

Research Article

A Method to Determine Generalized Basic Probability Assignment in the Open World

Wen Jiang, Jun Zhan, Deyun Zhou, and Xin Li

School of Electronics and Information, Northwestern Polytechnical University, Xi'an, Shaanxi 710072, China

Correspondence should be addressed to Wen Jiang; jiangwen@nwpu.edu.cn

Received 11 November 2015; Revised 14 February 2016; Accepted 27 April 2016

Academic Editor: Kishin Sadarangani

Copyright © 2016 Wen Jiang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Dempster-Shafer evidence theory (D-S theory) has been widely used in many information fusion systems since it was proposed by Dempster and extended by Shafer. However, how to determine the basic probability assignment (BPA), which is the main and first step in D-S theory, is still an open issue, especially when the given environment is in an open world, which means the frame of discernment is incomplete. In this paper, a method to determine generalized basic probability assignment in an open world is proposed. Frame of discernment in an open world is established first, and then the triangular fuzzy number models to identify target in the proposed frame of discernment are established. Pessimistic strategy based on the differentiation degree between model and sample is defined to yield the BPAs for known targets. If the sum of all the BPAs of known targets is over one, then they will be normalized and the BPA of unknown target is assigned to 0; otherwise the BPA of unknown target is equal to 1 minus the sum of all the known targets BPAs. IRIS classification examples illustrated the effectiveness of the proposed method.

1. Introduction

As one of the most important tools in multisources information fusion, Dempster-Shafer evidence theory (D-S theory) [1, 2] has been widely used since it has abilities to deal with uncertainty and unknown information and requires fewer conditions than probability theory. D-S theory has a significant advantage over the traditional probabilistic approach; that is, it allows for the allocation of probability mass to sets or intervals. And it can hence handle both aleatory uncertainty and epistemic (or subjective) uncertainty.

Although D-S theory is widely used in many application systems such as information fusion systems [3–6], evidential reasoning [7, 8], risk and reliability analysis [9–12], and decision making [13–15], some basic problems are still not completely clarified. One typical open issue is dependent evidence combination [16–18]. The other open issue is how to determine the basic probability assignment (BPA). So far, there is no general method to obtain BPA.

In the years, many researchers have investigated different approaches to solve this problem. Xu et al. [19] put forward a nonparametric method to determine BPA based on

the relationship between the test sample and the probability models. In their work, a nonparametric probability density function (PDF) is calculated firstly using the training data. Suh and Yook [20] presented a method to determine BPA through sensor data. Their study showed that a pedestrian could recognize a moving object with its method of determining BPA through multisensor data fusion. Yoon et al. [21] proposed a novel way to obtain BPA using sensor signals. The method enabled context inference even when there was no advanced information of the situation. Recently, with the generalized fuzzy numbers [22], we presented a fuzzy method to generate BPA [23].

Another vital issue is that the result of BPAs combination is always contradictory to common sense, using D-S theory of evidence to fuse highly conflicting evidences. This problem often appears in engineering applications because environmental noises and human disturbances often lead to conflict among the reports of multiple sensors.

Zadeh, the author of fuzzy mathematics, gives an example to analyze the deficiency of the combination rule of evidence theory, which has aroused the interest of many scholars [24]. How to handle conflicting evidence is heavily studied [25,

26]. Yager [27] holds on to the fact that the normalization of the combination rule is the main reason which leads to irrational fusion results. He advocates assigning the conflict coefficient to the unknown items in the frame of discernment (FOD) directly, without normalization step. Yager's method and the method in [28] only satisfy commutativity, but not associativity. This limits the real application of evidence theory. Smets and Kennes [29] present another ingenious approach. They argue that the conflict is mainly due to the incomplete knowledge base when the reports of sensors are all reliable, so they put forward the concepts of the closed world and the open world. In the transferable belief model (TBM), the so-called closed world means that the FOD consists of all the possible propositions, and the open world refers to the incomplete FOD due for lack of knowledge and various uncertainty. They point out that the methods of Dempster et al. combine BPAs on the premise of assumptions that the information fusion environment is in a closed world.

Based on the conception of closed world and open world, Deng [30] proposed the basic frame of generalized evidence theory (GET). In GET, \emptyset is regarded as an element with the same properties as the other elements. It represents unknown, but not a common empty. For example, in military applications, suppose there are three targets (a , b , and c) on the FOD. Then, the sensors can only recognize the different unions of these three targets. However, if there exists an unknown target (d), the sensors cannot distinguish whether it is one of the previous three targets. In GET, \emptyset can be interpreted as the unknown targets and $m(\emptyset) \neq 0$. GET is a proper tool to build the real world in a rational way. In the open world, the generalized combination rule (GCR) can solve the problem that the result of BPAs combination is counterintuitive because of the incompleteness of FOD. In the closed world, GET degenerates to D-S theory.

As mentioned above, there are already various methods in the closed world. However, there's no method to determine BPA in the open world. Determining BPA in the open world is a question worth to be explored. This paper proposed a new method to determine BPA in the open world. This method can be applied to classification problems and identification problems, especially when the FOD is not complete.

The remainder of this paper is organized as follows. Section 2 starts with a brief presentation of D-S theory and some necessary related concepts. The proposed method to construct BPA function is presented in Section 3. Section 4 investigates the effectiveness of the proposed method through a classical classification problem. Conclusions are presented in Section 5.

2. Preliminaries

2.1. Dempster-Shafer Evidence Theory. D-S theory, introduced by Dempster [1] and extended later by Shafer [2], is concerned with the question of belief in a proposition and systems of propositions. It mainly focuses on the epistemic uncertainty, but it is also valid for aleatoric uncertainty. D-S theory has many advantages, compared to probability theory. For instances, it can handle more uncertainty in real world. In probability theory, the belief can be only assigned to singleton

subsets, while in D-S theory the belief can be assigned to any subsets of FOD. And in D-S theory, prior distribution is not necessary before information fusion. Due to its ability to handle uncertainty or imprecision embedded in the evidence, D-S theory has been increasingly applied in many fields [31–35]. Formally, D-S theory concerns the following preliminary notations.

2.1.1. Frame of Discernment and Mass Function. Evidence theory first supposes the definition of a set of hypotheses H_i called the frame of discernment, defined as $\Theta = \{H_1, H_2, \dots, H_N\}$. The set Θ is composed of N exhaustive and exclusive hypotheses. Denote $P(\Theta)$, the power set composed of 2^N propositions of Θ , as

$$P(\Theta) = \{\emptyset, \{H_1\}, \{H_2\}, \dots, \{H_N\}, \{H_1 \cup H_2\}, \{H_1 \cup H_3\}, \dots, \Theta\}, \quad (1)$$

where \emptyset denotes the empty set. The N subsets containing only one element each are called singletons.

When the FOD is determined, the mass function m is defined as a mapping of the power set $P(\Theta)$ to a number between 0 and 1; that is,

$$m : P(\Theta) \mapsto [0, 1] \quad (2)$$

which satisfies the following conditions:

$$\begin{aligned} \sum_{A \in P(\Theta)} m(A) &= 1, \\ m(\emptyset) &= 0. \end{aligned} \quad (3)$$

The mass function m is also called the basic probability assignment (BPA) function. $m(A)$ expresses the proportion of all relevant and available evidence that supports the claim that a particular element of Θ belongs to the set A but to no particular subset of A . Any subset A of Θ such that $m(A) > 0$ is called a focal element.

2.1.2. Belief and Upper Probability Functions [2]. A function $\text{Bel} : 2^\Theta \rightarrow [0, 1]$ is called a belief function over Θ if it is given by (4) for some basic probability assignment $m : 2^\Theta \rightarrow [0, 1]$. Consider

$$\text{Bel}(A) = \sum_{B \subseteq A} m(B). \quad (4)$$

Whenever Bel is belief function over a frame Θ , the function $P^* : 2^\Theta \rightarrow [0, 1]$ defined by (5) is called the upper probability function of A :

$$P^*(A) = 1 - \text{Bel}(\bar{A}). \quad (5)$$

2.1.3. Dempster's Combination Rule [2]. Suppose Bel_1 and Bel_2 are belief functions over the same frame Θ , with BPA m_1 and m_2 and focal elements A_1, \dots, A_k and B_1, \dots, B_ℓ , respectively. Then the function $m : 2^\Theta \rightarrow [0, 1]$ defined by $m(\emptyset) = 0$ and

$$m(A) = \frac{\sum_{B \cap C = A} m_1(B) m_2(C)}{1 - k} \quad (6)$$

with

$$k = \sum_{B \cap C = \emptyset} m_1(B) m_2(C) \quad (7)$$

for all nonempty $A \subset \Theta$ is a BPA. The core of the belief function given by m is equal to the intersection of the cores of Bel_1 and Bel_2 . It is also called the *orthogonal sum* of Bel_1 and Bel_2 and is denoted by $\text{Bel}_1 \oplus \text{Bel}_2$.

2.2. Jousselme Distance. Jousselme et al. [36] proposed a new distance to measure the difference between two bodies of evidence, which is also called the evidence distance.

Let m_1 and m_2 be two BPAs on the same FOD Θ , containing N mutually exclusive and exhaustive hypotheses. The distance between m_1 and m_2 is

$$d_{\text{BPA}}(m_1, m_2) = \sqrt{\frac{1}{2} (\vec{m}_1 - \vec{m}_2)^T \underline{D} (\vec{m}_1 - \vec{m}_2)}, \quad (8)$$

where \vec{m}_1 and \vec{m}_2 are the BPAs according to (3) in Subsection 2.1 and \underline{D} is a $2^N \times 2^N$ matrix whose elements are

$$D(A, B) = \frac{|A \cap B|}{|A \cup B|}, \quad (9)$$

where $A, B \in P(\Theta)$ are derived from m_1 and m_2 , respectively.

2.3. Fuzzy Sets Theory. Fuzzy sets theory was first proposed by Zadeh [37] in 1965. The theory is widely used in many uncertain environments such as decision making [38–41] and optimization [42–44]. Some relative notions on fuzzy sets are given as follows.

Definition 1 (fuzzy set). Let X be a universe of discourse, where \tilde{A} is a fuzzy subset of X ; and for all $x \in X$, there is a number $\mu_{\tilde{A}}(x) \in [0, 1]$ which is assigned to represent the membership degree of x in \tilde{A} and is called the membership function of \tilde{A} [45].

Definition 2 (fuzzy number). A fuzzy number \tilde{A} is a normal and convex fuzzy subset of X [45].

Here, “normality” implies that

$$\bigvee_x \mu_{\tilde{A}}(x) = 1, \quad \exists x \in \mathbb{R}, \quad (10)$$

and “convex” means that

$$\mu_{\tilde{A}}(\alpha x_1 + (1 - \alpha) x_2) \geq \min(\mu_{\tilde{A}}(x_1), \mu_{\tilde{A}}(x_2)) \quad (11)$$

$$\forall x_1 \in X, x_2 \in X, \forall \alpha \in [0, 1].$$

Definition 3 (generalized fuzzy numbers). A generalized fuzzy number $A = (a_1, a_2, a_3, a_4; \omega)$ is described as any fuzzy subset of the real line R with membership function f_A that possesses the following features:

- (1) $f_A(x) : R \rightarrow [0, \omega]$ is continuous, $0 \leq \omega \leq 1$.
- (2) $f_A(x) = 0$, for all $x \in (-\infty, a_1]$.

(3) $f_A(x)$ is strictly increasing in $[a_1, a_2]$.

(4) $f_A(x) = \omega$, for $x \in [a_2, a_3]$, where ω is a constant and $0 < \omega \leq 1$.

(5) $f_A(x)$ is strictly decreasing in $[a_2, a_3]$.

(6) $f_A(x) = 0$, for all $x \in [a_4, \infty)$.

Particularly, a trapezoidal fuzzy number and triangular fuzzy number can be shown in Figures 1(a) and 1(b), respectively.

2.4. Generalized Evidence Theory [30]. Generalized evidence theory (GET), based on the classical Dempster-Shafer (D-S) theory, was proposed by Deng [30]. GET abolished the restriction on $m(\emptyset) = 0$; that is, $m(\emptyset)$ is not necessarily zero. In GET, the empty set (\emptyset) can be singleton subsets or multiple hypothesis sets. Moreover, GET can degenerate to the classical D-S theory if the value of $m(\emptyset)$ is zero. In other words, GET is the extension of the classical D-S theory and can express and deal with more uncertain information in the open world, comparing with D-S theory.

2.4.1. Generalized Basic Probability Assignment [30]. Suppose that U is a frame of discernment in an open world [30]. Its power set, 2_G^U , is composed of 2^U propositions, $\forall A \subset U$. If the function $m : 2_G^U \rightarrow [0, 1]$ meets the condition

$$\sum_{A \subset 2_G^U} m_G(A) = 1 \quad (12)$$

then m_G is the GBPA of the frame of discernment U . The difference between GBPA and traditional BPA is the restriction of \emptyset . Note that $m_G(\emptyset) = 0$ is not necessary in GBPA. If $m_G(\emptyset) = 0$, the GBPA reduces to a traditional BPA.

2.4.2. Generalized Combination Rule (GCR) [30]. In GET, $\emptyset_1 \cap \emptyset_2 = \emptyset$ means that the intersection between two empty sets is still an empty set. Given two GPBAs (m_1 and m_2), the GCR is defined as follows [30]:

$$m(A) = \frac{(1 - m(\emptyset)) \sum_{B \cap C = A} m_1(B) m_2(C)}{1 - K} \quad (13)$$

with

$$K = \sum_{B \cap C = \emptyset} m_1(B) m_2(C), \quad (14)$$

$$m(\emptyset) = m_1(\emptyset) m_2(\emptyset).$$

$m(\emptyset) = 1$ if and only if $K = 1$.

3. Proposed Method to Determine BPA Function

3.1. Sample Difference Degree. In order to reflect the difference between the sample and the target model, we proposed a sample difference degree function, which performs well in the measurement of difference between the sample data and

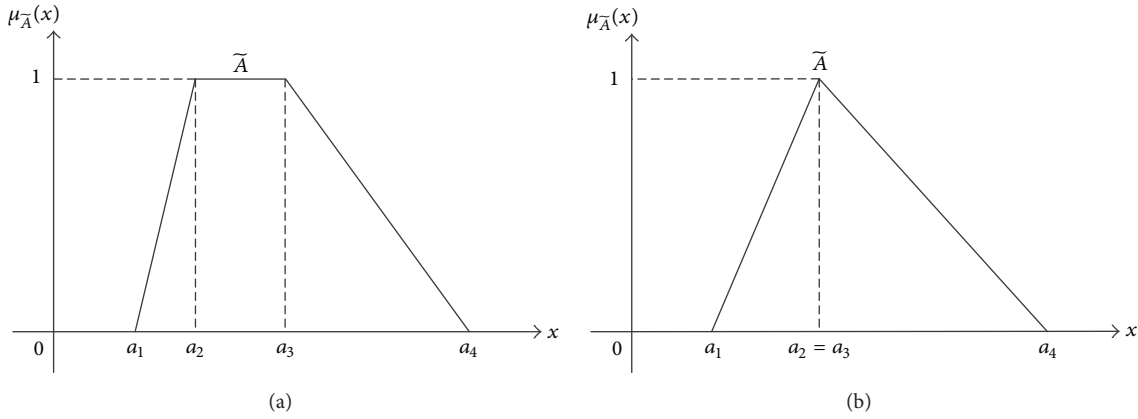


FIGURE 1: Trapezoidal and triangular fuzzy number.

the model data. And now, it will be amended to define pessimistic function in a more reasonable way.

In our method, the triangular fuzzy numbers, which denote the target model and the sample data, should firstly be normalized into the interval $[0, 1]$ as \tilde{A} and \tilde{a}_0 , respectively. Four normalized triangular fuzzy numbers are defined as follows:

- (1) Triangular fuzzy number of target model $\tilde{A} = (a, b, c; \omega)$.
- (2) Triangular fuzzy number of sample data $\tilde{a}_0 = (a_0, a_0, a_0; 1)$.
- (3) Left standard triangular fuzzy number $\tilde{A}_0 = (0, 0, 0; 1)$.
- (4) Right standard triangular fuzzy number $\tilde{A}_1 = (1, 1, 1; 1)$.

$\omega \in [0, 1]$, $0 \leq a \leq b \leq c \leq 1$. Then the Left and right average area, $S_L(\tilde{A})$, $S_L(\tilde{a}_0)$, $S_R(\tilde{A})$, and $S_R(\tilde{a}_0)$, are defined as follows and are shown in Figure 2. As can be seen from Figure 2, $g_1(x)$ is the left membership degree curve of \tilde{A} , and $g_1^{-1}(x)$ denotes inverse function of $g_1(x)$; $g_2(x)$ is the right membership degree curve of \tilde{A} , and $g_2^{-1}(x)$ denotes inverse function of $g_2(x)$. Then the left adjacent area S_{LA} is the area enclosed by $g_1(x)$ and left standard triangular fuzzy number $(0, 0, 0; 1)$; the left far area S_{LF} is the area enclosed by $g_2(x)$ and left standard triangular fuzzy number $(0, 0, 0; 1)$; the right adjacent area S_{RA} is the area enclosed by $g_2(x)$ and right standard triangular fuzzy number $(1, 1, 1; 1)$; the right far area S_{RF} is the area enclosed by $g_1(x)$ and right standard triangular fuzzy number $(1, 1, 1; 1)$. Obviously, the four kinds of area can be obtained by the following equations:

$$\begin{aligned}
 S_{LA} &= \int_0^\omega g_1^{-1}(x) dx, \\
 S_{LF} &= \int_0^\omega g_2^{-1}(x) dx, \\
 S_{RA} &= \int_0^\omega (1 - g_2^{-1}(x)) dx, \\
 S_{RF} &= \int_0^\omega (1 - g_1^{-1}(x)) dx.
 \end{aligned} \tag{15}$$

Based on the four kinds of area, S_{LA} , S_{LF} , S_{RA} , and S_{RF} , the left average area S_L and the right average area S_R are defined, respectively, as follows:

$$S_L = \frac{S_{LA} + S_{LF}}{2}, \tag{16}$$

$$S_R = \frac{S_{RA} + S_{RF}}{2}. \tag{17}$$

Figure 2 indicates that the larger S_L , the closer the fuzzy number \tilde{A} to \tilde{A}_1 ; the larger S_R , the closer the fuzzy number \tilde{A} to \tilde{A}_0 . That is to say, S_L and S_R can accurately represent the position and the state information of a triangular fuzzy number in the interval $[0, 1]$. Based on the fact that the shape and position of a fuzzy number can, to a large extent, be expressed as the credibility of the proposition, a conclusion can be made that the difference between the average area ($S_L(\tilde{A})$, $S_R(\tilde{A})$) of \tilde{A} and the average area ($S_L(\tilde{a}_0)$, $S_R(\tilde{a}_0)$) of \tilde{a}_0 reflects the difference between the sample \tilde{a}_0 and the model \tilde{A} . So, it is reasonable to define the difference degree *difa* to measure the difference between the sample and the model as follows:

$$difa = |S_L(\tilde{A}) - \omega S_L(\tilde{a}_0)| + |S_R(\tilde{A}) - \omega S_R(\tilde{a}_0)|. \tag{18}$$

3.2. Frame of Discernment in Open World. According to the basic framework of the generalized evidence theory [30], an open world is absolute and a closed world is relative. Assume the system to be concerned is not complete and the system FOD Θ is constructed as $\Theta = \{a_1, a_2, \dots, a_n, \emptyset\}$, where \emptyset denotes the unknown objects. \emptyset could be one unknown object or the conjunction of several unknown objects. During the procedure of applying GET, new target representation model is generated by machine learning method along with the accumulation of sensor reports to revise the existing target model library until the system is judged to be complete.

3.3. Pessimistic Function. As can be seen from (18), the *difa* reflects the possibility that the sample data may be distributed into the interval built by the training samples. The larger the *difa*, the greater the deviation between

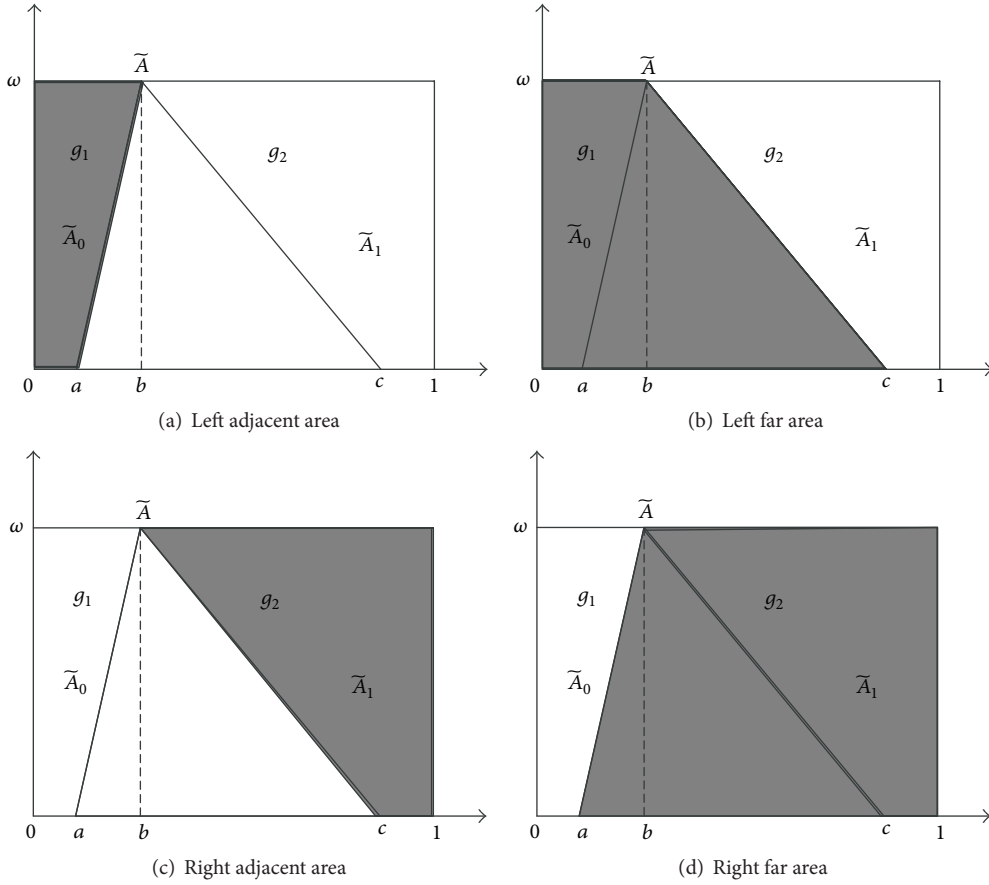


FIGURE 2: Left and right area between a triangle fuzzy number and standard fuzzy number.

the test sample and the training sample model, and vice versa. From this point of view, we use the *difa* as argument to define a pessimistic function and yield initial BPA, following a pessimistic strategy: when *difa* is greater than a threshold value, the incremental rate of BPA is less than the decreasing rate of *difa*; on the contrary, when *difa* is less than the threshold value, the incremental rate of BPA is greater than the decreasing rate of *difa*. In this way, pessimistic function can effectively reflect the difference between test sample and training sample models to generate the initial BPA.

Definition 4 (pessimistic function). Consider

$$S = \omega \cdot e^{-r \cdot (\text{difa}/2)}, \quad (19)$$

where ω denotes the height of sample triangular fuzzy number, r denotes the difference coefficient, which is used to revise the membership degree of unknown samples, and *difa* denotes the sample difference defined as (18).

3.4. Procedures to Determine BPA. A flow chart of the proposed method is shown in Figure 3 and details are as follows: Consider species $X = \{x_1, x_2, \dots, x_n, \emptyset\}$, where \emptyset denotes the unknown elements. Each species x_i has k

attributes $x_{i1}, x_{i2}, \dots, x_{ik}$, so the test sample ξ to be recognized also has k attributes $\xi_1, \xi_2, \dots, \xi_k$. We randomly choose m instances for each species x_i and build the model:

$$T^i = (t_1^i, t_2^i, \dots, t_m^i), \quad (20)$$

where T^i is a $k \times m$ matrix and the j th row $T^i(j, :)$ denotes j attribution value of each sample of species x_i .

3.4.1. Step 1: Establish the Triangular Fuzzy Number Model Matrix. A triangular fuzzy number model $\tilde{A}_{ij} = (a_{ij}, b_{ij}, c_{ij}; 1)$ for j attribution of species x_i is established according to the sample values, where

$$\begin{aligned} a_{ij} &= \min(T^i(j, :)), \\ b_{ij} &= \frac{\sum T^i(j, :)}{m}, \\ c_{ij} &= \max(T^i(j, :)). \end{aligned} \quad (21)$$

Then the triangular fuzzy number models for each attribution of species x_i can be represented as $M_i = (\tilde{A}_{i1}, \tilde{A}_{i2}, \dots, \tilde{A}_{ik})$.

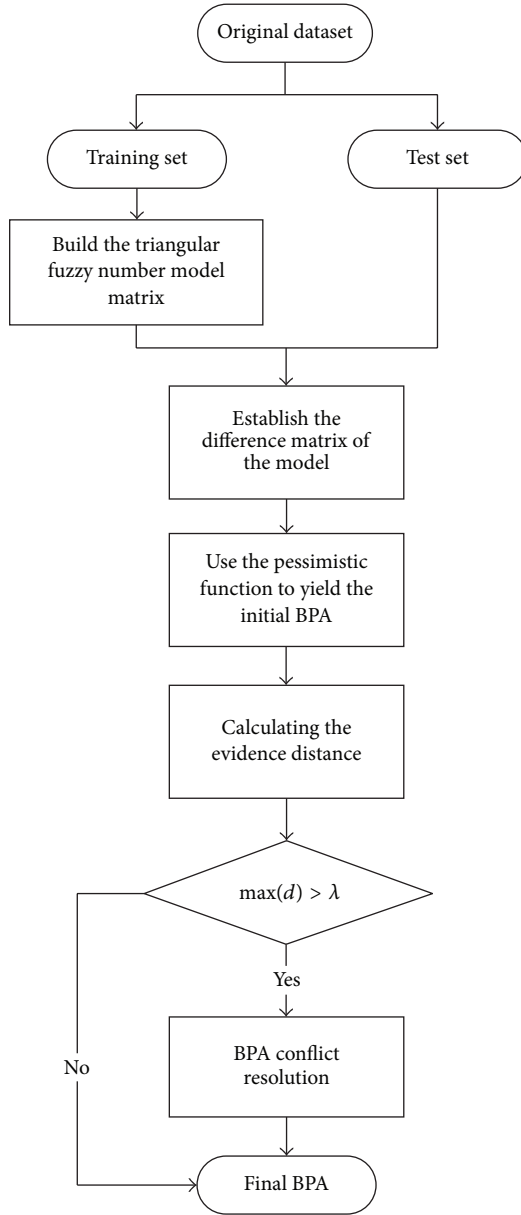


FIGURE 3: The steps to determine BPA.

Furthermore, all the triangular fuzzy number models for each attribution of species x_i ($i = 1, 2, \dots, n$) can be acquired and denoted as a $n \times k$ matrix $M = (M_1, M_2, \dots, M_n)'$, where each column of M represents the triangular fuzzy numbers belonging to the different species but the same attribution. As can be seen from Figure 4, there is often some intersection between two triangular fuzzy numbers. In most cases, the intersection is a generalized fuzzy number. But in some cases, it is not a generalized fuzzy number, as shown in Figure 4. For this particular case, it can be processed by the method proposed by Xiao et al. [46] to construct a generalized triangular fuzzy number. Besides, if there is no intersection between two fuzzy numbers, a special fuzzy number $(0, 0, 0; 0)$ could be used to represent

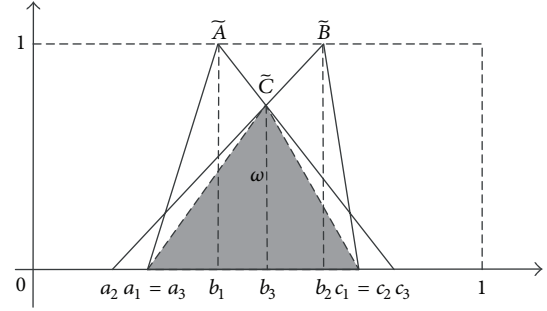


FIGURE 4: Generalized fuzzy number yield by two normal fuzzy numbers.

this case. To do the same operation on each column of M , we can get a fuzzy number matrix $M' = (M_1, M_2, \dots, M_{2^n-1})'$. The m th attribution value ξ_m ($1 \leq m \leq k$) of the test sample ξ to be recognized should be converted to a special triangular fuzzy number $\tilde{\xi}_m = (\xi_m, \xi_m, \xi_m; 1)$. Doing this processing k times, we can get a triangular fuzzy number matrix $\tilde{\xi} = (\tilde{\xi}_1, \tilde{\xi}_2, \dots, \tilde{\xi}_k)$. Then M' and $\tilde{\xi}$ can be merged into a matrix M_ξ . Now we need to normalize every triangular fuzzy number of the matrix M_ξ to $[0, 1]$. Each element of the matrix M_ξ is divided by n times of the maximum element value, where n represents normalization coefficient, which is adjustable according to different engineering application.

3.4.2. Step 2: Establish the Differentiation Matrix. In this step, we need to calculate the left and right average area (S_L, S_R) for each triangular fuzzy number in matrix M_ξ according to (16) and (17). Based on the area, the sample differentiation matrix $difa$ between each of the first $2^n - 1$ rows in M_ξ and the $(2^n - 1) \times k$ test sample ξ_0 could be calculated.

3.4.3. Step 3: Calculate the Similarity Matrix S . For each element in matrix $difa$, a pessimistic function can be defined according to (19):

$$S(i, j) = \omega_{ij} \cdot e^{-r \cdot (difa(i, j)/2)}; \quad (22)$$

now we could obtain a similarity matrix S_0 , which denotes the similarity degree between the test sample and the target species models. Assume FOD is incomplete; we first calculate the value of the j th ($j = 1, 2, \dots, k$) column of matrix S_0 . Then if the result is larger than 1, the j th column should be normalized and set $S(2^n, j) = 0$ which denotes the similarity degree between the test sample ξ_0 and \emptyset ; else if the result is less than 1, $S(2^n, j) = 1 - S(:, j)$ is regarded as the similarity between the test sample ξ_0 and \emptyset , where $S(:, j)$ represents the sum of the first $2^n - 1$ rows in similarity matrix S_0 . Finally, the modified $2^n \times k$ similarity matrix S_0 could be regarded as the initial BPA matrix $S_{2^n \times k}$.

3.4.4. Step 4: Calculate Evidence Distance. Taking each column $S(:, j)$ ($j = 1, 2, \dots, k$) in matrix $S_{2^n \times k}$ as a piece of evidence or as a vector, we could calculate the distance

between any two bodies of evidence based on (8) to obtain a distance matrix $[D_{ij}]_{k \times k}$, where D_{ij} represents the distance between the i th and j th column of the initial BPA matrix $S_{2^n \times k}$. Obviously, D_{ij} equals D_{ji} ; that is, distance matrix $[D_{ij}]_{k \times k}$ is a symmetrical matrix, which can simplify the calculation to a great extent.

3.4.5. Step 5: Determine the Extent of Conflict of Evidence.

First, we can sum each row $D(j, :)$ ($j = 1, 2, \dots, k$) in matrix $D_{k \times k}$ and obtain a vector d_0 ; it can be normalized as a vector $d = d_0 / \max(d_0)$, where $\max(d_0)$ denotes the maximum element of d_0 . As debated in Section 1, classical D-S theory has a vital problem that combining two highly conflicting evidences could get even a wrong result. So a conflict threshold λ ($0 \leq \lambda \leq 1$) should be set first according to concrete engineering application. If the maximum element value of vector d is less than conflict threshold λ , the conflict degree is acceptable and the initial BPAs could be combined with GCR (see (12) and (13)) directly; otherwise the initial BPAs should be adjusted. The specific process is in Section 3.4.6, with the discount coefficient method.

3.4.6. Step 6: Adjust the Conflicting Evidence by Using the Discount Coefficient Method. This step will be operated if evidence conflict is over threshold λ . In this step, conflict resolution between initial BPAs should firstly be done. Then GCR could be applied to combine the preprocessed BPAs. Computational procedure is summarized as follows:

- (1) Construct a comparative matrix P with the average of each row (\bar{d}_i) in distance matrix $D_{k \times k}$:

$$P(i, j) = \frac{\bar{d}_i}{\bar{d}_j}, \quad (23)$$

where $P(i, j)$ is the i th row and j th column element in matrix P , \bar{d}_i is the i th row of $D_{k \times k}$, and \bar{d}_j is the j th row of $D_{k \times k}$.

- (2) Calculate eigenvector (ω_i) corresponding to the maximum eigenvalue and normalize ω_i as discount coefficient ω .
- (3) Discount the initial BPAs with coefficient ω to obtain the final BPAs.
- (4) Combine the final BPAs with GCR and then the final recognized result will be acquired.

4. Experimental Analysis

In this paper, we take Iris dataset [47] to demonstrate the effectiveness of the proposed method. The Iris dataset, which is perhaps the best known database to be found in pattern recognition literature, involves classification of three species of the Iris flowers, *Iris setosa* (S), *Iris versicolour* (E), and *Iris virginica* (V), on the basis of four numeric attributes of the Iris flower: sepal length (SL) in cm, sepal width (SW) in cm, petal length (PL) in cm, and petal width (PW) in cm [47].

In the Iris dataset, there are 50 instances for each of three species. The data are obtained from the UCI repository (UCI Machine Learning Repository: <http://archive.ics.uci.edu/ml/datasets/Iris>) of machine learning databases. Among 50 instances of each species, 30 instances are randomly selected as the training set, and the remaining 20 instances serve as the test set. Each of four attributes is regarded as an information source, and correspondingly there are three training sets and three test sets.

4.1. To Recognize Known Species of Iris Dataset. Before the experiment, the fully optimized parameter settings should be obtained first by training the dataset. The conflict threshold λ can be set according to experts' experience usually. In this experiment, $\lambda = 0.2$. It means that the conflict is in the acceptable scale only if the maximum conflict between bodies of evidence is less than 0.2. Otherwise the initial BPAs should be adjusted. The difference coefficient r can be obtained by an iterative program. The difference coefficient r would be adjusted in each iterative step. When the average recognition rate of known species is up to an acceptable scale and the deviation of the average recognition rate between current step and last step was tiny enough, the parameter r would be obtained. After the training process, we have obtained the optimized parameter settings as $\{r = 12.0, \lambda = 0.2\}$. Because these optimized parameters fit the dataset well, parameters are all set as the optimized settings in the following.

Following the steps in Section 3.4, triangular fuzzy numbers of the training samples are built, shown in Table 1 and in Figure 5.

According to Step 5 in Section 3.4.5, we can obtain BPAs for each attribute of a *Setosa* instance (5.1, 3.5, 1.4, 0.2) shown in Table 2.

After conflict resolution (if needed), these BPAs in Table 2 could be combined with GCR. The results are as follows:

$$\begin{aligned} m(S) &= 0.9655, \\ m(E) &= 0.0344, \\ m(V) &= 0.0002, \\ m(SE) &= 0, \\ m(SV) &= 0, \\ m(EV) &= 0, \\ m(SEV) &= 0, \\ m(\emptyset) &= 0. \end{aligned} \quad (24)$$

As can be seen, the combination result illustrates that the test instance (5.1, 3.5, 1.4, 0.2) belongs to species *setosa*, which is consistent with the actual situation.

4.2. To Recognize Unknown Species of Iris Dataset. In order to check the proposed method's ability to recognize the unknown species of data in an open world, triangular fuzzy number models are constructed only by two species of Iris dataset selected randomly from the three species this time.

TABLE 1: Fuzzy number model constructed by 90 Iris data instances.

Item	Attributes	
	SL	SW
S	(4.4, 4.94, 5.8; 1.0)	(3.0, 3.47, 4.0; 1.0)
E	(4.9, 5.48, 6.2; 1.0)	(2.0, 2.65, 3.4; 1.0)
V	(6.1, 6.85, 7.7; 1.0)	(2.6, 2.94, 3.2; 1.0)
SE	(4.9, 5.2625, 5.8; 0.625)	(2.0, 2.65, 3.4; 1.0)
SV	(0, 0, 0; 0)	(2.6, 2.94, 3.2; 1.0)
EV	(6.1, 6.151, 6.20; 0.068)	(2.6, 2.849, 3.2; 0.733)
SEV	(0, 0, 0; 0)	(2.6, 2.849, 3.2; 0.733)

Item	Attributes	
	PL	PW
S	(1.0, 1.43, 1.9; 1.0)	(0.1, 0.24, 0.6; 1.0)
E	(3.3, 3.9, 4.5; 1.0)	(1.0, 1.2, 1.6; 1.0)
V	(4.8, 5.71, 6.7; 1.0)	(1.4, 1.99, 2.3; 1.0)
SE	(0, 0, 0; 0)	(0, 0, 0; 0)
SV	(0, 0, 0; 0)	(0, 0, 0; 0)
EV	(0, 0, 0; 0)	(1.4, 1.519, 1.6; 0.202)
SEV	(0, 0, 0; 0)	(0, 0, 0; 0)

Thus, the remaining species of Iris dataset could be regarded as test sets, which are unknown to the recognition system. For example, we randomly selected 30 instances from *Setosa* (S) and *Versicolour* (E), respectively, as training samples to construct triangular fuzzy number model and the remaining 20 instances in each species as test samples. And parameters should also be set as the optimized settings obtained in the training process. The training samples' triangular fuzzy number model is shown in Table 3.

For a known *Setosa* instance (4.9, 3.1, 1.4, 0.2), we can obtain BPAs for each attribute in Table 4, according to Step 5 in Section 3.4.5.

After conflict resolution (if needed), these BPAs in Table 4 should be combined with GCR. The results are as follows:

$$\begin{aligned}
 m(S) &= 0.9629, \\
 m(E) &= 0.0371, \\
 m(SE) &= 0, \\
 m(\emptyset) &= 0.
 \end{aligned} \tag{25}$$

As can be seen, the combination result illustrates that the test instance (4.9, 3.1, 1.4, 0.2) belongs to species *setosa*.

Similarly, to recognize an instance from the open world, we can obtain BPAs for each attribute of a *Virginica* (V) instance (unknown species) (6.3, 3.3, 6.0, 2.5) in Table 5. As can be seen from Table 5, the difference function $difa$ (see (18)) and pessimistic function (see (19)) indeed recognized this instance as a member of species V.

After conflict resolution (if needed), these BPAs in Table 5 should be combined again with GCR. The results are as follows:

$$\begin{aligned}
 m(S) &= 0.0009, \\
 m(E) &= 0.3955, \\
 m(SE) &= 0, \\
 m(\emptyset) &= 0.6036.
 \end{aligned} \tag{26}$$

TABLE 2: BPAs for each attribute of a *Setosa* instance.

Item	Attributes			
	SL	SW	PL	PW
S	0.3232	0.1828	0.8279	0.8321
E	0.2790	0.1369	0.1536	0.1595
V	0.2104	0.1579	0.0185	0.0084
SE	0.0717	0.1369	0	0
SV	0	0.1579	0	0
EV	0.1157	0.1138	0	0
SEV	0	0.1138	0	0
\emptyset	0	0	0	0

TABLE 3: Fuzzy number model constructed by 60 Iris data instances.

Item	Attributes	
	SL	SW
S	(4.7, 5.19, 5.7; 1.0)	(3.1, 3.62, 4.4; 1.0)
E	(5.1, 5.97, 6.7; 1.0)	(2.5, 2.84, 3.1; 1.0)
SE	(5.1, 5.478, 5.7; 0.4383)	(0, 0, 0; 0)

Item	Attributes	
	PL	PW
S	(1.3, 1.5, 1.7; 1.0)	(0.2, 0.28, 0.4; 1.0)
E	(3.0, 4.23, 5.0; 1.0)	(1.1, 1.36, 1.7; 1.0)
SE	(0, 0, 0; 0)	(0, 0, 0; 0)

TABLE 4: BPAs for each attribute of a *Setosa* instance.

Item	Attributes			
	SL	SW	PL	PW
S	0.4448	0.3888	0.8036	0.8256
E	0.3352	0.3888	0.1964	0.1559
SE	0.2200	0.2224	0	0
\emptyset (V)	0	0	0	0.0185

And the recognition result suggests that the test instance (6.3, 3.3, 6.0, 2.5) is an unknown species of Iris datasets.

In order to further illustrate the validity and accuracy of the proposed method, the same experiments have been done 100 times. And to be satisfied, the average recognition rate about known-species Iris dataset (in closed world) is up to 81.55% and the average recognition rate about unknown-species Iris dataset (in open world) is up to 73.40%. Several unrecognized data in one experiment are shown in Table 6. As can be seen from Table 6, two instances of species E were recognized as species V, and two datasets of species V were recognized as species E and \emptyset (that is, species S), respectively. And two instances of species \emptyset (species S) were recognized as species E.

It is not difficult to explain the wrong results. From Table 1 and Figure 5 we can see that only attributes PL and PW of species *setosa* are totally separated from species E and species V, and the two remaining attributes are all intersected with others. At the same time, species E and species V intersect with each other for each of attributes. Especially since the length of interval overlapping in the SL and SW attribute

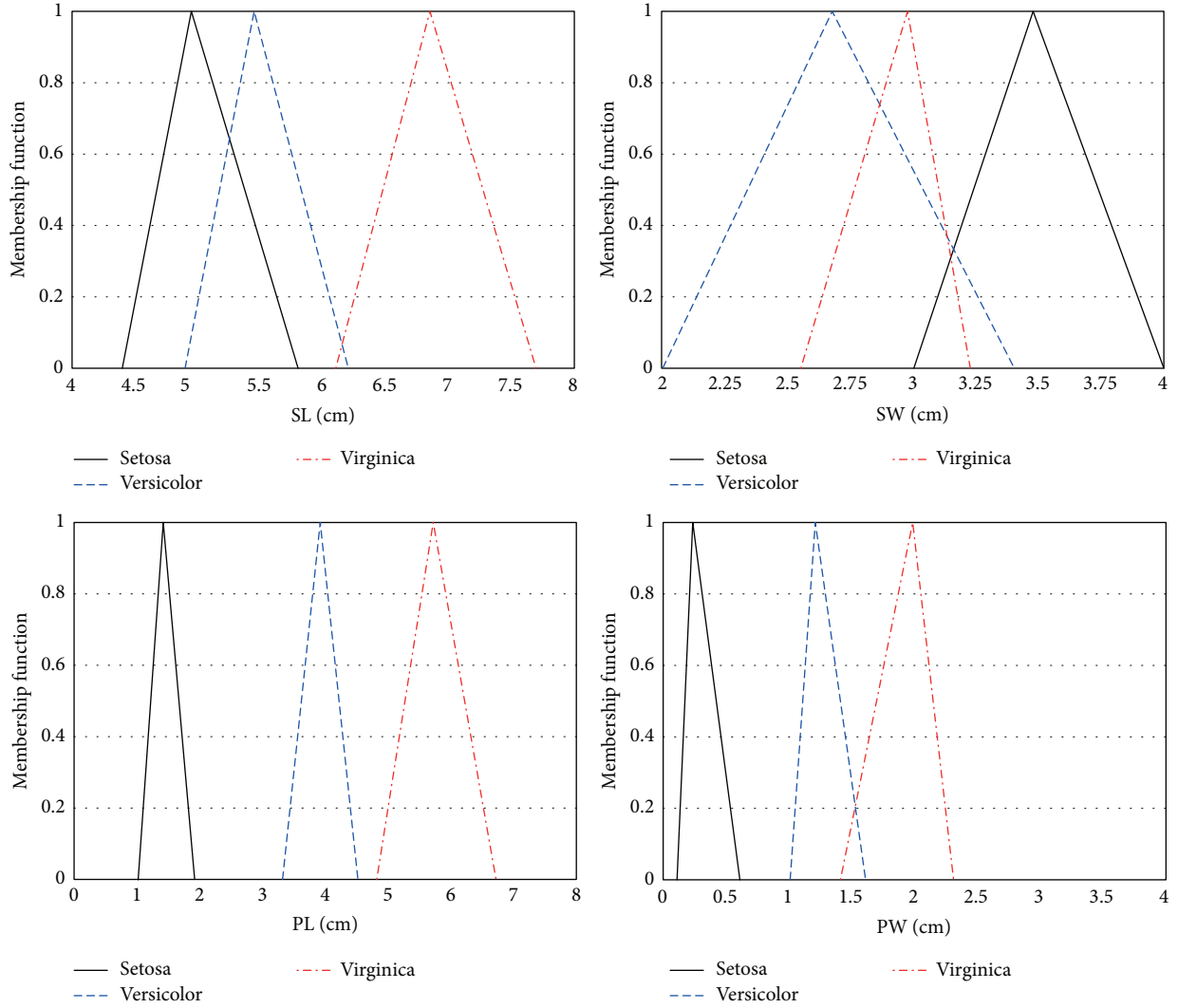


FIGURE 5: The fuzzy number representation of each attribute of each species.

TABLE 5: BPAs for each attribute of a Virginica instance.

Item	Attributes			
	SL	SW	PL	PW
S	0.0118	0.2214	0	0
E	0.0434	0.0188	0.0011	0.0001
SE	0.0313	0.0983	0	0
\emptyset (V)	0.9136	0.6615	0.9989	0.9999

TABLE 6: Unrecognized instances in an experiment.

SL	SW	PL	PW	Sample species	Recognition result
5.9	3.2	4.8	1.8	E	V
6.7	3.1	4.7	1.5	E	V
7.7	2.6	6.9	2.3	V	\emptyset
6.1	2.6	5.6	1.4	V	E
5.0	3.0	1.6	0.2	\emptyset	E
5.4	3.4	1.5	0.4	\emptyset	E

is large, all the attributes of some data intersect with each other and it is difficult to distinguish these attributes using pessimistic function. Even so, the simulation examples show that the average recognition rate for the instances of the known species is up to 81.55% and the average recognition rate of the unknown-species instances is up to 73.40%. Moreover, during the process of recognition, the number of training samples is small (only 30 instances in each species), and test samples are totally separated from training samples.

It scientifically proves that the proposed method to determine BPA has great effectiveness and could work well with GET.

5. Conclusion

In the application of data fusion, the generalized evidence theory (GET) has more advantages than the classical Dempster-Shafer evidence theory due to its ability to deal with evidence conflict when the frame of discernment is

incomplete. How to determine the generalized basic probability assignment (GBPA) in an open world is still an open issue. A method to construct GBPA is proposed in this paper. This method uses training samples to build triangular fuzzy number models for each attribute of the multiattribute dataset. Then, the differentiation function and the similarity function are defined. The initial GBPAs are generated by the similarity function, and bodies of evidence are fused with Dempster's rule or the generalized combination rule (GCR) according to the actual target environment. This method makes full use of the advantages of GET to deal with these targets in the open world. In order to reduce the impact of conflicting evidence on the fusion results, the distance between each body of evidence is calculated and conflict resolution is to be done in the initial stage of determining GBPAs to eliminate human interference and environment noise. Several numerical examples show that the method is concise and effective, and this method has a very significant data processing capacity of small samples based on a good theoretical foundation. The proposed method to obtain GBPA can effectively overcome the problem of subjectivity, which has strong generality. The classification of Iris data is used to illustrate the efficiency and the low computational complexity of the proposed method. This method will help to promote GET and use GCR effectively.

Competing Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The work is partially supported by National Natural Science Foundation of China (Grant nos. 60904099 and 61501377) and Natural Science Basic Research Plan in Shaanxi Province of China (Program no. 2016JM6018).

References

- [1] A. P. Dempster, "Upper and lower probabilities induced by a multivalued mapping," *Annals of Mathematical Statistics*, vol. 38, pp. 325–339, 1967.
- [2] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, New Jersey, NJ, USA, 1976.
- [3] C. Mellish and X. Sun, "The semantic web as a linguistic resource: opportunities for natural language generation," *Knowledge-Based Systems*, vol. 19, no. 5, pp. 298–303, 2006.
- [4] Y.-Z. Liu, Y.-C. Jiang, X. Liu, and S.-L. Yang, "CSMC: a combination strategy for multi-class classification based on multiple association rules," *Knowledge-Based Systems*, vol. 21, no. 8, pp. 786–793, 2008.
- [5] E. Bagheri, R. Zafarani, and M. Barouni-Ebrahimi, "Can reputation migrate? On the propagation of reputation in multi-context communities," *Knowledge-Based Systems*, vol. 22, no. 6, pp. 410–420, 2009.
- [6] X. Zhao, R. Wang, H. Gu, G. Song, and Y. L. Mo, "Innovative data fusion enabled structural health monitoring approach," *Mathematical Problems in Engineering*, vol. 2014, Article ID 369540, 11 pages, 2014.
- [7] C. Fu and K.-S. Chin, "Robust evidential reasoning approach with unknown attribute weights," *Knowledge-Based Systems*, vol. 59, pp. 9–20, 2014.
- [8] C. Fu and S. Yang, "An evidential reasoning based consensus model for multiple attribute group decision analysis problems with interval-valued group consensus requirements," *European Journal of Operational Research*, vol. 223, no. 1, pp. 167–176, 2012.
- [9] X. Su, S. Mahadevan, P. Xu, and Y. Deng, "Dependence assessment in human reliability analysis using evidence theory and AHP," *Risk Analysis*, vol. 35, no. 7, pp. 1296–1316, 2015.
- [10] H.-C. Liu, J.-X. You, X.-J. Fan, and Q.-L. Lin, "Failure mode and effects analysis using D numbers and grey relational projection method," *Expert Systems with Applications*, vol. 41, no. 10, pp. 4670–4679, 2014.
- [11] W. Jiang, B. Wei, C. Xie, and D. Zhou, "An evidential sensor fusion method in fault diagnosis," *Advances in Mechanical Engineering*, vol. 8, no. 3, pp. 1–7, 2016.
- [12] W. Jiang, C. Xie, B. Wei, and D. Zhou, "A modified method for risk evaluation in failure modes and effects analysis of aircraft turbine rotor blades," *Advances in Mechanical Engineering*, vol. 8, no. 4, pp. 1–16, 2016.
- [13] F. Lolli, A. Ishizaka, R. Gamberini, B. Rimini, and M. Messori, "FlowSort-GDSS—a novel group multi-criteria decision support system for sorting problems with application to FMEA," *Expert Systems with Applications*, vol. 42, no. 17–18, pp. 6342–6349, 2015.
- [14] N. Rikhtegar, N. Mansouri, A. A. Oroumieh, A. Yazdani-Chamzini, E. K. Zavadskas, and S. Kildienė, "Environmental impact assessment based on group decision-making methods in mining projects," *Economic Research*, vol. 27, no. 1, pp. 378–392, 2014.
- [15] D. Niu, Y. Wei, Y. Shi, and H. R. Karimi, "A novel evaluation model for hybrid power system based on vague set and Dempster-Shafer evidence theory," *Mathematical Problems in Engineering*, vol. 2012, Article ID 784389, 12 pages, 2012.
- [16] R. R. Yager, "On the fusion of non-independent belief structures," *International Journal of General Systems*, vol. 38, no. 5, pp. 505–531, 2009.
- [17] X. Su, S. Mahadevan, W. Han, and Y. Deng, "Combining dependent bodies of evidence," *Applied Intelligence*, vol. 44, pp. 634–644, 2016.
- [18] X. Su, S. Mahadevan, P. Xu, and Y. Deng, "Handling of dependence in Dempster-Shafer theory," *International Journal of Intelligent Systems*, vol. 30, no. 4, pp. 441–467, 2015.
- [19] P. D. Xu, X. Y. Su, S. Mahadevan, C. Z. Li, and Y. Deng, "A non-parametric method to determine basic probability assignment for classification problems," *Applied Intelligence*, vol. 41, no. 3, pp. 681–693, 2014.
- [20] D. Suh and J. Yook, "A method to determine basic probability assignment in context awareness of a moving object," *International Journal of Distributed Sensor Networks*, vol. 2013, Article ID 972641, 7 pages, 2013.
- [21] S. Yoon, C.-K. Ryu, and D. Suh, "A novel way of BPA calculation for context inference using sensor signals," *International Journal of Smart Home*, vol. 8, no. 1, pp. 1–8, 2014.
- [22] W. Jiang, Y. Luo, X. Qin, and J. Zhan, "An improved method to rank generalized fuzzy numbers with different left heights and right heights," *Journal of Intelligent and Fuzzy Systems*, vol. 28, no. 5, pp. 2343–2355, 2015.

- [23] W. Jiang, Y. Yang, Y. Luo, and X. Qin, "Determining basic probability assignment based on the improved similarity measures of generalized fuzzy numbers," *International Journal of Computers, Communications & Control*, vol. 10, no. 3, pp. 333–347, 2015.
- [24] 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, 1986.
- [25] C. Yu, J. Yang, D. Yang, X. Ma, and H. Min, "An improved conflicting evidence combination approach based on a new supporting probability distance," *Expert Systems with Applications*, vol. 42, no. 12, pp. 5139–5149, 2015.
- [26] W. Jiang, B. Wei, X. Qin, J. Zhan, and Y. Tang, "Sensor data fusion based on a new conflict measure," *Mathematical Problems in Engineering*, Article ID 5769061, 2016.
- [27] R. R. Yager, "On the Dempster-Shafer framework and new combination rules," *Information Sciences*, vol. 41, no. 2, pp. 93–137, 1987.
- [28] D. Dubois and H. Prade, "A survey of belief revision and updating rules in various uncertainty models," *International Journal of Intelligent Systems*, vol. 9, no. 1, pp. 61–100, 1994.
- [29] P. Smets and R. Kennes, "The transferable belief model," *Artificial Intelligence*, vol. 66, no. 2, pp. 191–234, 1994.
- [30] Y. Deng, "Generalized evidence theory," *Applied Intelligence*, vol. 43, no. 3, pp. 530–543, 2015.
- [31] Z. Xu and M. Xia, "Induced generalized intuitionistic fuzzy operators," *Knowledge-Based Systems*, vol. 24, no. 2, pp. 197–209, 2011.
- [32] Y. Deng, S. Mahadevan, and D. Zhou, "Vulnerability assessment of physical protection systems: a bio-inspired approach," *International Journal of Unconventional Computing*, vol. 11, pp. 227–243, 2015.
- [33] J.-B. Yang and D.-L. Xu, "Evidential reasoning rule for evidence combination," *Artificial Intelligence*, vol. 205, pp. 1–29, 2013.
- [34] Y. Yang and D. Han, "A new distance-based total uncertainty measure in the theory of belief functions," *Knowledge-Based Systems*, vol. 94, pp. 114–123, 2016.
- [35] R. R. Yager and N. Alajlan, "Decision making with ordinal payoffs under Dempster-Shafer type uncertainty," *International Journal of Intelligent Systems*, vol. 28, no. 11, pp. 1039–1053, 2013.
- [36] A.-L. Jousselme, D. Grenier, and É. Bossé, "A new distance between two bodies of evidence," *Information Fusion*, vol. 2, no. 2, pp. 91–101, 2001.
- [37] L. A. Zadeh, "Fuzzy sets," *Information and Computation*, vol. 8, pp. 338–353, 1965.
- [38] G. Fan, D. Zhong, F. Yan, and P. Yue, "A hybrid fuzzy evaluation method for curtain grouting efficiency assessment based on an AHP method extended by D numbers," *Expert Systems with Applications*, vol. 44, pp. 289–303, 2016.
- [39] Y. Deng, "A threat assessment model under uncertain environment," *Mathematical Problems in Engineering*, vol. 2015, Article ID 878024, 12 pages, 2015.
- [40] Y. Deng, "Fuzzy analytical hierarchy process based on canonical representation on fuzzy numbers," *Journal of Computational Analysis and Applications*, In press.
- [41] S.-B. Tsai, M.-F. Chien, Y. Xue et al., "Using the fuzzy DEMATEL to determine environmental performance: a case of printed circuit board industry in Taiwan," *PLoS ONE*, vol. 10, no. 6, Article ID e0129153, 2015.
- [42] A. Mardani, A. Jusoh, and E. K. Zavadskas, "Fuzzy multiple criteria decision-making techniques and applications—two decades review from 1994 to 2014," *Expert Systems with Applications*, vol. 42, no. 8, pp. 4126–4148, 2015.
- [43] C. Changdar, G. S. Mahapatra, and R. K. Pal, "An ant colony optimization approach for binary knapsack problem under fuzziness," *Applied Mathematics and Computation*, vol. 223, pp. 243–253, 2013.
- [44] Y. Deng, Y. Liu, and D. Zhou, "An improved genetic algorithm with initial population strategy for symmetric TSP," *Mathematical Problems in Engineering*, vol. 2015, Article ID 212794, 6 pages, 2015.
- [45] D. Dubois and H. Prade, *Fuzzy Sets and Systems: Theory and Applications*, Academic Press, New York, NY, USA, 1980.
- [46] J. Y. Xiao, M. M. Tong, and C. J. Zhu, "Basic probability assignment construction method based on generalized triangular fuzzy number," *Journal of Scientific Instrument*, vol. 33, pp. 430–434, 2012.
- [47] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of Human Genetics*, vol. 7, no. 2, pp. 179–188, 1936.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

