BPA's using Dempster's rule.

$$m_1 \oplus m_2(C) = \frac{m'_{12}(C)}{1 - k_{12}}$$

where

$$m'_{12}(C) = \sum_{A \cap B = C} m_1(A)m_2(B) \quad (30)$$

and

$$k_{12} = \sum_{A \cap B = \phi} m_1(A)m_2(B) \quad (31)$$

The left-hand side of (29) thus becomes

$$(m_1 \oplus m_2) \oplus m_3(E) = \frac{\frac{1}{1 - k_{12}} \sum_{C \cap D = E} m'_{12}(C)m_3(D)}{1 - \frac{1}{1 - k_{12}} \sum_{C \cap D = \phi} m'_{12}(C)m_3(D)}$$

Substituting $m'_{12}(C)$ with the right-hand side of (30), we get

$$= \frac{\frac{1}{1 - k_{12}} \sum_{A \cap B \cap D = E} m_1(A)m_2(B)m_3(D)}{1 - \frac{1}{1 - k_{12}} \sum_{A \cap B \cap D = \phi, A \cap B \neq \phi} m_1(A)m_2(B)m_3(D)}$$

Multiplying both the numerator and the denominator by $1 - k_{12}$, we have

$$= \frac{\sum_{A \cap B \cap D = E} m_1(A)m_2(B)m_3(D)}{1 - k_{12} - \sum_{A \cap B \cap D = \phi, A \cap B \neq \phi} m_1(A)m_2(B)m_3(D)}$$

Substituting k_{12} with the right-hand side of (31), we get

$$= \frac{\sum_{A \cap B \cap D = E} m_1(A)m_2(B)m_3(D)}{1 - \left[\sum_{A \cap B = \phi} m_1(A)m_2(B) + \sum_{A \cap B \cap D = \phi, A \cap B \neq \phi} m_1(A)m_2(B)m_3(D) \right]}$$

Since $\sum_D m_3(D) = 1$, we can reformulate the normalization factor:

$$= \frac{\sum_{A \cap B \cap D = E} m_1(A)m_2(B)m_3(D)}{1 - \left[\sum_{A \cap B \cap D = \phi, A \cap B = \phi} m_1(A)m_2(B)m_3(D) + \sum_{A \cap B \cap D = \phi, A \cap B \neq \phi} m_1(A)m_2(B)m_3(D) \right]}$$

Finally, we get

$$= \frac{\sum_{A \cap B \cap D = E} m_1(A)m_2(B)m_3(D)}{1 - \sum_{A \cap B \cap D = \phi} m_1(A)m_2(B)m_3(D)} = N[(m_1 \otimes m_2) \otimes m_3].$$

Hence, we have shown that the normalization step in Dempster's rule can be delayed without changing the result of combination.

Our generalized rule of combination consists of two operations: a cross-product operation and a normalization process. Fuzzy BPA's are first combined by performing the following generalized cross product:

$$m'_{12}(C) = m_1 \otimes m_2(C) = \sum_{A \cap B = C} m_1(A)m_2(B). \quad (32)$$

where \cap denotes the fuzzy intersection operator and C is an unnormalized intersection of focal elements, which could be a subnormal fuzzy subset of the frame of discernment. The empty set is a special kind of subnormal focal elements. To compute the normalized combined BPA (e.g., for computing its belief function), we apply the following normalization process (discussed in Section V-F-2)) to the unnormalized combined BPA:

$$N[m'](D) = \frac{\sum_{\bar{C} = D} \max_{x_i} \mu_C(x_i) m'(C)}{1 - \sum_{C \subset T} \left(1 - \max_{x_i} \mu_C(x_i) \right) m'(C)}. \quad (33)$$

For example, if we need to combine three bpa's of the frame of discernment T , the result of combination is computed by first combining the three bpa's without normalization using (32), and then normalizing the final

result:

$$m_1 \oplus m_2 \oplus m_3 = N[(m_1 \otimes m_2) \otimes m_3].$$

It is obvious that the generalized cross-product operation is commutative, e.g.,

$$N[(m_1 \otimes m_2) \otimes m_3] = N[m_1 \otimes (m_2 \otimes m_3)].$$

Thus, through delaying the normalization process, we are able to combine fuzzy BPA's in an order-independent fashion.

In the special case where there are only two fuzzy BPA's to be combined, the combined BPA using the generalized Dempster's rule of combination is

$$\begin{aligned} m_1 \oplus m_2(C) &= N[m_1 \otimes m_2](C) \\ &= \frac{\sum_{(A \cap B) = C} \max_{x_i} \mu_{A \cap B}(x_i) m_1(A) m_2(B)}{1 - \sum_{A, B} \left(1 - \max_{x_i} \mu_{A \cap B}(x_i)\right) m_1(A) m_2(B)} \end{aligned} \quad (34)$$

The normalization process (i.e., (33)) generalizes the notion of *conflicting evidence* in the D-S theory to that of *partially conflicting evidence*. In Dempster's original rule, two pieces of evidence are either in conflict (i.e., the intersection of their focals is empty) or not in conflict at all (i.e., the intersection of their focals is not empty). In our generalized combining rule, two pieces of evidence are partially in conflict if the intersection of their focals is subnormal. The *degree of conflict* is measured by the difference between one and the peak (i.e., the maximum value) of the focal's membership function. The case of peak being zero corresponds to the case of total conflict in the D-S theory.

Our extension to Dempster's rule differs from Ishizuka's extension (discussed in Section IV) in its handling of subnormal intersections of focal elements. Ishizuka's degree of intersection $J(A, B)$ becomes $\max_{x_i} \mu_{A \cap B}(x_i)$ in (34) when both fuzzy set A and fuzzy set B are normal; therefore, it is analogous to the factor that scales down the basic probability in the normalization step of our approach. While we use the reciprocal of the factor to scale up the membership function of the focals' intersection, Ishizuka does not normalize the intersection. More importantly, Ishizuka's approach appeals to intuition without rigorous justification, whereas our approach is derived from the principle that the normalization step should preserve the relative probabilistic constraints imposed by focal elements, whether it is normal or not.⁴

One of the most controversial issues regarding Dempster's rule of combination has been its normalization process. Zadeh, for instance, has questioned the validity of discarding the probability mass assigned to the empty

set because the probability mass is an indication of the degree of conflict between the evidential sources that are combined [19]. However, to be consistent with axioms of probability theory, the probability of empty set has to be zero. In our approach, this dilemma is solved by delaying the normalization process. By computing the unnormalized BPA of the frame of discernment, our generalized rule of combination is able to use the basic probability of the empty set as a measure of the degree of conflict, which influences the credibility of the combined evidential sources. In the meantime, we can obtain the normalized BPA, which is needed for computing the belief function, by applying the normalization step to the unnormalized BPA. Hence, the generalized Dempster's rule not only allows the combination of vague evidential opinions, but also provides information regarding the credibility of the combined opinion.

VI. DISCUSSION

A. Consonant Focals and Fuzzy Focals

Several authors have discussed the similarity between possibility distribution and one specific instance of the D-S plausibility function called *consonant support function*—when the focal elements are nested, i.e., when they can be arranged in order so that each focal is contained in the following one [10]. Based on this observation, we have defined the probabilistic constraint of a fuzzy focal to be that of a set of consonant crisp focals in Section V-D. Here, we will focus on the differences between the consonant focal elements and the fuzzy focal element.

A set of consonant focal elements differs from a fuzzy focal element in two important ways. First, consonant focal elements are more restrictive in the kinds of fuzzy evidential support they can represent. More specifically, they are limited to representing single vague evidential support. A fuzzy basic probability assignment (BPA), however, may consist of several fuzzy focal elements. Hence, it can express multiple fuzzy evidential supports. Second, each fuzzy focal element is induced by single evidential elements, while consonant focals are induced by several evidential elements that form an *inferential evidence* [9]. This difference between fuzzy focals and consonant focals explains their different combination results. The combination of two consonant BPA's is a result of combining their evidential elements pairwise. Therefore, the combined focals are, in general, no longer consonant. However, the combination of two fuzzy focal elements, which involves the combination of underlying fuzzy compatibility relations, always yields another fuzzy focal element.

Due to these significant differences between fuzzy focals and consonant crisp focals, we should emphasize that we do not view fuzzy focal elements as identical to consonant crisp focals. In other words, the decomposition of a fuzzy focal element is not equivalent to the fuzzy focal itself. A fuzzy focal and its decomposition are only equivalent in the probabilistic constraints they imposed on the probability distribution of the frame of discernment.

⁴Obviously, the absolute probabilistic constraints of non-empty focal elements are not preserved by the normalization process because their basic probabilities are increased by the normalization factor (i.e., the denominator in (33)).

VII. CONCLUSION

We have described a generalization of the Dempster-Shafer theory to fuzzy sets. Rather than generalizing the formula for computing belief function, we generalize the basic constructs of the D-S theory: the compatibility relations, the objective functions of the optimization problem for calculating belief functions, and the probabilistic constraints imposed by focal elements. As a result, we can compute the lower probability (i.e., the belief function) directly from these generalized constructs. Moreover, by employing the noninteractive assumption in possibility theory, we have modified Dempster's rule to combine evidence that may be partially in conflict.

Our approach offers several advantages over previous work. First, the semantics of the D-S theory is maintained. Belief functions are treated as lower probabilities in our extension. Second, we avoid the problem of "choosing the right inclusion operators" faced by all previous approaches. Third, the generalized belief function is determined by the whole membership function of the focal element, not just by some critical points as used in some of the previous work. Any change of the membership function of a focal element is directly reflected in a change of the focal's probabilistic constraint, which in turn affects the belief function. Fourth, the generalized rule of combination provides information about the degree of conflict between the evidence combined by delaying the normalization step in original Dempster's rule. Finally, our generalization is well-justified using possibility theory and probability theory. Therefore, it serves as a bridge that brings together the Dempster-Shafer theory and fuzzy set theory into a hybrid approach to reasoning under various kinds of uncertainty in intelligent systems.

ACKNOWLEDGMENT

The author would like to thank Prof. Lotfi Zadeh for his continuous support and encouragement, which have made this work possible. Many thanks also go to Henri Prade, Piero Bonissone, and the reviewers for their valuable comments that help me in revising the paper.

REFERENCES

- [1] J. Pearl, "On evidential reasoning in a hierarchy of hypothesis," *Artificial Intell.*, vol. 28, no. 1, Feb. 1986, pp. 9-16.
- [2] S. A. Lesh, *An Evidential Theory Approach to Judgment-based Decision Making*, Ph.D. dissertation, Department of Forestry and Environmental Studies, Duke Univ., 1986.
- [3] "Fuzzy control ensures a smooth ride," *Hitachi* 1987, pp. 12-13.
- [4] L. A. Zadeh, "Fuzzy sets and information granularity," in *Advances in Fuzzy Set Theory and Applications*, 1979, pp. 3-18.
- [5] M. Ishizuka, K. S. Fu, and J. T. P. Yao, "Inference procedures and uncertainty for the problem-reduction method," *Inform. Sci.*, vol. 28, 1982, pp. 179-206.
- [6] R. Yager, "Generalized probabilities of fuzzy events from fuzzy belief structures," *Inform. Sci.*, vol. 28, 1982, pp. 45-62.
- [7] H. Ogawa and K. S. Fu, "An inexact inference for damage assessment of existing structures," *International Journal of Man-Machine Studies*, vol. 22, 1985, pp. 295-306.
- [8] A. P. Dempster, "Upper and lower probabilities induced by a multivalued mapping," *Ann. Math. Stat.*, vol. 38, pp. 325-339, 1967.
- [9] G. Shafer, *Mathematical Theory of Evidence*. Princeton, N.J.: Princeton Univ. Press, 1976.
- [10] —, "Belief functions and possibility measures," tech. report working paper no. 163, University of Kansas, School of Business, 1984.
- [11] L. A. Zadeh, "Possibility theory and soft data analysis," in *Mathematical Frontiers of the Social and Policy Sciences*, L. Cobb and R. M. Thrall, Eds. Boulder, CO: Westview Press, 1981, pp. 69-129.
- [12] —, "Fuzzy sets as a basis for a theory of possibility," *Fuzzy Sets and Systems*, vol. 1, pp. 3-28, 1978.
- [13] D. Dubois and H. Prade, *Possibility Theory*. New York: Plenum Press, 1988.
- [14] P. Smets, "The degree of belief in a fuzzy event," *Inform. Sci.*, vol. 25, pp. 1-19, 1981.
- [15] L. A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning-I," *Inform. Sci.*, vol. 8, pp. 199-249, 1975.
- [16] —, "Fuzzy languages and their relation to human and machine intelligence," *Proc. Int. Conf. Man Comput.*, S. Karger, Ed. Bordeaux, France: Basel, 1972, pp. 130-165.
- [17] —, "Similarity relations and fuzzy orderings," *Information Sci.*, vol. 3, pp. 177-200, 1971.
- [18] 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, Ed. New York: North-Holland, 1982, pp. 167-181.
- [19] L. A. Zadeh, "A simple view of the Dempster-Shafer theory of evidence and its implication for the Rule of Combination," *AI Magazine*, vol. 7, no. 2, pp. 85-90, Summer 1986.



John Yen received his B.S. in electrical engineering from the National Taiwan University in 1980, his M.S. in Computer Science from the University of Santa Clara, Santa Clara, CA, in 1982, and his Ph.D. in computer science from the University of California, Berkeley, CA, in 1986. His dissertation research on evidential reasoning developed a number of extensions to the Dempster-Shafer model of reasoning about uncertain evidence, and produced a medical diagnostic system called Gertis, which is the first operational expert system that implements evidential reasoning mechanisms based on the Dempster-Shafer approach.

He is an assistant professor of the Department of Computer Science at Texas A&M University. Before joining Texas A&M, he has been a Research Scientist at Information Sciences Institute of University of Southern California, Los Angeles, CA, since 1986. His research interests have been hybrid reasoning architecture and programming methodology, approximate reasoning, expert systems, intelligent user interfaces, knowledge representations, and neural networks. He has played a major role in the development of Backbord, a general-purpose knowledge base browser which was based on concepts developed in Rabbit at Xerox Parc. He has also developed a classification-based production system (Clasp), which fuses the best aspects of frames, rules, and object-oriented programming, on top of a hybrid knowledge representation system (Loom) developed at USC/ISI.