INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 10(3):333-347, June, 2015.

# Determining Basic Probability Assignment Based on the Improved Similarity Measures of Generalized Fuzzy Numbers

W. Jiang, Y. Yang, Y. Luo, X.Y. Qin

**Wen Jiang*, Yan Yang, Yu Luo, Xiyun Qin**
Northwestern Polytechnical University
School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710072, China
jiangwen@nwpu.edu.cn, yangyan7003@nwpu.edu.cn
345255046@qq.com, 945766782@qq.com
*Corresponding author

**Abstract:** Dempster-Shafer theory of evidence has been widely used in many data fusion application systems. However, how to determine basic probability assignment, which is the main and the first step in evidence theory, is still an open issue. In this paper, an improved method to determine the similarity measure between generalized fuzzy numbers is presented. The proposed method can overcome the drawbacks of the existing similarity measures. Then, we propose a new method for obtaining basic probability assignment (BPA) based on the proposed similarity measure method between generalized fuzzy numbers. Finally, the efficiency of the proposed method is illustrated by the classification of Iris data.
**Keywords:** data fusion, dempster-Shafer evidence theory, basic probability assignment (BPA), generalized fuzzy numbers, similarity measures

# 1 Introduction

Dempster-Shafer (D-S) theory of evidence is widely used in many fields of information fusion due to its efficiency in dealing with uncertain information. In real data fusion application systems based on DS theory, the basic probability assignment function should be given so that the combined BPA can be obtained through Dempster's rule of combination [1, 2]. However, how to determine basic probability assignment, which is the main and the first step in evidence theory, is still an open issue. A number of authors have addressed this problem using different approaches. Zhu et al. proposed a method to derive mass values from fuzzy membership degrees. For this purpose, fuzzy c-means (FCM) clustering is used to represent the grey levels as fuzzy sets [3]. Bendjebbour et al. [4] proposed a probabilistic model where the frame of discernment contained the individual clusters and the ignorance that was modeled by the union of all individual clusters. In that work, the authors derived the mass value of ignorance from the mixture of distributions of the individual clusters composing it. Guan et al. [5] came up with three methods to construct the BPA function based on gray correlation analysis, fuzzy sets, and attribute measure respectively. Chen et al. [6] and Li et al. [7] used back-propagation (BP) neural network to obtain basic probability assignment. According as neural network can gain stronger generalization ability, the measured data being processed by neural network can be used as the BPA value of every sensor. Xu et al. [8] used the difference matrix of deviation degree to represent quantitatively the degree of similarity between interval numbers, and constructed an expression of basic probability assignment function. Since the basic probability assignment of evidence theory obtained by using neural network has high computational complexity, Zuo et al. [9] put forward a method of rough set theory based on random set and BP neural network to obtain the basic probability assignment. In the framework of random set, the ability of attribute reduction of rough set was made use of to reduce the neural network input dimension. In papers [10-13], in order to solve the different practical problems, we proposed several different approaches to obtain BPA. These more pragmatic methods are proposed to generate BPAs from uncertain information.

In this paper, an improved method to determine the similarity measure between generalized fuzzy numbers is presented. Then, a new method to obtain basic probability assignment (BPA) is proposed based on the improved similarity measure between generalized fuzzy numbers. An experiment of Iris data classification is used to illustrate the efficiency of our method.

## 2 Preliminaries

### 2.1 Dempster Shafer Evidence Theory.

Evidence theory first supposes the definition of a set of hypotheses $\Theta = \{H_1, H_2, \cdots, H_N\}$ called the frame of discernment. It is composed of $N$ exhaustive and exclusive hypotheses. Let us denote $P(\Theta)$, the power set composed with the $2^N$ propositions A of $\Theta$:

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

Where $\emptyset$ denotes the empty set. The $N$ subsets containing only one element are called singletons. A key point of evidence theory is the basic probability assignment (BPA). A BPA is a function from $P(\Theta)$ to [0, 1] ,and which satisfies the following conditions:

$$\sum_{A \in P(\Theta)} m(A) = 1; m(\emptyset) = 0, \tag{2}$$

Dempster's rule of combination (also called orthogonal sum), noted by $m = m_1 \oplus m_2$, is the first one within the framework of evidence theory which can combine two BPAs $m_1$ and $m_2$ to yield a new BPA:

$$m(A) = \frac{\sum_{B \cap C = A} m_1(B) m_2(C)}{1 - k} \quad \text{and} \quad k = \sum\nolimits_{B \cap C = \emptyset} m_1(B) m_2(C) \tag{3}$$

Where $k$ is a normalization constant, called conflict because it measures the degree of conflict between $m_1$ and $m_2$. $k = 0$ corresponds to the absence of conflict between $m_1$ and $m_2$, whereas $k = 1$ implies complete contradiction between $m_1$ and $m_2$. The belief function resulting from the combination of $J$ information sources $S_J$ defined as

$$m = m_1 \oplus m_2 \cdots \oplus m_j \cdots\cdots \oplus m_J, \tag{4}$$

As can be seen from above, multi source information can be easily fused in the framework of evidence theory, if we can obtain the BPA functions.

### 2.2 Generalized Fuzzy numbers

A generalized fuzzy number $\widetilde{A} = (a, b, c, d; w)$ is described as any fuzzy subset of the real line $R$ with membership function $\mu_{\widetilde{A}}$ which has the following properties [14]:

(1) $\mu_{\widetilde{A}}$ is a continuous mapping from $R$ to the closed interval in $[0, w]$, $0 \leq w \leq 1$;
(2) $\mu_{\widetilde{A}}(x) = 0$ for all $x \in (-\infty, a]$;
(3) $\mu_{\widetilde{A}}(x)$ is strictly increasing in $[a, b]$;
(4) $\mu_{\widetilde{A}}(x) = w$ for all $x \in [b, c]$, where $w$ is a constant and $0 < w \leq 1$;
(5) $\mu_{\widetilde{A}}(x)$ is strictly decreasing in $[c, d]$;
(6) $\mu_{\widetilde{A}}(x) = 0$ for all $x \in [d, +\infty)$.

Where $0 < w \leq 1$, $a, b, c$ and $d$ are real numbers. Especially, a generalized trapezoidal fuzzy number can be defined as $\widetilde{A} = (a, b, c, d; w)$, where $a \leq b \leq c \leq d$, $0 \leq w \leq 1$, its membership function is defined by

$$\mu_{\widetilde{A}}(x) = \begin{cases} \dfrac{(x-a)}{b-a} & a \leq x \leq b \\ w & b \leq x \leq c \\ \dfrac{(x-c)}{d-c} & c \leq x \leq d \\ 0 & \text{else} \end{cases} \tag{5}$$

If $w = 1$, then the generalized fuzzy number $\widetilde{A}$ is called a normal trapezoidal fuzzy number, denote as $\widetilde{A} = (a, b, c, d)$. If $a = b$ and $c = d$, then $\widetilde{A}$ is called a crisp interval. If $b = c$, then $\widetilde{A}$ is called a generalized triangular fuzzy number. If $a = b = c = d$, then $\widetilde{A}$ is called a real number.

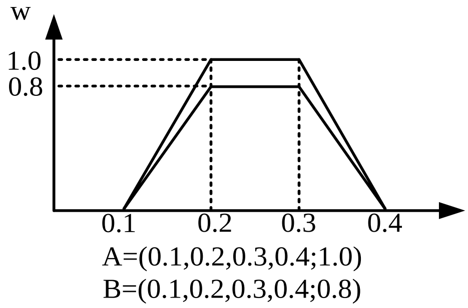


Figure 1: Two generalized trapezoidal fuzzy numbers $\widetilde{A}$ and $\widetilde{B}$

Figure 1 shows two different generalized trapezoidal fuzzy numbers $\widetilde{A} = (0.1, 0.2, 0.3, 0.4; 1.0)$ and$\widetilde{B} = (0.1, 0.2, 0.3, 0.4; 0.8)$. Compared with normal fuzzy numbers, the generalized fuzzy numbers can deal with uncertain information in a more flexible manner. For example, in decision making situation, the values $w_1$ and $w_2$ represent the degree of confidence of the opinions of the decision-makers' $\widetilde{A}$ and$\widetilde{B}$, respectively, where $w_1 = 1$ and $w_2 = 0.8$.

## 2.3 A Review of the Existing Similarity Measures between Fuzzy Numbers

In this section, we briefly introduce some existing similarity measures between fuzzy numbers from Chen [15], Lee [16], Chen and Chen [17], Wei & Chen [18] and Hejazi, et al. [19].

Let $\widetilde{A}$ and $\widetilde{B}$ be two trapezoidal fuzzy numbers, where $\widetilde{A} = (a_1, a_2, a_3, a_4)$ and $\widetilde{B} = (b_1, b_2, b_3, b_4)$, as shown in Figure 2. Chen [15] presented a similarity measure between fuzzy numbers $\widetilde{A}$ and $\widetilde{B}$ based on the geometric distance, where the degree of similarity $S(\widetilde{A}, \widetilde{B})$ between the fuzzy numbers $\widetilde{A}$ and $\widetilde{B}$ is calculated as follows:

$$S(\widetilde{A}, \widetilde{B}) = 1 - \frac{\sum_{i=1}^{4} |a_i - b_i|}{4}. \tag{6}$$

Where $S(\widetilde{A}, \widetilde{B}) \in [0, 1]$. The larger the value of $S(\widetilde{A}, \widetilde{B})$, the more the similarity between the fuzzy numbers $\widetilde{A}$ and $\widetilde{B}$.

If $\widetilde{A}$ and $\widetilde{B}$ are two triangular fuzzy numbers, where $\widetilde{A} = (a_1, a_2, a_3)$ and $\widetilde{B} = (b_1, b_2, b_3)$. The degree of similarity $S(\widetilde{A}, \widetilde{B})$ between the triangular fuzzy numbers $\widetilde{A}$ and $\widetilde{B}$ is calculated as

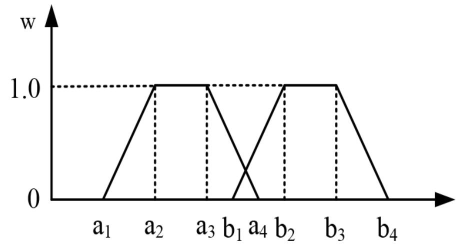


Figure 2: Trapezoidal fuzzy numbers $\widetilde{A}$ and $\widetilde{B}$

follows [15]:

$$S(\widetilde{A},\widetilde{B}) = 1 - \frac{\sum_{i=1}^{3}|a_i - b_i|}{3}. \tag{7}$$

Where$S(\widetilde{A},\widetilde{B}) \in [0,1]$. The larger the value of $S(\widetilde{A},\widetilde{B})$, the more the similarity between the fuzzy numbers $\widetilde{A}$ and $\widetilde{B}$.

Lee [16] presented a similarity measure between trapezoidal fuzzy numbers, where the degree of similarity $S(\widetilde{A},\widetilde{B})$ between the trapezoidal fuzzy numbers $\widetilde{A}$ and $\widetilde{B}$ is calculated as follows:

$$S(\widetilde{A},\widetilde{B}) = 1 - \frac{\left\|\widetilde{A}-\widetilde{B}\right\|_{l_p}}{\|U\|} \times 4^{-1/p}. \tag{8}$$

Where

$$\left\|\widetilde{A}-\widetilde{B}\right\|_{l_p} = \left(\sum_{i=1}^{4}(|a_i - b_i|)^p\right)^{1/p}. \tag{9}$$

and

$$\|U\| = \max(U) - \min(U). \tag{10}$$

In order to optimally aggregate experts' fuzzy opinions, Chen and Chen [17] presented a similarity measure between generalized trapezoidal fuzzy numbers. First, they calculate the center-of-gravity (COG) point $(x^*_{\widetilde{A}}, y^*_{\widetilde{A}})$ and $(x^*_{\widetilde{B}}, y^*_{\widetilde{B}})$ of the generalized trapezoidal fuzzy numbers $\widetilde{A}$ and $\widetilde{B}$, respectively. The COG point $(x^*_{\widetilde{A}}, y^*_{\widetilde{A}})$ of the generalized trapezoidal fuzzy numbers $\widetilde{A} = (a_1, a_2, a_3, a_4; w_{\widetilde{A}})$ is calculated as follows:

$$y^*_{\widetilde{A}} = \begin{cases} \dfrac{w_{\widetilde{A}} \times \left(\dfrac{a_3 - a_2}{a_4 - a_1} + 2\right)}{6} & \text{if} \quad a_1 \neq a_4 \quad \text{and} \quad 0 < w_{\widetilde{A}} \leq 1, \\ \dfrac{w_{\widetilde{A}}}{2} & \text{if} \quad a_1 = a_4 \quad \text{and} \quad 0 < w_{\widetilde{A}} \leq 1, \end{cases} \tag{11}$$

$$x^*_{\widetilde{A}} = \frac{y^*_{\widetilde{A}}(a_3 + a_2) + (a_4 + a_1)(w_{\widetilde{A}} - y^*_{\widetilde{A}})}{2w_{\widetilde{A}}}. \tag{12}$$

Then the degree of similarity $S(\widetilde{A},\widetilde{B})$ between the two generalized trapezoidal fuzzy numbers $\widetilde{A}$ and $\widetilde{B}$ can be calculated as follows [17]:

$$S(\widetilde{A},\widetilde{B}) = \left[1 - \frac{\sum_{i=1}^{4}|a_i - b_i|}{4}\right] \times \left(1 - \left|x^*_{\widetilde{A}} - x^*_{\widetilde{B}}\right|\right)^{B(S_{\widetilde{A}}, S_{\widetilde{B}})} \times \frac{\min(y^*_{\widetilde{A}}, y^*_{\widetilde{B}})}{\max(y^*_{\widetilde{A}}, y^*_{\widetilde{B}})} \tag{13}$$

Where $B(S_{\widetilde{A}}, S_{\widetilde{B}})$ are defined as follows:

$$B(S_{\widetilde{A}}, S_{\widetilde{B}}) = \begin{cases} 1 & S_{\widetilde{A}} + S_{\widetilde{B}} > 0 \\ 0 & S_{\widetilde{A}} + S_{\widetilde{B}} = 0 \end{cases} \tag{14}$$

Where $S_{\widetilde{A}} = a_4 - a_1$ and $S_{\widetilde{B}} = b_4 - b_1$ are the lengths of the generalized trapezoidal fuzzy numbers $\widetilde{A}$ and $\widetilde{B}$. The larger the value of $S(\widetilde{A}, \widetilde{B})$, the more the similarity measure between two fuzzy numbers.

Wei & Chen [18] proposed a method for calculating the similarity of two fuzzy numbers $\widetilde{A}$ and $\widetilde{B}$, where $\widetilde{A} = (a_1, a_2, a_3, a_4; w_{\widetilde{A}})$ and $\widetilde{B} = (b_1, b_2, b_3, b_4; w_{\widetilde{B}})$. If $0 \leq a_1 \leq a_2 \leq a_3 \leq a_4 \leq 1$ and $0 \leq b_1 \leq b_2 \leq b_3 \leq b_4 \leq 1$, then the degree of similarity $S(\widetilde{A}, \widetilde{B})$ between the generalized trapezoidal fuzzy numbers $\widetilde{A}$ and $\widetilde{B}$ is calculated as follows:

$$S(\widetilde{A}, \widetilde{B}) = \left[1 - \frac{\sum_{i=1}^{4} |a_i - b_i|}{4}\right] \times \frac{\min(P(\widetilde{A}), P(\widetilde{B})) + \min(w_{\widetilde{A}}, w_{\widetilde{B}})}{\max(P(\widetilde{A}), P(\widetilde{B})) + \max(w_{\widetilde{A}}, w_{\widetilde{B}})} \tag{15}$$

Where $S(\widetilde{A}, \widetilde{B}) \in [0, 1]$; $P(\widetilde{A})$ and $P(\widetilde{B})$ are defined as follows:

$$P(\widetilde{A}) = \sqrt{(a_1 - a_2)^2 + w_{\widetilde{A}}^2} + \sqrt{(a_3 - a_4)^2 + w_{\widetilde{A}}^2} + (a_3 - a_2) + (a_4 - a_1). \tag{16}$$

$$P(\widetilde{B}) = \sqrt{(b_1 - b_2)^2 + w_{\widetilde{B}}^2} + \sqrt{(b_3 - b_4)^2 + w_{\widetilde{B}}^2} + (b_3 - b_2) + (b_4 - b_1). \tag{17}$$

$P(\widetilde{A})$ and $P(\widetilde{B})$ are the perimeters of generalized trapezoidal fuzzy numbers $\widetilde{A}$ and $\widetilde{B}$, respectively. The larger the value of $S(\widetilde{A}, \widetilde{B})$, the more the similarity measure between two fuzzy numbers.

Hejazi, etc.[19] presented an improved similarity measure between two fuzzy numbers $\widetilde{A}$ and $\widetilde{B}$ combining the concept of geometric distance, height, areas and perimeters of generalized fuzzy numbers. The degree of similarity $S(\widetilde{A}, \widetilde{B})$ between the generalized trapezoidal fuzzy numbers $\widetilde{A}$ and $\widetilde{B}$ is calculated as follows:

$$S(\widetilde{A}, \widetilde{B}) = \left[1 - \frac{\sum_{i=1}^{4} |a_i - b_i|}{4}\right] \times \frac{\min(P(\widetilde{A}), P(\widetilde{B}))}{\max(P(\widetilde{A}), P(\widetilde{B}))} \times \frac{\min(A(\widetilde{A}), A(\widetilde{B})) + \min(w_{\widetilde{A}}, w_{\widetilde{B}})}{\max(A(\widetilde{A}), A(\widetilde{B})) + \max(w_{\widetilde{A}}, w_{\widetilde{B}})} \tag{18}$$

$P(\widetilde{A})$ and $P(\widetilde{B})$ are the perimeters of two generalized trapezoidal fuzzy numbers which are calculated by Eqs.(16),(17).On the other hand they have $A(\widetilde{A})$ and $A(\widetilde{B})$ which are the areas of the two fuzzy numbers and that are calculated as follows:

$$A(\widetilde{A}) = \frac{1}{2} w_{\widetilde{A}} (a_3 - a_2 + a_4 - a_1). \tag{19}$$

$$A(\widetilde{B}) = \frac{1}{2} w_{\widetilde{B}} (b_3 - b_2 + b_4 - b_1). \tag{20}$$

The larger the value of $S(\widetilde{A}, \widetilde{B})$, the more the similarity measure between two fuzzy numbers.

# 3 An Improved Similarity Measure of Generalized Fuzzy Numbers

Many similarity measures between fuzzy numbers have been proposed [15-19]. However, it has been found that the existing methods cannot correctly calculate the degree of similarity between two generalized fuzzy numbers in some situations. In this section, we present an improved method to calculate the degree of similarity between generalized fuzzy numbers[20],which gives consideration to the horizontal center of gravity, the perimeter, the height and the area of the two fuzzy numbers. The proposed similarity measure can overcome the drawbacks of the existing methods.

Assume there are two generalized trapezoidal fuzzy numbers $\widetilde{A}$ and $\widetilde{B}$, where $\widetilde{A} = (a_1, a_2, a_3, a_4; w_{\widetilde{A}})$ and $\widetilde{B} = (b_1, b_2, b_3, b_4; w_{\widetilde{B}})$, $0 \leq a_1 \leq a_2 \leq a_3 \leq a_4 \leq 1$ and $0 \leq b_1 \leq b_2 \leq b_3 \leq b_4 \leq 1$. Then the degree of similarity $S(\widetilde{A}, \widetilde{B})$ between the generalized trapezoidal fuzzy numbers $\widetilde{A}$ and $\widetilde{B}$ is calculated as follows:

$$S(\widetilde{A}, \widetilde{B}) = \left[1 - \left|x^*_{\widetilde{A}} - x^*_{\widetilde{B}}\right|\right] \times \left[1 - \left|w_{\widetilde{A}} - w_{\widetilde{B}}\right|\right] \times \frac{\min(P(\widetilde{A}), P(\widetilde{B})) + \min(A(\widetilde{A}), A(\widetilde{B}))}{\max(P(\widetilde{A}), P(\widetilde{B})) + \max(A(\widetilde{A}), A(\widetilde{B}))} \tag{21}$$

Where $x^*_{\widetilde{A}}$ and $x^*_{\widetilde{B}}$ are the horizontal center-of-gravity (COG) of the generalized trapezoidal fuzzy numbers $\widetilde{A}$ and $\widetilde{B}$, respectively. The COG point $(x^*_{\widetilde{A}}, y^*_{\widetilde{A}})$ of the generalized trapezoidal fuzzy numbers $\widetilde{A}$ is calculated as follows:

$$y^*_{\widetilde{A}} = \begin{cases} \dfrac{w_{\widetilde{A}} \times \left(\dfrac{a_3 - a_2}{a_4 - a_1} + 2\right)}{6} & \text{if} \quad a_1 \neq a_4 \quad \text{and} \quad 0 < w_{\widetilde{A}} \leq 1, \\ \dfrac{w_{\widetilde{A}}}{2} & \text{if} \quad a_1 = a_4 \quad \text{and} \quad 0 < w_{\widetilde{A}} \leq 1, \end{cases} \tag{22}$$

$$x^*_{\widetilde{A}} = \frac{y^*_{\widetilde{A}}(a_3 + a_2) + (a_4 + a_1)(w_{\widetilde{A}} - y^*_{\widetilde{A}})}{2w_{\widetilde{A}}}, \tag{23}$$

$P(\widetilde{A})$ and $P(\widetilde{B})$ are the perimeters of two generalized trapezoidal fuzzy numbers which are calculated as follows:

$$P(\widetilde{A}) = \sqrt{(a_1 - a_2)^2 + w^2_{\widetilde{A}}} + \sqrt{(a_3 - a_4)^2 + w^2_{\widetilde{A}}} + (a_3 - a_2) + (a_4 - a_1) \tag{24}$$

$$P(\widetilde{B}) = \sqrt{(b_1 - b_2)^2 + w^2_{\widetilde{B}}} + \sqrt{(b_3 - b_4)^2 + w^2_{\widetilde{B}}} + (b_3 - b_2) + (b_4 - b_1) \tag{25}$$

$A(\widetilde{A})$ and $A(\widetilde{B})$ are the areas of two generalized trapezoidal fuzzy numbers which are calculated as follows:

$$A(\widetilde{A}) = \frac{1}{2} w_{\widetilde{A}}(a_3 - a_2 + a_4 - a_1), \tag{26}$$

$$A(\widetilde{B}) = \frac{1}{2} w_{\widetilde{B}}(b_3 - b_2 + b_4 - b_1), \tag{27}$$

The larger the value of $S(\widetilde{A}, \widetilde{B})$ is, the more the similarity measure between two generalized trapezoidal fuzzy numbers $\widetilde{A}$ and $\widetilde{B}$ will be.

In the following sections, we will introduce some of properties that our model has:

*Theorem* 3.1. Two generalized trapezoidal fuzzy numbers $\widetilde{A}$ and $\widetilde{B}$ are identical if and only if $S(\widetilde{A}, \widetilde{B}) = 1$.

**Proof:**

(i) If $\widetilde{A}$ and $\widetilde{B}$ are identical, $x^*_{\widetilde{A}} = x^*_{\widetilde{B}}$, $w_{\widetilde{A}} = w_{\widetilde{B}}$, $\min(P(\widetilde{A}), P(\widetilde{B})) = \max(P(\widetilde{A}), P(\widetilde{B}))$, $\min(A(\widetilde{A}), A(\widetilde{B})) = \max(A(\widetilde{A}), A(\widetilde{B}))$.

The degree of similarity between two generalized trapezoidal fuzzy numbers is calculated as follows:

$$\begin{aligned} S(\widetilde{A},\widetilde{B}) &= \left[1-\left|x^*_{\widetilde{A}}-x^*_{\widetilde{B}}\right|\right]\times\left[1-\left|w_{\widetilde{A}}-w_{\widetilde{B}}\right|\right]\times\frac{\min(P(\widetilde{A}),P(\widetilde{B}))+\min(A(\widetilde{A}),A(\widetilde{B}))}{\max(P(\widetilde{A}),P(\widetilde{B}))+\max(A(\widetilde{A}),A(\widetilde{B}))} \\ &= [1-0]\times[1-0]\times 1 = 1 \end{aligned} \tag{28}$$

(ii) If $S(\widetilde{A},\widetilde{B})=1$, then

$$\begin{aligned} S(\widetilde{A},\widetilde{B}) = &\left[1-\left|x^*_{\widetilde{A}}-x^*_{\widetilde{B}}\right|\right]\times\left[1-\left|w_{\widetilde{A}}-w_{\widetilde{B}}\right|\right] \\ &\times\frac{\min(P(\widetilde{A}),P(\widetilde{B}))+\min(A(\widetilde{A}),A(\widetilde{B}))}{\max(P(\widetilde{A}),P(\widetilde{B}))+\max(A(\widetilde{A}),A(\widetilde{B}))} = 1 \end{aligned} \tag{29}$$

It implies that $x^*_{\widetilde{A}}=x^*_{\widetilde{B}}$, $w_{\widetilde{A}}=w_{\widetilde{B}}$, $\min(P(\widetilde{A}),P(\widetilde{B}))=\max(P(\widetilde{A}),P(\widetilde{B}))$ and $\min(A(\widetilde{A}),A(\widetilde{B}))=\max(A(\widetilde{A}),A(\widetilde{B}))$. Therefore, the generalized trapezoidal fuzzy numbers $\widetilde{A}$ and $\widetilde{B}$ are identical. □

*Theorem* 3.2. $S(\widetilde{A},\widetilde{B})=S(\widetilde{B},\widetilde{A})$.

**Proof:** Because

$$S(\widetilde{A},\widetilde{B}) = \left[1-\left|x^*_{\widetilde{A}}-x^*_{\widetilde{B}}\right|\right]\times\left[1-\left|w_{\widetilde{A}}-w_{\widetilde{B}}\right|\right]\times\frac{\min(P(\widetilde{A}),P(\widetilde{B}))+\min(A(\widetilde{A}),A(\widetilde{B}))}{\max(P(\widetilde{A}),P(\widetilde{B}))+\max(A(\widetilde{A}),A(\widetilde{B}))} \tag{30}$$

$$S(\widetilde{B},\widetilde{A}) = \left[1-\left|x^*_{\widetilde{B}}-x^*_{\widetilde{A}}\right|\right]\times\left[1-\left|w_{\widetilde{B}}-w_{\widetilde{A}}\right|\right]\times\frac{\min(P(\widetilde{B}),P(\widetilde{A}))+\min(A(\widetilde{B}),A(\widetilde{A}))}{\max(P(\widetilde{B}),P(\widetilde{A}))+\max(A(\widetilde{B}),A(\widetilde{A}))} \tag{31}$$

We can see that $\left|x^*_{\widetilde{A}}-x^*_{\widetilde{B}}\right| = \left|x^*_{\widetilde{B}}-x^*_{\widetilde{A}}\right|$, $\left|w_{\widetilde{A}}-w_{\widetilde{B}}\right| = \left|w_{\widetilde{B}}-w_{\widetilde{A}}\right|$, $\min(P(\widetilde{A}),P(\widetilde{B})) = \min(P(\widetilde{B}),P(\widetilde{A}))$, $\min(A(\widetilde{A}),A(\widetilde{B}))=\min(A(\widetilde{B}),A(\widetilde{A}))$, and $\max(A(\widetilde{A}),A(\widetilde{B}))=\max(A(\widetilde{B}),A(\widetilde{A}))$.

Therefore, $S(\widetilde{A},\widetilde{B})=S(\widetilde{B},\widetilde{A})$. □

*Theorem* 3.3. If $\widetilde{A}=(a_1,a_2,a_3,a_4;w_{\widetilde{A}})$ and $\widetilde{B}=(b_1,b_2,b_3,b_4;w_{\widetilde{B}})$ are two generalized trapezoidal fuzzy numbers with the same geometric shape and height, then $S(\widetilde{A},\widetilde{B})=1-d$, where $d=\left|x^*_{\widetilde{A}}-x^*_{\widetilde{B}}\right|$ is the offset between $\widetilde{A}$ and $\widetilde{B}$.

**Proof:** Because $w_{\widetilde{A}}=w_{\widetilde{B}}$, and based on Eq.(24) - Eq.(27),

we can get $\min(P(\widetilde{A}),P(\widetilde{B}))=\max(P(\widetilde{A}),P(\widetilde{B}))$ and $\min(A(\widetilde{A}),A(\widetilde{B}))=\max(A(\widetilde{A}),A(\widetilde{B}))$;

Therefore, due to Eq.(21), the degree of similarity between $\widetilde{A}$ and $\widetilde{B}$ is calculated as follows:

$$\begin{aligned} S(\widetilde{A},\widetilde{B}) &= \left[1-\left|x^*_{\widetilde{A}}-x^*_{\widetilde{B}}\right|\right]\times\left[1-\left|w_{\widetilde{A}}-w_{\widetilde{B}}\right|\right]\times\frac{\min(P(\widetilde{A}),P(\widetilde{B}))+\min(A(\widetilde{A}),A(\widetilde{B}))}{\max(P(\widetilde{A}),P(\widetilde{B}))+\max(A(\widetilde{A}),A(\widetilde{B}))} \\ &= \left[1-\left|x^*_{\widetilde{A}}-x^*_{\widetilde{B}}\right|\right]\times[1-0]\times 1 = 1-d \end{aligned} \tag{32}$$

□

## 4 A Comparison of The Similarity Measures

In this section, we extend 15 sets of fuzzy numbers presented by Wei & Chen [18] into 18 sets of fuzzy numbers, as shown in Figure 3, and compare the calculation results of the proposed method with the results of the existing similarity measures, as shown in Table 1. From Figure 3 and Table 1, we can see the drawbacks of the existing similarity measures:

(1) From Figure 3, we can see that Set 3 and Set 4 are different sets of fuzzy numbers. However, from Table 1, we can see that if we apply Chen's method (Chen, 1996) and Lee's method (Lee, 2002), Set 3 and Set 4 get the same degree of similarity, respectively.

(2) From Set 5 of Figure 3, we can see that $\widetilde{A}$ and $\widetilde{B}$ are different generalized fuzzy numbers. However, from Table 1, we can see that if we apply Chen's method (Chen, 1996) and Lee's method (Lee, 2002), their result is a degree of similarity equal to 1, respectively, which is an incorrect result.

(3) From Set 6 of Figure 3 and Table 1, we can see that if we apply Lee's method (Lee, 2002), we cannot calculate the degree of similarity between two identical real values due to the fact that the denominator will become zero, such that $S(\widetilde{A}, \widetilde{B}) = \infty$, which is an incorrect result.

(4) From Set 7 of Figure 3 and Table 1, we can see that if we apply Lee's method (Lee, 2002), we can see that Lee's method cannot correctly calculate the degree of similarity between two identical real values due to the fact that the degree of similarity of the real values become zero, which is an incorrect result.

(5) From Set 8 and Set 9 of Figure 3, we can see that they are two different sets of fuzzy numbers. However, from Table 1, we can see that if we apply Chen's method (Chen, 1996), they get the same degree of similarity, respectively, which does not coincide with the intuition of human being.

(6) From Set 10, Set 11 and Set 12 of Figure 3, we can see that they are different sets of generalized fuzzy numbers. However, from Table 1, we can see that if we apply Chen's method (Chen, 1996), they get the same degree of similarity, respectively, which does not coincide with the intuition of human being.

(7) From Set 7, Set 9 and Set 13 of Figure 3, we can see that $\widetilde{A}$ and $\widetilde{B}$ have the same shape and the offset $d = 0.1$ in the X-axis, respectively. By applying the proposed method, we can see that the proposed method has the good property that the degree of similarity between $\widetilde{A}$ and $\widetilde{B}$ is equal to$1 - |d| = 1 - 0.1 = 0.9$. However, from Table 1, we can see that if we apply Chen and Chen's method (Chen & Chen, 2003), the degree of similarity is equal to 0.81, which is an incorrect result.

(8) From Set 14 of Figure 3, using Chen's Method (Chen, 1996) and Lee's Method (Lee, 2002), the result is a degree of similarity equal to 1, respectively, which is an incorrect result.

(9) From Set 14 and Set 15 of Figure 3, we can see that Set 14 is more similar than Set 15 by the intuition of human being. However, from Table 1, we can see that if we apply Chen and Chen's method (Chen & Chen, 2003), we can see that it gets an incorrect result.

(10) From Figure 3, we can see that Set 10 and Set 16 are different sets of generalized fuzzy numbers and Set 10 is more similar than set 16 by the intuition of human being. However, from Table 1, we can see that if we apply the methods presented by Chen (1996), Lee (2002) and Hejazi et al. (2011), Set 10 and Set 16 get the same degree of similarity, respectively, and if we apply the method presented by Wei & Chen (2009), the result shows that Set 16 is more similar than Set 10. They are not the correct results.

(11) From Figure 3, we can see that Set 11 and Set 17 are different sets of generalized fuzzy numbers and Set 11 is more similar than Set 17 by the intuition of human being. However, from Table 1, we can see that if we apply the methods presented by Chen (1996), Lee (2002) and Hejazi et al. (2011), Set 11 and Set 17 get the same degree of similarity, respectively, and if we

apply the method presented by Chen & Chen (2003) and Wei & Chen (2009), the result shows that Set 17 is more similar than Set 11. They are not the correct results.

(12) From Figure 3, we can see that Set 11 and Set 18 are different sets of generalized fuzzy numbers. However, from Table 1, we can see that if we apply the methods presented by Chen (1996), Lee (2002), Hejazi et al. (2011) and Wei & Chen (2009), Set 11 and Set 18 get the same degree of similarity, respectively.

In summary, from Figure 3 and Table 1, we can see that the proposed method can overcome the drawbacks of the existing similarity measures.

Table 1: The fuzzy model of attribute SL

| Sets | Chen's method (1996)[15] | Lee's method (2002)[16] | Chen & Chen's method (2003)[17] | Wei & Chen's method (2009)[18] | Hejazi et al. method (2011)[19] | Proposed method |
|---|---|---|---|---|---|---|
| Set 1 | 0.9750 | 0.9617 | 0.8357 | 0.9500 | 0.9004 | 0.9473 |
| Set 2 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| Set 3 | **0.7000** | **0.5000** | 0.4200 | 0.6820 | 0.6465 | 0.6631 |
| Set 4 | **0.7000** | **0.5000** | 0.4900 | 0.700 | 0.7000 | 0.7000 |
| Set 5 | **1.0000** | **1.0000** | 0.8000 | 0.8248 | 0.6681 | 0.6659 |
| Set 6 | 1.0000 | * | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| Set 7 | 0.9000 | **0.0000** | **0.8100** | 0.9000 | 0.9000 | 0.9000 |
| Set 8 | **0.9000** | 0.5000 | 0.5400 | 0.8411 | 0.3700 | 0.3896 |
| Set 9 | **0.9000** | 0.6667 | **0.8100** | 0.9000 | 0.9000 | 0.9000 |
| Set 10 | **0.9000** | **0.8333** | 0.9000 | **0.7833** | **0.6261** | 0.7731 |
| Set 11 | **0.9000** | **0.7500** | **0.7200** | **0.8003** | **0.6448** | 0.7938 |
| Set 12 | **0.9000** | 0.8000 | 0.8325 | 0.8289 | 0.7361 | 0.7478 |
| Set 13 | 0.9000 | 0.7500 | **0.8100** | 0.9000 | 0.9000 | 0.9000 |
| Set 14 | **1.0000** | 1.0000 | **0.7000** | 0.7209 | 0.5113 | 0.5104 |
| Set 15 | 0.9500 | 0.7500 | **0.9048** | 0.6215 | 0.3830 | 0.4242 |
| Set 16 | **0.9000** | **0.8333** | 0.7425 | 0.8140 | 0.6261 | 0.7321 |
| Set 17 | **0.9000** | **0.7500** | **0.8911** | **0.8380** | **0.6448** | 0.7432 |
| Set 18 | **0.9000** | **0.7500** | 0.6976 | **0.8003** | **0.6448** | 0.7144 |

Note. "*" means that the similarity measure cannot calculate the degree of similarity between two generalized fuzzy numbers and the results that are not satisfactory are given in bold.

## 5 A New Method to Obtain BPA

In fact, some samples exist in many systems, which often approximatively submit the triangular distribution. Therefore, we use the existing sample data to build a triangular distribution to describe model of attribute categories, and then generate the BPA function based on the similarity between the collected attribute and the model attribute.

In order to be understood easily, the following Iris data classification problem shows the detailed approach of the proposed method.

In the Iris data, there are 3 species of Iris flower, i.e., *Setosa*, *Versicolor*, and *Virginica* [21]. The Iris data contain 150 instances, and each species contains 50 instances. There are four

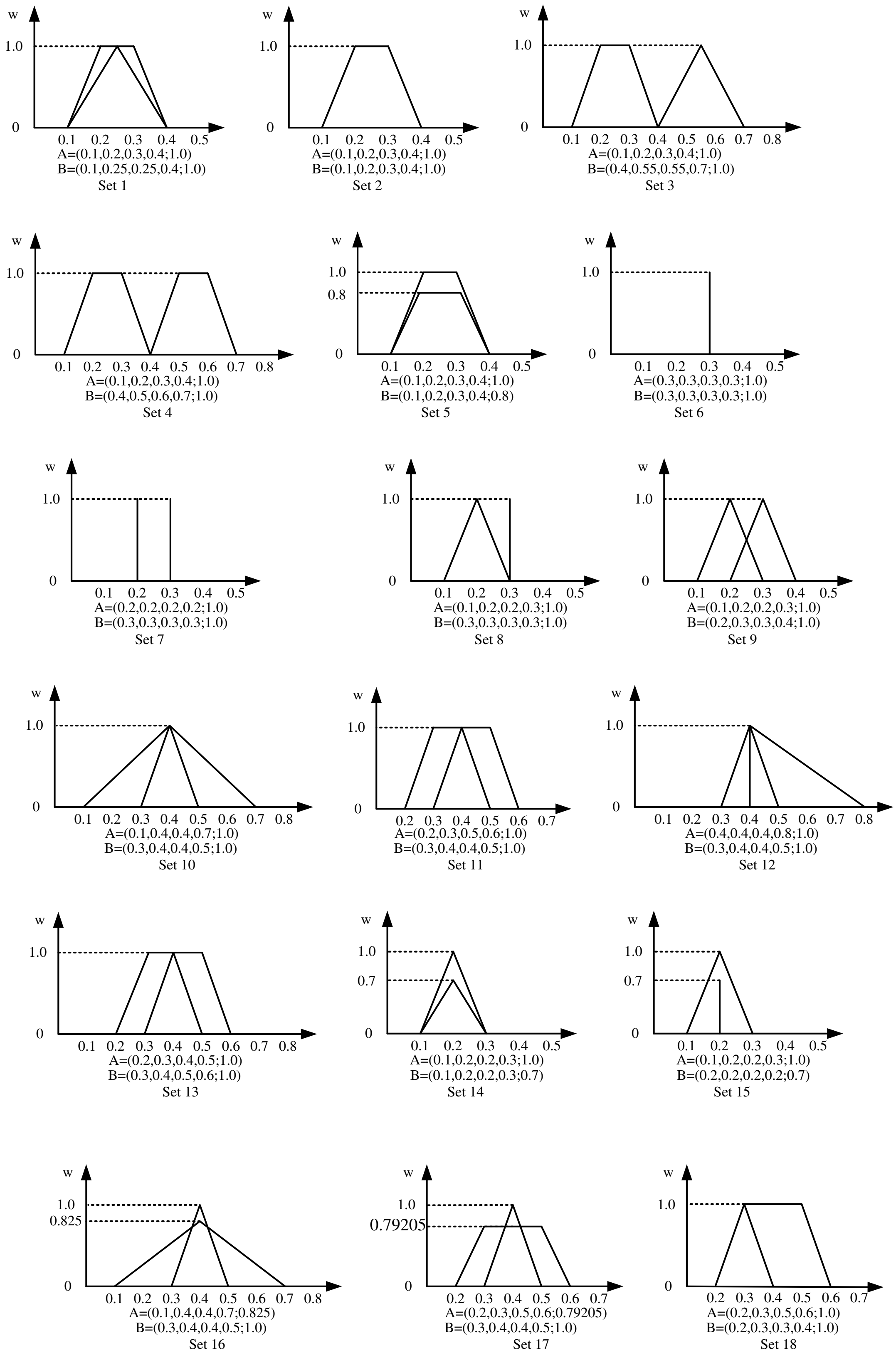


Figure 3: Eighteen sets of generalized fuzzy number $\widetilde{A}$ and $\widetilde{B}$

attributes in the Iris data, i.e., Sepal Length (SL), Sepal Width (SW), Petal Length (PL), and Petal Width (PW).

We randomly chose 40 instances from *Setosa*, the min(SL)=4.30; the average(SL)=5.04; the max(SL)=5.80 can be obtained. Hence, we can construct the fuzzy model of SL attribute of *Setosa* in Figure 4. In the same way, we can construct fuzzy models of Sepal Length(SL) of *Versicolor* and *Virginica*, as shown in Table 2 and Figure 4.

As can be seen from Figure 4, there are some crossing areas. For example, the crossing area of fuzzy number of *Setosa and Versicolour* can be shown in Figure 5. All the crossing areas can be modeled as generalized fuzzy numbers shown in Table 3.

Table 2: The fuzzy model of attribute SL

| Species | Setosa(S) | Versicolor(C) | Virginica(V) |
|---|---|---|---|
| Attribute (SL) | (4.30,5.04,5.80;1.0) | (5.0,5.90,6.8;1.0) | (5.6,6.59,7.90;1.0) |

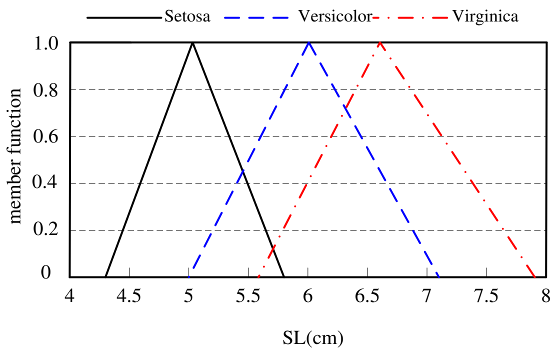


Figure 4: The fuzzy number representation of SL attribute of each species

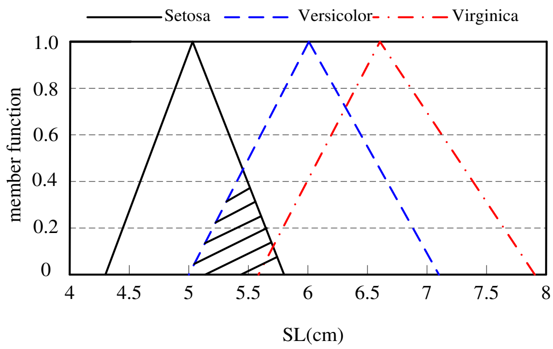


Figure 5: The generalized fuzzy number model of crossing area of *Setosa* and Versicolour

In a similar way, the fuzzy models of Sepal Width (SW) attribute, Petal Length (PL) attribute, and Petal Width (PW) attribute of each species can be constructed.

Table 3: The generalized fuzzy number model of crossing area of three species Iris flowers

| Species Attribute (SL) | S&C | S&V | V&C | S&C&V |
|---|---|---|---|---|
| | (5.00,5.43,5.80;0.48) | (5.60,5.71,5.80;0.11) | (5.60,6.23,6.80;0.63) | (5.60,5.71,5.80;0.11) |

We randomly chose a datum from Iris source; for example, a new instance (NI) of *Setosa* could be shown as (5.1cm, 3.5cm, 1.4cm, 0.2cm). In Figure 6, the relation between SL attribute of NI and the fuzzy number representation of SL attribute of each species is distinctly shown.

To calculate the similarities between NI and each generalized fuzzy number and normalize the obtained similarities, the BPA can be obtained, as shown in table 4.

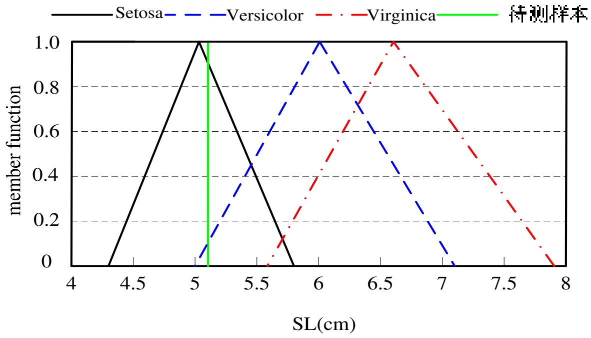


Figure 6: The relation between the new instance and the model of SL attribute of each species

Table 4: The BPA of an instance attribute SL

| | m(S) | m(C) | m(V) | m(S,C) | m(V,S) | m(C,V) | m(S,V,C) |
|---|---|---|---|---|---|---|---|
| SL | 0.2955 | 0.2605 | 0.2215 | 0.0842 | 0.0046 | 0.1291 | 0.0046 |

In the same way, the four attributes and their relative BPAs can be obtained, as shown in Table 5.

With the combination rule in Eq.(3) used, the fusion results can be shown as follows:

$$m(S) = 0.5324, \quad m(C) = 0.2607, \quad m(V) = 0.1879, \quad m(S,C) = 0.0112,$$
$$m(V,S) = 0.0015, \quad m(C,V) = 0.0063, \quad m(S,V,C) = 0.0001$$

Hence, the instance can be classed as *Setosa*. The result is consistent with the actual situation.

The algorithm of our proposed method can be listed step by step as follows.

*Step 1:* Use the existing sample data to obtain the min, average and max value to construct the triangular fuzzy models, which describe the model attributes of instances.

*Step 2:* Calculate the similarity between the collected attribute and the model attribute.

*Step 3:* Normalize the similarity measure to obtain the BPA function.

We randomly selected 120 instances, 40 instances for each of the 3 species, to construct species models. The remaining 30 instances, 10 instances for each of the 3 species, were used as collected instances whose class was unknown. By applying the method proposed in the above section to classify the 30 instances, the correct rate of Iris data classification was calculated as 96.67%.

Table 5: The BPA of an instance

| | m(S) | m(C) | m(V) | m(S,C) | m(V,S) | m(C,V) | m(S,V,C) |
|---|---|---|---|---|---|---|---|
| SL | 0.2955 | 0.2605 | 0.2215 | 0.0842 | 0.0046 | 0.1291 | 0.0046 |
| SW | 0.1727 | 0.1614 | 0.1687 | 0.0925 | 0.1510 | 0.1611 | 0.0925 |
| PL | 0.2198 | 0.1193 | 0.0731 | 0.1943 | 0.1943 | 0.0049 | 0.1943 |
| PW | 0.1921 | 0.0988 | 0.0499 | 0.2143 | 0.2143 | 0.0162 | 0.2143 |

Further, we applied the proposed method 10 times; the average correct rate of Iris data classification was up to 95.67%. It can be seen that our proposed method has good results in classification problem.

## 6 Conclusions

The estimation of BPA plays a very important role in the application of Dempster-Shafer theory in complex uncertain problems. The fusion performance depends on the method of BPA construction. To solve this problem, firstly, the paper presents an improved method to calculate the degree of similarity between generalized fuzzy numbers, which gives consideration to the horizontal center of gravity, the perimeter, the height and the area of the two fuzzy numbers. The method can overcome the drawbacks of the existing similarity measures. Then, a new method to obtain BPA is proposed based on the improved similarity measure between generalized fuzzy numbers. The proposed method to obtain BPA can effectively overcome the problem of subjectivity, which also has strong generality. The classification of Iris data is used to illustrate the efficiency and the low computational complexity of the proposed method. The proposed method provides a simple technique that will help to use the classical Dempster combination rule effectively.

### Acknowledgment

This work was supported in part by a grant from National Natural Science Foundation of China (No. 61104214) and Foundation for Fundament Research of Northwestern Polytechnical University, Grant No.JC20120235.

## Bibliography

[1] A. Dempster(1967), Upper and lower probabilities induced by multivalued mapping, *Annals of Mathematical Statistics*, ISSN 0003-4851, 38(2): 325-339.

[2] G. Shafer(1976), *A mathematical theory of evidence*, Princeton University Press, ISBN 978-069-11-0042-5.

[3] Y.M. Zhu, L. Bentabet, M. Rombaut, O. Dupuis, V. Kaftandjian, D. Babot(2002), Automatic determination of mass functions in DS theory using FCM and spatial neighbourhood information for image segmentation, *Optical Engineering*, ISSN 0091-3286, 41(4): 760-770.

[4] A. Bendjebbour, Y. Delignon, L. Fouque, V. Samson, W. Pieczynski(2001), Multisensor image segmentation using DS fusion in Markov fields context, *IEEE Trans. Geosci. Remote Sensing*, ISSN 0196-2892, 39(8): 1789-1798.

[5] X. Guan, X. Yi, Y. He(2008), Study on algorithms of determining basic probability assignment function in Dempster-Shafer evidence theory, *Proc. of the 7th Int. Conf. on Machine Learning and Cybernetics*, 121-126.

[6] B. Chen, J.F. Wang, S.B. Chen(2010), Prediction of pulsed GTAW penetration status based on BP neural network and D-S evidence theory information fusion, *International Journal of Advanced Manufacturing Technology*, ISSN 0268-3768, 48(1-4): 83-94.

[7] X.M. Li, L.X. Ding, Y. Li, G. Xu, J.B. Li(2009), HVAC fan mechinery fault diagnosis based on ANN and D-S evidence theory, *IITA Int. Conf. on Control, Automation and Systems Engineering*, Zhangjiajie, China, 603-606.

[8] Z. Xu, M. Liu, G. Yang, N. Li(2009), Application of interval analysis and evidence theory to fault location, *IET Electric Power Application*, ISSN 1751-8660, 3(1): 77-84.

[9] Z.Y. Zuo, Y.F. Xu, G.C. Chen(2009), A new method of obtaining BPA and application to the bearing fault diagnosis of wind turbine, *Proc. of the 2009 Int. Symposium on Information Processing*, Huangshan, China, 368-371.

[10] W. Jiang, J.Y. Peng, Y. Deng(2011), A new method to determine BPA in evidence theory, *Journal of Computers*, ISSN 1796-203X, 6(6): 1162-1167.

[11] Y. Deng, W. Jiang, R. Sadiq(2011), Modeling contaminant intrusion in water distribution networks: A new similarity-based DST method, *Expert Systems with Applications*, ISSN 0957-4174, 38(1): 571-578.

[12] Y. Deng, R. Sadiq, W. Jiang, S. Tesfamariam(2011), Risk analysis in a linguistic environment: A fuzzy evidential reasoning-based approach, *Expert Systems with Applications*, ISSN 0957-4174, 38(12): 15438-15446.

[13] Y. Deng, W. Jiang, X.B. Xu(2009), Determinging BPA under uncertainty environments and its application in data fusion, *Journal of Electronics (China)*, ISSN 0217-9822, 26(1): 13-17.

[14] W. Jiang, Y. Luo, X.Y. Qin, J. Zhan(2015), An Improved Method to Rank Generalized Fuzzy Numbers with Different Left Heights and Right Heights, *Journal of Intelligent and Fuzzy Systems*, Accepted.

[15] S. M. Chen(1996), Foreword, New methods for subjective mental workload assessment and fuzzy risk analysis, *Cybernetics and Systems: An International Journal*, ISSN 0196-9722, 27(5): 449-472.

[16] H.S. Lee(2002), Optimal consensus of fuzzy opinions under group decision making environment, *Fuzzy Sets and Systems*, ISSN 0196-9722, 132(3): 303-315.

[17] S.J. Chen, S.M. Chen(2003), Fuzzy risk analysis based on similarity measures of generalized fuzzy numbers, *IEEE Transaction on Fuzzy Systems*, ISSN 1063-6706, 11(1): 45-56.

[18] S.H. Wei, S.M. Chen(2009), A new approach for fuzzy risk analysis based on similarity measures of generalized fuzzy numbers, *Expert Systems with Applications*, ISSN 0957-4174, 36(1): 589-598.

[19] S.R. Hejazi, A. Doostparast, S.M. Hosseini(2011), An improved fuzzy risk analysis based on a new similarity measures of generalized fuzzy numbers, *Expert Systems with Applications*, ISSN 0957-4174, 38(8): 9179-9185.

[20] W. Jiang, X. Fan, D.J. Duanmu, Y.Deng(2011), A Modified Similarity Measure of Generalized Fuzzy Numbers, *2011 Int. Conf. on Advanced in Control Engineering and Information Science*, 2773-2777.

[21] R.A. Fisher(1936), The use of multiple measurements in taxonomic problems, *Annals of Eugenics*, ISSN 1469-1809, 7(2): 179-188.