



An evidence clustering DSMT approximate reasoning method based on convex functions analysis



Qiang Guo^{a,*}, You He^a, Xin Guan^b, Li Deng^c, Lina Pan^d, Tao Jian^a

^a Research Institute of Information Fusion, Naval Aeronautical and Astronautical University, Yantai, Shandong 264001, China

^b Electronics and Information Department, Naval Aeronautical and Astronautical University, Yantai, Shandong 264001, China

^c Department of Armament Science and Technology, Naval Aeronautical and Astronautical University, Yantai, Shandong 264001, China

^d Department of Basic Science, Naval Aeronautical and Astronautical University, Yantai, Shandong 264001, China

ARTICLE INFO

Article history:

Available online 23 June 2015

Keywords:

Evidence clustering
Approximate reasoning
Information fusion
Convex functions analysis
Dezert–Smarandache Theory

ABSTRACT

The computational complexity of Dezert–Smarandache Theory (DSMT) increases exponentially with the linear increment of element number in the discernment frame, and it limits the wide applications and development of DSMT. In order to efficiently reduce the computational complexity and remain high accuracy, a new Evidence Clustering DSMT Approximate Reasoning Method for two sources of information is proposed based on convex function analysis. This new method consists of three steps. First, the belief masses of focal elements in each evidence are clustered by the Evidence Clustering method. Second, the un-normalized approximate fusion results are obtained using the DSMT approximate convex function formula, which is acquired based on the mathematical analysis of Proportional Conflict Redistribution 5 (PCR5) rule in DSMT. Finally, the normalization step is applied. The computational complexity of this new method increases linearly rather than exponentially with the linear growth of the elements. The simulations show that the approximate fusion results of the new method have higher Euclidean similarity to the exact fusion results of PCR5 based information fusion rule in DSMT framework (DSMT + PCR5), and it requires lower computational complexity as well than the existing approximate methods, especially for the case of large data and complex fusion problems with big number of focal elements.

© 2015 Elsevier Inc. All rights reserved.

1. Introduction

As a novel key technology with vigorous development, information fusion can integrate multiple-source incomplete information and reduce uncertainty of information which always has the contradiction and redundancy. Information fusion can improve rapid correct decision capacity of intelligent systems and has been successfully used in the military and economy fields, thus great attention has been paid to its development and application by scholars in recent years [1–9]. As information environment becomes more and more complex, greater demands for efficient fusion of highly conflicting and uncertain evidence are being placed on information fusion. Belief function theory (also called evidence theory) referred by Dezert–Smarandache Theory (DSMT) [9] and Dempster–Shafer Theory (DST) [10,11] can well deal with the uncertain and conflict information. DSMT, jointly proposed by Dezert and Smarandache, is considered as the general extension of DST, since it beyonds the exclusiveness limitation of elements in DST. DSMT can ob-

tain precise results for dealing with complex fusion problems in which the conflict is high and the refinement of the frame is not accessible [9]. Recently, DSMT (belief function theory) has been successfully applied in many areas, such as, Map Reconstruction of Robot [12,13], Decision Making Support [14], Target Type Tracking [15,16], Image Processing [17], Sonar Imagery [18], Data Classification [19–21], Clustering [22,23], and so on. Particularly, the very recent credal classification methods [12,15,20] working with belief functions have been introduced by Liu, Dezert, et al. for dealing with uncertain data, and the object is allowed to belong to any singleton class and set of classes (called meta-class) with different belief masses. By doing this, the credal classifiers are able to well capture the uncertainty of classification and also efficiently reduce the errors. However, the main problem of the application (e.g. classification task) of DSMT is that when the focal elements' number increases linearly, computational complexity increases exponentially.

Many approximate reasoning methods of evidence combination in DST framework were presented in [24–26]. But these methods cannot satisfy the small amount of computational complexity and less loss of information requirements at the same time. In recent years, there are some important articles [27–34] dealing with the

* Corresponding author.

E-mail address: gq19860209@163.com (Q. Guo).

computational complexity of the combination algorithms formulated in DSMT framework in different ways. Djiknavorian [27] has proposed a novel method and a Matlab program to reduce the DSMT hybrid rule complexity. Smarandache has proposed Smarandache's codification [35] to describe the parts of a Venn diagram in DSMT framework, which is easy to read and apply in unions and intersections of sets [28]. For manipulating the focal elements easily, Martin [28] has proposed a Venn diagram codification, which is more complex for the readers than Smarandache's codification. However, the DSMT (belief function theory) complexity can be reduced by using this practical Martin's codification and only considering the reduced hyper-power set D_r^Θ after integrating the constraints in the codification at the beginning of the general belief function framework proposed in [28]. Abbas [29,30] has proposed a DSMT based combination scheme for multi-class classification which also reduces the number of focal elements. Li [31] has proposed a method for reducing the information fusion complexity, which is different from the above methods by reducing the combined sources numbers instead of reducing the number of focal elements. Li and other scholars [32–34] also proposed an approximate reasoning method for reducing the complexity of the Proportional Conflict Redistribution 5 (PCR5) based information fusion rule within DSMT framework. However, when processing highly conflict evidences by the method in [32], the belief assignments of correct main focal elements transfer to the other focal elements, which leads to low Euclidean similarity of the results in this case.

Aiming at reducing the computational complexity of PCR5 based information fusion rule within DSMT framework (DSMT + PCR5) and obtaining accurate results in any case, a new Evidence Clustering DSMT Approximate Reasoning Method for two sources of information is proposed in this paper. In Section 2, the basics knowledge on DST, DSMT and the dissimilarity measure method of multi evidences are introduced briefly. In Section 3, mathematical analysis of PCR5 formula is conducted, which discovers every conflict mass product satisfies the properties of convex function. A new DSMT approximate convex function formula is proposed and error analysis of the proposed formula is also presented. Based on the error analysis, an Evidence Clustering method is proposed as the preprocessing step and the normalization method is applied as the final step of the proposed method for reducing the approximate error. The process of the proposed method is given, then analysis of computation complexity of DSMT + PCR5 and the proposed method are presented. In Section 4, the results of simulation show that the approximate fusion results of the method proposed in this paper have higher Euclidean similarity with the exact fusion results of DSMT + PCR5, and lower computational complexity than existing DSMT approximate reasoning method in [32]. The conclusions are given in Section 5.

2. Basic knowledge

In this section, we will give an overview of the basics knowledge on DST and DSMT, which are closely related to our work in this paper.

2.1. Dempster–Shafer Theory (DST)

Let us consider a discernment frame $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ containing n elements $\theta_1, \theta_2, \dots, \theta_n$, which is the refinement of the discernment based on Shafer's model. The basic belief assignment (bba) is defined over the power-set 2^Θ which consists of all subsets of Θ . For example, if one has $\Theta = \{\theta_1, \theta_2, \theta_3\}$, the power set is given by $2^\Theta = \{\emptyset, \theta_1, \theta_2, \theta_3, \theta_1 \cup \theta_2, \theta_1 \cup \theta_3, \theta_2 \cup \theta_3, \theta_1 \cup \theta_2 \cup \theta_3\}$, and the bba $m(\cdot) : 2^\Theta \rightarrow [0, 1]$ on the power set is defined by [10,11]

$$m(X_i) = 0, \quad X_i = \emptyset \tag{1}$$

$$\sum_{X_i \in 2^\Theta, 1 \leq i \leq n} m(X_i) = 1 \tag{2}$$

The element X_i is called focal elements if it holds $m(X_i) > 0$. Dempster's rule is often used for the combination of multiple sources of evidence represented by bba's in Shafer's model, and it requires that the bba's must be independent. The bba of the i th source of evidence is denoted m_i . The Dempster combination rule is defined by Equation (3) and the conflict in Dempster combination rule, denoted by C , is defined by Equation (4) [10,11]

$$m_{DS}(Z) = \frac{1}{1-C} \sum_{X_i \cap X_j = Z, i \neq j} m_1(X_i) \cdot m_2(X_j) \quad \forall Z \subseteq \Theta \tag{3}$$

$$C = \sum_{\substack{X_i, X_j \subseteq \Theta, i \neq j \\ X_i \cap X_j = \emptyset}} m_1(X_i) \cdot m_2(X_j) \tag{4}$$

One can see that all the conflicting beliefs C has been redistributed to other focal elements. Dempster's rule usually produces very unreasonable results in the fusion of high conflicting information due to the redistribution of conflicting beliefs. In order to solve this problem, many alternative combination rules like Proportional Conflict Redistribution 1–6 (PCR1-6) rules [36,38,39] have been developed.

2.2. Dezert-Smarandache Theory (DSMT)

DSMT [29] overcomes the exclusiveness limitation in Shafer's model. In many fusion problems, the hypotheses can be vague in reality and the elements are not precisely separated which don't satisfy Shafer's model. The hyper-power set denoted by D^Θ is built by applying operator \cap and \cup to the elements in Θ [36,37]. Let us consider a simple frame of discernment $\Theta = \{\theta_1, \theta_2\}$, then one gets $D^\Theta = \{\emptyset, \theta_1, \theta_2, \theta_3, \theta_1 \cup \theta_2, \theta_1 \cap \theta_2\}$. The bba in DSMT is defined over the hyper-power set as $m(\cdot) : D^\Theta \rightarrow [0, 1]$.

In the combination of multiple sources of evidence, there exist two models in DSMT [36,37]: 1) free combination model and 2) hybrid combination model which is often used in real application because it takes into account some integrity constraints. In hybrid combination rule, it transfers partial conflicting beliefs to the corresponding intersected elements, but this increases the uncertainty of fusion results. The Proportional Conflict Redistribution 1–6 (PCR1-6) rules [37,39,40] provide proper conflict redistribution ways, and they proportionally transfer conflicting masses to the involved elements.

The difference of PCR1-6 rules mainly lies in the redistribution of conflicts, and PCR5 is considered as the most precise redistribution way [37,39,40]. The combination of two independent sources of evidences by PCR5 rule is given as follows [37,39,40]

$$m_{1 \oplus 2}(X_i) = \sum_{\substack{Y, Z \in G^\Theta \text{ and } Y, Z \neq \emptyset \\ Y \cap Z = X_i}} m_1(Y) \cdot m_2(Z) \tag{5}$$

$$m_{PCR5}(X_i) = \begin{cases} m_{1 \oplus 2}(X_i) + \sum_{\substack{X_j \in G^\Theta \text{ and } i \neq j \\ X_i \cap X_j = \emptyset}} \left[\frac{m_1(X_i)^2 \cdot m_2(X_j)}{m_1(X_i) + m_2(X_j)} \right. \\ \left. + \frac{m_2(X_i)^2 \cdot m_1(X_j)}{m_2(X_i) + m_1(X_j)} \right] & X_i \in G^\Theta \text{ and } X_i \neq \emptyset \\ 0 & X_i = \emptyset \end{cases} \tag{6}$$

where G^Θ can be seen as the power set 2^Θ , the hyper-power set D^Θ and the super-power set S^Θ , if discernment of the fusion problem satisfies Shafer's model, the hybrid DSMT model, and the minimal refinement Θ^{ref} of Θ respectively and where all denominators are more than zero and the fraction is discarded when the denominator of it is zero [37,39,40].

Nevertheless, PCR5 rule still has some disadvantages, such as, firstly, it is not associative in the fusion of multiple (more than 2) sources of evidences, so the combination order may have influence on the results, secondly, its computational complexity increases exponentially, when the focal elements' number increases. Our research in this paper is mainly for reducing the complexity of PCR5 within DSMT framework.

2.3. The dissimilarity measure method of multi evidences

The dissimilarity measure method of multi evidences and several Evidence Support Measure of Similarity (ESMS) functions have been given in [31,41]. The often used Euclidean ESMS function and Josselme ESMS function are briefly recalled.

1) Euclidean ESMS function $Sim_E(m_1, m_2)$

Let $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$, $n > 1$, $|G^\Theta|$ be the cardinality of G^Θ , $m_1(\cdot)$ and $m_2(\cdot)$ be two bba's. The Euclidean ESMS function is defined by [31]

$$Sim_E(m_1, m_2) = 1 - \frac{1}{\sqrt{2}} \sqrt{\sum_{i=1}^{|G^\Theta|} [m_1(X_i) - m_2(X_i)]^2} \quad (7)$$

2) Josselme ESMS function $Sim_J(m_1, m_2)$

The Josselme ESMS function [31] is defined based on the Josselme et al. measure [41]

$$Sim_J(m_1, m_2) = 1 - \frac{1}{\sqrt{2}} \sqrt{(m_1 - m_2)^T D (m_1 - m_2)} \quad (8)$$

where $D = [D_{ij}]$ is a $|G^\Theta| \times |G^\Theta|$ positively definite matrix, and $D_{ij} = |X_i \cap X_j| / |X_i \cup X_j|$ with $X_i, X_j \in G^\Theta$.

Some more ESMS functions can be seen in [31] for details. $Sim_E(m_1, m_2)$ is considered with the fastest convergence speed [31], and it is adopted here as the dissimilarity measure for comparison of the method proposed in this paper with the other methods.

3. An evidence clustering DSMT approximate reasoning method

3.1. Mathematical analysis of PCR5 formula

As shown in Equation (6), $\sum_{\substack{X_j \in G^\Theta \\ X_i \cap X_j = \emptyset}} [\frac{m_1(X_i)^2 \cdot m_2(X_j)}{m_1(X_i) + m_2(X_j)} + \frac{m_2(X_j)^2 \cdot m_1(X_i)}{m_2(X_j) + m_1(X_i)}]$ has symmetry.

Due to the symmetry, one item $\frac{m_1(X_i)^2 \cdot m_2(X_j)}{m_1(X_i) + m_2(X_j)}$ is analyzed.

Let $m_1(X_i) = a$ and $m_2(X_j) = x$ get

$$\frac{m_1(X_i)^2 \cdot m_2(X_j)}{m_1(X_i) + m_2(X_j)} = \frac{a^2 x}{a + x} = a^2 \left[1 - a \left(\frac{1}{a + x} \right) \right]. \quad (9)$$

Let $x_1, x_2, \dots, x_n \in \{m_2(X_j) | i \neq j, X_j \in G^\Theta \text{ and } X_i \cap X_j = \emptyset\}$, then

$$\sum_{\substack{X_j \in G^\Theta \\ X_i \cap X_j = \emptyset}} \left[\frac{m_1(X_i)^2 \cdot m_2(X_j)}{m_1(X_i) + m_2(X_j)} \right] = a^2 \left[n - a \left(\frac{1}{a + x_1} + \frac{1}{a + x_2} + \dots + \frac{1}{a + x_n} \right) \right]. \quad (10)$$

Let $f(x) = \frac{1}{a+x}$, since $f(x)$ is continuous function on $(0, 1)$, it has a second order derivatives on $(0, 1)$, and $f''(x) > 0$ on $(0, 1)$, $f(x)$ is a convex function.

So $\frac{1}{n}(f(x_1) + f(x_2) + \dots + f(x_n)) \geq f(\frac{x_1+x_2+\dots+x_n}{n})$, the equation holds iff $x_1 = x_2 = \dots = x_n$.

The approximate convex function formula is given by

$$\begin{aligned} & \frac{1}{a + x_1} + \frac{1}{a + x_2} + \dots + \frac{1}{a + x_n} \\ &= \frac{n}{a + (x_1 + x_2 + \dots + x_n)/n} + \Delta, \\ & \Delta \geq 0, \Delta = 0 \text{ iff } x_1 = x_2 = \dots = x_n. \end{aligned} \quad (11)$$

Let $x_1 \leq x_2 \leq \dots \leq x_i \leq \dots \leq x_n$, carry out analysis of convex function formula errors

$$\begin{aligned} \Delta &= \left[\frac{1}{a + x_1} - \frac{1}{a + (x_1 + x_2 + \dots + x_n)/n} \right] \\ &+ \left[\frac{1}{a + x_2} - \frac{1}{a + (x_1 + x_2 + \dots + x_n)/n} \right] + \dots \\ &+ \left[\frac{1}{a + x_n} - \frac{1}{a + (x_1 + x_2 + \dots + x_n)/n} \right]. \end{aligned} \quad (12)$$

Analysis of the i item in Equation (12).

Let $(x_1 + x_2 + \dots + x_n)/n = x_0$, then

$$\frac{1}{a + x_i} - \frac{1}{a + (x_1 + x_2 + \dots + x_n)/n} = \frac{1}{a + x_i} - \frac{1}{a + x_0}. \quad (13)$$

By Taylor expansion theorem

$$\begin{aligned} & \frac{1}{a + x_i} - \frac{1}{a + x_0} \\ &= f'(x_0)(x_i - x_0) + \frac{f''(x_0)}{2}(x_i - x_0)^2 + \frac{f'''(\delta)}{3!}(x_i - x_0)^3 + \dots, \\ & \delta \in (x_i, x_0) \text{ or } (x_0, x_i), \end{aligned} \quad (14)$$

then Equation (14) is transformed to

$$\begin{aligned} & \frac{1}{a + x_1} + \frac{1}{a + x_2} + \dots + \frac{1}{a + x_n} - \frac{n}{a + (x_1 + x_2 + \dots + x_n)/n} \\ &= f'(x_0)[(x_1 - x_0) + (x_2 - x_0) + \dots + (x_n - x_0)] \\ &+ \frac{f''(x_0)}{2} [(x_1 - x_0)^2 + (x_2 - x_0)^2 + \dots + (x_n - x_0)^2] \\ &+ \sum_{i=1}^n o(x_i - x_0)^2. \end{aligned} \quad (15)$$

Since $f'(x_0)[(x_1 - x_0) + (x_2 - x_0) + \dots + (x_n - x_0)] = 0$, then

$$\begin{aligned} & \frac{1}{a + x_1} + \frac{1}{a + x_2} + \dots + \frac{1}{a + x_n} - \frac{n}{a + (x_1 + x_2 + \dots + x_n)/n} \\ &= \frac{f''(x_0)}{2} \sum_{i=1}^n (x_i - x_0)^2 + \sum_{i=1}^n o(x_i - x_0)^2, \end{aligned} \quad (16)$$

where

$$\begin{aligned} & \sum_{i=1}^n o(x_i - x_0)^2 \\ &= \frac{f'''(x_0)}{3!} [(x_1 - x_0)^3 + (x_2 - x_0)^3 + \dots + (x_n - x_0)^3] \\ &+ \frac{f''''(\delta_1)}{4!} (x_1 - x_0)^4 + \frac{f''''(\delta_2)}{4!} (x_2 - x_0)^4 + \dots \\ &+ \frac{f''''(\delta_n)}{4!} (x_n - x_0)^4 + \dots. \end{aligned}$$

Analysis of $|f^{(m)}(x)|$, $m = 2, 3, \dots, \infty$

$$|f^{(m)}(x)| = \left| \left(\frac{1}{a+x} \right)^{(m)} \right| = m \left(\frac{1}{a+x} \right)^{(m-1)}, \quad (17)$$

then

$$\begin{aligned} & \left| \frac{f^{(m-1)}(x_0)}{(m-1)!} (x-x_0)^{m-1} \right| - \left| \frac{f^{(m)}(x_0)}{m!} (x-x_0)^m \right| \\ &= \frac{1}{(m-2)!} \left(\frac{1}{a+x_0} \right)^{(m-2)} |(x-x_0)^{m-1}| \\ & \quad - \frac{1}{(m-1)!} \left(\frac{1}{a+x_0} \right)^{(m-1)} |(x-x_0)^m| \\ &= \frac{1}{(m-2)!} \left(\frac{1}{a+x_0} \right)^{(m-2)} |(x-x_0)^{m-1}| \\ & \quad \times \left[1 - \frac{1}{(m-1)} \left(\frac{1}{a+x_0} \right) |x-x_0| \right]. \end{aligned} \quad (18)$$

If $x \leq x_0$, $|x-x_0| = x_0 - x < x_0 + a$, then

$$\left| \frac{f^{(m-1)}(x_0)}{(m-1)!} (x-x_0)^{m-1} \right| > \left| \frac{f^{(m)}(x_0)}{m!} (x-x_0)^m \right|. \quad (19)$$

If $m \geq 2$, $x > x_0$, $x < a + 2x_0$, then

$$\frac{1}{m-1} \cdot \left(\frac{1}{a+x_0} \right) |x-x_0| > 0. \quad (20)$$

So if $x_i < a + 2x_0$, $i = 1, 2, \dots, n$,

$$\left| \frac{f^{(m-1)}(x_0)}{(m-1)!} (x_i-x_0)^{m-1} \right| > \left| \frac{f^{(m)}(x_0)}{m!} (x_i-x_0)^m \right|, m \geq 2, \quad (21)$$

namely,

$$\begin{aligned} & \left| \frac{f''(x_0)}{2} (x_i-x_0)^2 \right| > \left| \frac{f'''(x_0)}{3!} (x_i-x_0)^3 \right| > \dots \\ & > \left| \frac{f^{(m)}(x_0)}{m!} (x_i-x_0)^m \right|. \end{aligned} \quad (22)$$

Neglect the fourth order item errors and more order item errors.

For the third order item is odd number item, for each x_i , $i = 1, 2, \dots, n$, $\frac{f^3(x_0)}{3!} (x_i-x_0)^3$ can be positive and negative. Then the sum of the third order items is much smaller than the sum of the second order items if $x_i < a + 2x_0$, $i = 1, 2, \dots, n$. Neglect the third order item and more order item errors if $x_i < a + 2x_0$, $i = 1, 2, \dots, n$.

So,

$$\begin{aligned} & \frac{1}{a+x_1} + \frac{1}{a+x_2} + \dots + \frac{1}{a+x_n} - \frac{n}{a+(x_1+x_2+\dots+x_n)/n} \\ & \approx \frac{M}{2} \sum_{i=1}^n (x_i-x_0)^2 = \frac{\sum_{i=1}^n (x_i-x_0)^2}{2(a+x_0)^3}, \quad x_i < a + 2x_0. \end{aligned} \quad (23)$$

Then

$$\begin{aligned} & a^2 \left(\frac{x_1}{a+x_1} + \frac{x_2}{a+x_2} + \dots + \frac{x_n}{a+x_n} \right) \\ & \quad - a^2 \left[\frac{x_1+x_2+\dots+x_n}{a+(x_1+x_2+\dots+x_n)/n} \right] \\ & \approx a^2 \frac{f''(x_0)}{2} \sum_{i=1}^n (x_i-x_0)^2 = a^2 \frac{\sum_{i=1}^n (x_i-x_0)^2}{2(a+x_0)^3}. \end{aligned} \quad (24)$$

From the above analysis, the errors are related to $\sum_{i=1}^n (x_i-x_0)^2$ and $\frac{a^2}{2(a+x_0)^3}$.

By the properties of $\frac{a^2}{2(a+x_0)^3}$, if the mean point x_0 increases, $\frac{a^2}{2(a+x_0)^3}$ decreases quickly accordingly. When the cluster set $\{x_i\}$

is not particularly divergent, $\sum_{i=1}^n (x_i-x_0)^2$ is much smaller than divergent cluster. So get the conclusion that if the distribution of the cluster set $\{x_i\}$ is concentrated and the mean point x_0 is large, the errors can be smaller.

Based on the above error analysis, for reducing approximate error of the DSMT approximate convex function formula, a new Evidence Clustering method is proposed as follows:

- 1) Force the mass assignments of focal elements in the evidence to two sets by the standard of $\frac{2}{n}$.
- 2) If $x_i \geq \frac{2}{n}$, x_i is forced to one set, denoted by $\{x_i^L\}$, and the sum of mass assignments for $\{x_i^L\}$ is denoted by S_L , the number of points in $\{x_i^L\}$ is denoted by n_L ; otherwise, x_i is forced to the other set, denoted by $\{x_i^S\}$.
- 3) If $x_i \in \{x_i^S\}$, pick the focal element x_i with the maximal value $x_{i \max}$; if $x_{i \max} \geq \frac{2(1-S_L)}{n-n_L}$, x_i is forced to one set $\{x_i^L\}$.
- 4) Go on the step 3), until $x_{i \max} < \frac{2(1-S_L)}{n-n_L}$; the sum of mass assignments for $\{x_i^S\}$ is denoted by S_S .

After the above cluster steps in the evidence, the mass assignments of focal elements are forced to 2 sets denoted by $\{x_i^L\}$ and $\{x_i^S\}$. Compared to $\{x_i^L\}$, the distribution of $\{x_i^S\}$ may be more concentrated. Compared to $\{x_i^S\}$, the main point of $\{x_i^L\}$ is large. So carrying out the above evidence clustering method in front of the approximate convex function formula can make errors of approximate results much smaller.

Let $m_1(X) = a$, $x_1, x_2, \dots, x_n \in \{m_2(X_j) | X_j \in G^\Theta, i \neq j \text{ and } X_i \cap X_j = \emptyset\}$, the approximate convex function formula of DSMT + PCR5 is given by

$$\begin{aligned} & \sum_{\substack{X_j \in G^\Theta \text{ and } i \neq j \\ X \cap Y = \emptyset}} \left[\frac{m_1(X_i)^2 \cdot m_2(X_j)}{m_1(X_i) + m_2(X_j)} \right] \\ & = a^2 \left[n - \frac{a \cdot n}{a + (x_1 + x_2 + \dots + x_n)/n} \right] + \Delta \end{aligned} \quad (25)$$

Finally, analysis of relationship between the approximate computation item and its errors item is as follows

$$\begin{aligned} & \sum_{\substack{X_j \in G^\Theta \text{ and } i \neq j \\ X \cap Y = \emptyset}} \left[\frac{m_1(X_i)^2 \cdot m_2(X_j)}{m_1(X_i) + m_2(X_j)} \right] \\ & = a^2 \left[n - a \left(\frac{1}{a+x_1} + \frac{1}{a+x_2} + \dots + \frac{1}{a+x_n} \right) \right] \\ & \approx a^2 \left[n - \frac{a \cdot n}{a + (x_1 + x_2 + \dots + x_n)/n} \right] \\ & \quad + a^2 \frac{\sum_{i=1}^n [x_i - (x_1 + x_2 + \dots + x_n)/n]^2}{2[a + (x_1 + x_2 + \dots + x_n)/n]^3} \\ & = a^2 \left[\frac{x_1 + x_2 + \dots + x_n}{a + (x_1 + x_2 + \dots + x_n)/n} \right] \\ & \quad + a^2 \frac{\sum_{i=1}^n [x_i - (x_1 + x_2 + \dots + x_n)/n]^2}{2[a + (x_1 + x_2 + \dots + x_n)/n]^3} \end{aligned} \quad (26)$$

After evidence clustering method, the influence of numerator $x_1 + x_2 + \dots + x_n$ to the approximate computation item and the influence of numerator $\sum_{i=1}^n (x_i - (x_1 + x_2 + \dots + x_n)/n)^2$ to errors item is much smaller than the influence of their denominators. So the approximate computation item $a^2 \left[\frac{x_1 + x_2 + \dots + x_n}{a + (x_1 + x_2 + \dots + x_n)/n} \right]$ is mainly proportional to the error item $a^2 \frac{\sum_{i=1}^n [x_i - (x_1 + x_2 + \dots + x_n)/n]^2}{2[a + (x_1 + x_2 + \dots + x_n)/n]^3}$. By the properties of convex function, all the errors items of focal elements are negative. Based on the above analysis, the normalization

method is applied as the final step of the proposed method for errors redistribution.

3.2. The process of evidence clustering DSMT approximate reasoning method

Based on the mathematical analysis of PCR5 in Section 3.1, the process of Evidence Clustering DSMT Approximate Reasoning Method is proposed as follows

Definition 1. Assuming the existence of a cluster set $\{x\}$, the definition of the total number of $\{x\}$ is $Num(\{x\}) = number\{x\}$, the sum of each point in $\{x\}$ is $Sum(\{x\}) = \sum\{x\}$ and the mean point of $\{x\}$ is $Mean(\{x\}) = \frac{\sum\{x\}}{number\{x\}}$.

1) Carry out Clustering Evidence method proposed in Section 3.1. Force mass assignments of focal elements in each evidence to two cluster sets. Giving an example of two evidences, denoted by $x = \{x_i\}$, $y = \{y_i\}$, $i = 1, 2, \dots, n$. Then force each evidence to two cluster sets, respectively denoted by $x^L = \{x_i^L\}$, $x^S = \{x_i^S\}$, $x^L \cup x^S = x$, $y_i^L = \{y_i^L\}$, $y^S = \{y_i^S\}$, $y^L \cup y^S = y$.

2) Un-normalized approximate fusion results are obtained by the approximate convex formula as follows

$$m_{CONVEXi} = \begin{cases} x_i \cdot y_i + \frac{x_i^2 Sum(\{y_i^L\}/y_i)}{x_i + Mean(\{y_i^L\}/y_i)} + \frac{x_i^2 Sum(\{y_i^S\})}{x_i + Mean(\{y_i^S\})} \\ + \frac{y_i^2 Sum(\{x_i^L\}/x_i)}{y_i + Mean(\{x_i^L\}/x_i)} + \frac{y_i^2 Sum(\{x_i^S\})}{y_i + Mean(\{x_i^S\})}, \\ y_i \in \{y_i^L\}, x_i \in \{x_i^L\} \\ x_i \cdot y_i + \frac{x_i^2 Sum(\{y_i^L\})}{x_i + Mean(\{y_i^L\})} + \frac{x_i^2 Sum(\{y_i^S\}/y_i)}{x_i + Mean(\{y_i^S\}/y_i)} \\ + \frac{y_i^2 Sum(\{x_i^L\}/x_i)}{y_i + Mean(\{x_i^L\}/x_i)} + \frac{y_i^2 Sum(\{x_i^S\})}{y_i + Mean(\{x_i^S\})}, \\ y_i \in \{y_i^L\}, x_i \in \{x_i^S\} \\ x_i \cdot y_i + \frac{x_i^2 Sum(\{y_i^L\})}{x_i + Mean(\{y_i^L\})} + \frac{x_i^2 Sum(\{y_i^S\}/y_i)}{x_i + Mean(\{y_i^S\}/y_i)} \\ + \frac{y_i^2 Sum(\{x_i^L\})}{y_i + Mean(\{x_i^L\})} + \frac{y_i^2 Sum(\{x_i^S\}/x_i)}{y_i + Mean(\{x_i^S\}/x_i)}, \\ y_i \in \{y_i^S\}, x_i \in \{x_i^S\} \\ x_i \cdot y_i + \frac{x_i^2 Sum(\{y_i^L\}/y_i)}{x_i + Mean(\{y_i^L\}/y_i)} + \frac{x_i^2 Sum(\{y_i^S\})}{x_i + Mean(\{y_i^S\})} \\ + \frac{y_i^2 Sum(\{x_i^L\})}{y_i + Mean(\{x_i^L\})} + \frac{y_i^2 Sum(\{x_i^S\}/x_i)}{y_i + Mean(\{x_i^S\}/x_i)}, \\ y_i \in \{y_i^S\}, x_i \in \{x_i^L\} \end{cases} \quad (27)$$

3) Final approximate fusion results are obtained by the normalization step as follows

$$m_{GH} = \left(1 - \sum_{\substack{X=x_i \cap y_j \\ X \in G^\ominus}} m(X)\right) \frac{m_{CONVEXi}}{\sum_{i=1}^n m_{CONVEXi}} \quad (28)$$

3.3. Analysis of computation complexity

If there are 2 evidence sources, assume that all possible unions of the elements in the discernment framework are empty for convenient computation of complexity. All singleton focal elements and multiple intersection focal elements have mass assignments in hyper-power sets of 2 evidences, denoted by $G^\ominus = \{\theta_1, \theta_2, \dots, \theta_n, \theta_a \cap \dots \cap \theta_b, \dots, \theta_c \cap \theta_d \cap \dots \cap \theta_e\}$, $\{a, b, c, d, e\} \in [1, \dots, n]$, n denotes the number of singleton focal elements, c denotes the number of multiple focal elements. First, the computation complexity of 2 evidence sources fusion based on DSMT + PCR5 is analyzed. Then, the computation complexity of the same 2 evidence sources fusion

problem based on the method in this paper is also analyzed. Computation complexity comparison of two methods can be obtained from the analysis.

Assume that the computation complexity of one time multiplication is denoted by K , the computation complexity of one time addition is denoted by Σ , the computation complexity of one time division is denoted by ψ and the computation complexity of one time subtraction is denoted by B . The computation complexity of the processing procedure of 2 evidence sources fusion based on DSMT + PCR5, denoted by $o_{DSMT+PCR5}$, is given as follows

$$\begin{aligned} o_{DSMT+PCR5}[n+c] &= [K + (4K + 2\psi + 4\Sigma)(n+c-1)](n+c) \\ &\quad - x(2K + 2\psi + 4\Sigma) + y\Sigma \\ &= (4n+4c-3)(n+c)K \\ &\quad + (2n+2c-2)(n+c)\psi \\ &\quad + (4n+4c-4)(n+c)\Sigma \\ &\quad - x(2K + 2\psi + 4\Sigma) + y\Sigma \end{aligned} \quad (29)$$

where x denotes the number of multiple focal elements in the results, y denotes the number of the same multiple focal elements generated in the procedure of mass assignments combination product.

The computation complexity of the same problem based on the method in this paper, denoted by $o_{GH}[n+c]$, is given as follows

$$\begin{aligned} o_{GH}[n+c] &= (n+c)K + 2(n+c)[2(3K + \Sigma + \psi) + \Sigma] \\ &\quad + \Sigma + n\psi + 2(n+c)B - x(2K + 2\psi + 4\Sigma) + y\Sigma \\ &= 13(n+c)K + [4(n+c) + 1]\Sigma + (5n+4c)\psi \\ &\quad + 2(n+c)B - x(2K + 2\psi + 4\Sigma) + y\Sigma \end{aligned} \quad (30)$$

where x denotes the number of multiple focal elements in the results, y denotes the number of the same multiple focal elements generated in the procedure of mass assignments combination product.

Computation complexity comparison of two methods obtained from Equation (29) and Equation (30) shows that the computation complexity of DSMT is almost proportion to $(n+c)^2$ and the computation complexity of the method in this paper is almost proportion to $(n+c)$. Analysis of computation complexity shows that the computational complexity of the proposed method increases linearly instead of exponentially when the elements number of the discernment framework increases linearly.

4. Simulation experiments

4.1. Simple cases of cluster sets in each evidence

Example 1. If there are 2 evidence sources, assume that only singleton focal elements have mass assignments in hyper-power sets, denoted by $G_k^\ominus = \{\theta_1, \theta_2, \dots, \theta_7\}$, $k = 1$ or 2. The mass assignments in each evidence are $x = \{0.1, 0.1, 0.05, 0.3, 0.2, 0.2, 0.05\}$, $y = \{0.2, 0.05, 0.05, 0.2, 0.15, 0.3, 0.05\}$, the process of the method is given as follows:

1) The mass assignments of focal elements in each evidence are clustered to two sets, denoted by $x = \{x_1, x_2, x_3, x_5, x_6, x_7\} \cup \{x_4\}$, $y = \{y_1, y_2, y_3, y_4, y_5, y_7\} \cup \{y_6\}$.

2) Un-normalized approximate fusion results are calculated by Equation (27)

$$\begin{aligned} m_{CONVEX1} &= x_1 \cdot y_1 + \frac{y^2 \cdot Sum(\{x_2, x_3, x_5, x_6, x_7\})}{y_1 + Mean(\{x_2, x_3, x_5, x_6, x_7\})} + \frac{y_1^2 \cdot x_4}{y_1 + x_4} \\ &\quad + \frac{x_1^2 \cdot Sum(\{y_2, y_3, y_4, y_5, y_7\})}{x_1 + Mean(\{y_2, y_3, y_4, y_5, y_7\})} + \frac{x_1^2 \cdot y_6}{x_1 + y_6} \end{aligned}$$

Table 1
Euclidean similarities and average computing time comparison of the methods.

	Euclidean similarities with results of DSMT + PCR5						Average computing times (s)
	1	2	3	4	5	6	
The method in this paper	0.9939	0.9911	0.9955	0.9938	0.9965	0.9960	0.0028
The method in [32]	0.9583	0.9791	0.9588	0.9530	0.9443	0.9347	0.0105

$$m_{\text{CONVEX2}} = x_2 \cdot y_2 + \frac{y_2^2 \cdot \text{Sum}(\{x_1, x_3, x_5, x_6, x_7\})}{y_2^2 + \text{Mean}(\{x_1, x_3, x_5, x_6, x_7\})} + \frac{y_2^2 \cdot x_4}{y_2^2 + x_4} \\ + \frac{x_2^2 \cdot \text{Sum}(\{y_1, y_3, y_4, y_5, y_7\})}{x_2^2 + \text{Mean}(\{y_1, y_3, y_4, y_5, y_7\})} + \frac{x_2^2 \cdot y_6}{x_2^2 + y_6}$$

$$m_{\text{CONVEX3}} = x_3 \cdot y_3 + \frac{y_3^2 \cdot \text{Sum}(\{x_1, x_2, x_5, x_6, x_7\})}{y_3^2 + \text{Mean}(\{x_1, x_2, x_5, x_6, x_7\})} + \frac{y_3^2 \cdot x_4}{y_3^2 + x_4} \\ + \frac{x_3^2 \cdot \text{Sum}(\{y_1, y_2, y_4, y_5, y_7\})}{x_3^2 + \text{Mean}(\{y_1, y_2, y_4, y_5, y_7\})} + \frac{x_3^2 \cdot y_6}{x_3^2 + y_6}$$

$$m_{\text{CONVEX4}} = x_4 \cdot y_4 + \frac{y_4^2 \cdot \text{Sum}(\{x_1, x_2, x_3, x_5, x_6, x_7\})}{y_4^2 + \text{Mean}(\{x_1, x_2, x_3, x_5, x_6, x_7\})} \\ + \frac{x_4^2 \cdot \text{Sum}(\{y_1, y_2, y_3, y_5, y_7\})}{x_4^2 + \text{Mean}(\{y_1, y_2, y_3, y_5, y_7\})} + \frac{x_4^2 \cdot y_6}{x_4^2 + y_6}$$

$$m_{\text{CONVEX5}} = x_5 \cdot y_5 + \frac{y_5^2 \cdot \text{Sum}(\{x_1, x_2, x_3, x_6, x_7\})}{y_5^2 + \text{Mean}(\{x_1, x_2, x_3, x_6, x_7\})} + \frac{y_5^2 \cdot x_4}{y_5^2 + x_4} \\ + \frac{x_5^2 \cdot \text{Sum}(\{y_1, y_2, y_3, y_4, y_7\})}{x_5^2 + \text{Mean}(\{y_1, y_2, y_3, y_4, y_7\})} + \frac{x_5^2 \cdot y_6}{x_5^2 + y_6}$$

$$m_{\text{CONVEX6}} = x_6 \cdot y_6 + \frac{y_6^2 \cdot \text{Sum}(\{x_1, x_2, x_3, x_5, x_7\})}{y_6^2 + \text{Mean}(\{x_1, x_2, x_3, x_5, x_7\})} + \frac{y_6^2 \cdot x_4}{y_6^2 + x_4} \\ + \frac{x_6^2 \cdot \text{Sum}(\{y_1, y_2, y_3, y_4, y_5, y_7\})}{x_6^2 + \text{Mean}(\{y_1, y_2, y_3, y_4, y_5, y_7\})}$$

$$m_{\text{CONVEX7}} = x_7 \cdot y_7 + \frac{y_7^2 \cdot \text{Sum}(\{x_1, x_2, x_3, x_5, x_6\})}{y_7^2 + \text{Mean}(\{x_1, x_2, x_3, x_5, x_6\})} + \frac{y_7^2 \cdot x_4}{y_7^2 + x_4} \\ + \frac{x_7^2 \cdot \text{Sum}(\{y_1, y_2, y_3, y_4, y_5\})}{x_7^2 + \text{Mean}(\{y_1, y_2, y_3, y_4, y_5\})} + \frac{x_7^2 \cdot y_6}{x_7^2 + y_6}$$

Get $m_{\text{CONVEX}} = \{0.1588, 0.0558, 0.0273, 0.3108, 0.1926, 0.3108, 0.0273\}$.

3) The final approximate results are obtained by Equation (28)

$$m_{\text{GH}} = \{0.1465, 0.0515, 0.0252, 0.2869, 0.1778, 0.2869, 0.0252\}.$$

4) The results of DSMT + PCR5 and the method in [32] are also calculated for the comparison with the method in this paper.

The results of DSMT + PCR5

$$m_{\text{DSMT+PCR5}} = \{0.1435, 0.0488, 0.0237, 0.2922, 0.1751, 0.2929, 0.0237\}.$$

The results of the method in [32]

$$m_{\text{XDL}} = \{0.1536, 0.0605, 0.0253, 0.2980, 0.1670, 0.2738, 0.0217\}.$$

The Euclidean similarity between results m_{GH} obtained by the method in this paper and the results $m_{\text{DSMT+PCR5}}$ of DSMT + PCR5, is obtained by Equation (7)

$$E_{\text{GH}} = 0.9932.$$

In the same way, the Euclidean similarity between the results m_{XDL} of the method in [32] and the results $m_{\text{DSMT+PCR5}}$ of DSMT + PCR5 is obtained by Equation (7)

$$E_{\text{XDL}} = 0.9812.$$

From the above results of this example, the results obtained by the method proposed in this paper have higher Euclidean Similarity with DSMT + PCR5 than the existing DSMT approximate reasoning method in [32]. The Euclidean Similarity which remains over 99% shows that the method proposed in this paper has high accuracy and practical meaning.

Example 2. If there are the same 2 evidence sources in Example 1, the hyper-power sets are denoted by $G_k^\Theta = \{\theta_1, \theta_2, \dots, \theta_7\}$, $k = 1$ or 2. The mass assignments in each evidence are $x = \{0.1, 0.1, 0.05, 0.3, 0.2, 0.2, 0.05\}$, $y = \{0.2, 0.05, 0.05, 0.2, 0.15, 0.3, 0.05\}$. The mass assignments in the second evidence source are unchanged, denoted by $y = \{0.2, 0.05, 0.05, 0.2, 0.15, 0.3, 0.05\}$, and the mass assignments in the first evidence change and the sequence of the mass belief of each focal element moves one position backward at one time to procedure 6 new evidences, such as:

$$x_1 = \{0.1, 0.1, 0.05, 0.3, 0.2, 0.2, 0.05\},$$

$$x_2 = \{0.05, 0.1, 0.1, 0.05, 0.3, 0.2, 0.2\},$$

$$x_3 = \{0.2, 0.05, 0.1, 0.1, 0.05, 0.3, 0.2\},$$

$$x_4 = \{0.2, 0.2, 0.05, 0.1, 0.1, 0.05, 0.3\},$$

$$x_5 = \{0.3, 0.2, 0.2, 0.05, 0.1, 0.1, 0.05\},$$

$$x_6 = \{0.05, 0.3, 0.2, 0.2, 0.05, 0.1, 0.1\}.$$

Each new evidence and the evidence y are calculated to obtain the fusion results by DSMT + PCR5 and the approximate fusion results by the method in this paper and the method in [32]. Then Euclidean similarities of the approximate results of different methods with the results of DSMT + PCR5 are obtained by Equation (7) and the average computing times of the methods are also taken record as Table 1. (In this paper, all the simulation experiments are implemented by Matlab simulation in the hardware condition of Pentium(R) Dual-Core CPU E5300 2.6 GHz 2.59 GHz, memory 1.99 GB.)

As shown in Table 1, in the simple cases of cluster sets in each evidence, the accuracies of the method in this paper all remain over 99% and much higher than the method in [32]. Average computing time of the method in this paper is much lower than the method in [32]. At the same time, the accuracies of the method in this paper in different evidence cases change little, which prove that the method in this paper has higher performance stability.

4.2. Complex cases of cluster sets in each evidence

Example 3. If there are 2 evidence sources, assume that only singleton focal elements have mass assignments in hyper-power sets, denoted by $G_k^\Theta = \{\theta_1, \theta_2, \dots, \theta_{12}\}$, $k = 1$ or 2. The mass assignments in each evidence are $x = \{0.3, 0.35, 0.05, 0.05, 0.04, 0.06, 0.02, 0.01, 0.02, 0.01, 0.04, 0.05\}$, $y = \{0.2, 0.05, 0.04, 0.21, 0.15,$

Table 2
Euclidean similarities and average computing time comparison of the methods.

	Euclidean similarities with results of DSMT + PCR5											Average computing times (s)
	1	2	3	4	5	6	7	8	9	10	11	
The method in this paper	0.9987	0.9983	0.9982	0.9979	0.9981	0.9985	0.9985	0.9983	0.9984	0.9983	0.9983	0.0038
The method in [32]	0.8795	0.9330	0.9514	0.9484	0.8112	0.8636	0.8253	0.8342	0.8331	0.8189	0.8483	0.0186

0.25, 0.05, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01}, the method process is given as follows:

1) The mass assignments of focal elements in each evidence are clustered to two sets, denoted by $x = \{x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}\} \cup \{x_1, x_2\}$, $y = \{y_2, y_3, y_7, y_8, y_9, y_{10}, y_{11}, y_{12}\} \cup \{y_1, y_4, y_5, y_6\}$.

2) Un-normalized approximate fusion results are calculated by Equation (27)

$$m_{CONVEX} = \{0.3069, 0.2559, 0.0247, 0.1309, 0.0834, 0.1662, 0.0177, 0.0019, 0.0041, 0.0019, 0.0110, 0.0153\}.$$

3) The final approximate results are obtained by Equation (28)

$$m_{GH} = \{0.3009, 0.2509, 0.0242, 0.1283, 0.0818, 0.1630, 0.0174, 0.0019, 0.0041, 0.0019, 0.0108, 0.0150\}.$$

4) The results of DSMT + PCR5 and the method in [32] are also calculated for the comparison with the method in this paper.

The results of DSMT + PCR5

$$m_{DSMT+PCR5} = \{0.3019, 0.2524, 0.0235, 0.1282, 0.0811, 0.1635, 0.0169, 0.0018, 0.0039, 0.0018, 0.0104, 0.0146\}.$$

The results of the method in [32]

$$m_{XDL} = \{0.3710, 0.1834, 0.0276, 0.1269, 0.0828, 0.1651, 0.0003, 0.0000, 0.0001, 0.0001, 0.0002, 0.0002\}.$$

The Euclidean similarity between the results m_{GH} and $m_{DSMT+PCR5}$ is calculated, denoted by $E_{GH} = 0.9984$ and computing time is 0.0035 s.

In the same way, the Euclidean similarity between the results m_{XDL} and $m_{DSMT+PCR5}$ is calculated, denoted by $E_{XDL} = 0.9287$ and computing time is 0.0185 s.

As shown in the above experiment results, the results obtained by the method proposed in this paper have higher Euclidean similarity and lower computation complexity than the existing method in [32]. The Euclidean similarity which remains over 99% shows that the method proposed in this paper has higher accuracy and practical meaning.

Example 4. If there are the same 2 evidence sources in Example 3, the hyper-power sets are denoted by $G_k^\Theta = \{\theta_1, \theta_2, \dots, \theta_{12}\}$, $k = 1$ or 2. The mass assignments in the second evidence source are unchanged, denoted by $y = \{0.2, 0.05, 0.04, 0.21, 0.15, 0.25, 0.05, 0.01, 0.01, 0.01, 0.01, 0.01\}$. The mass assignments in the first source change and the sequence of the mass beliefs of focal elements moves one position backward at one time to procedure 11 new evidences.

Each new evidence and the second evidence y are calculated to get the fusion results of DSMT + PCR5 and the approximate results by the method in this paper and the method in [32]. Then Euclidean similarities of the approximate results of different methods with the results of DSMT + PCR5 and the average computing times of the methods are taken record as Table 2.

As shown in Table 2, under complex cases of cluster sets in each evidence, the accuracy of the method in this paper also remains over 99% and much higher than the method in [32]. Average computing time of the method in this paper is lower than the method in [32]. At the same time, the accuracies of the method in this pa-

Table 3
The mass assignments of highly conflict evidence sources.

Conflict evidence sources	a	b	c	d
S_1	$x - \varepsilon$	ε	$1 - x - \varepsilon$	ε
S_2	ε	$y - \varepsilon$	ε	$1 - y - \varepsilon$

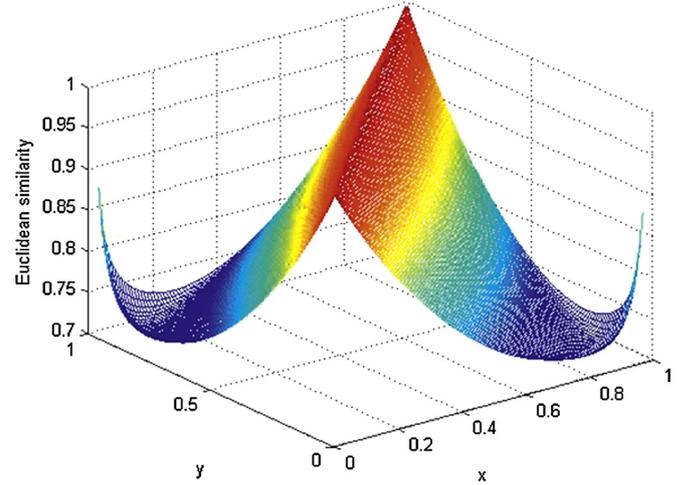


Fig. 1. Euclidean similarity of the method in [32] with DSMT + PCR5.

per in different evidence cases change little, which prove that the method in this paper has higher performance stability.

4.3. Cases of highly conflict evidence sources

Example 5. In order to verify that information fusion of highly conflict evidence sources can be effectively solved by the method in this paper, assume there are two highly conflict evidence sources with the hyper-power set denoted by $D^\Theta = \{a, b, c, d\}$ and the mass assignments of two evidence sources are shown as Table 3.

Let $\varepsilon = 0.01$, $x, y \in [0.02, 0.98]$. The fusion results are obtained by different methods when x, y is increasing from 0.02 to 0.98 by 0.01 step at the same time. Euclidean similarity of the method in [32] with DSMT + PCR5 is shown in Fig. 1. Euclidean similarity of the method in this paper with DSMT + PCR5 is shown in Fig. 2.

The average Euclidean similarity of the method in this paper is 0.9873 and the average Euclidean similarity of the method in [32] is 0.8513. It's shown that information fusion problem of highly conflict evidence sources can be effectively solved by the method proposed in this paper.

4.4. Convergence analysis

Example 6. If there are 2 evidence sources, assume that only singleton focal elements have mass assignments in hyper-power sets, denoted by $G_k^\Theta = \{\theta_1, \theta_2, \dots, \theta_{12}\}$, $k = 1$ or 2. The mass assignments in each evidence are $x = \{0.1, 0.01, 0.02, 0.25, 0.15, 0.05, 0.1, 0.1, 0.1, 0.05, 0.05, 0.02\}$, $y = \{0.5, 0.35, 0.02, 0.02, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01, 0.04\}$. First, fusion results of two evidences x and y are obtained by different fusion methods. Then, fusion of the prior fusion results with y is carried out repeatedly.

Table 4
Fusion results comparison of the methods for convergence analysis.

		Fusion results
1	DSmT + PCR5	0.4287, 0.2600, 0.0058, 0.1163, 0.0539, 0.0109, 0.0297, 0.0297, 0.0297, 0.0109, 0.0109, 0.0136
	The method in [32]	0.4067, 0.0994, 0.0212, 0.1888, 0.0798, 0.0385, 0.0027, 0.0027, 0.0027, 0.0010, 0.0010, 0.0019
	The method in this paper	0.4285, 0.2610, 0.0060, 0.1151, 0.0535, 0.0110, 0.0296, 0.0296, 0.0296, 0.0110, 0.0110, 0.0141
2	DSmT + PCR5	0.5829, 0.3324, 0.0031, 0.0388, 0.0120, 0.0017, 0.0053, 0.0053, 0.0053, 0.0017, 0.0017, 0.0099
	The method in [32]	0.5687, 0.1686, 0.0036, 0.0642, 0.0202, 0.0127, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000
	The method in this paper	0.5705, 0.3459, 0.0031, 0.0371, 0.0116, 0.0018, 0.0052, 0.0052, 0.0052, 0.0018, 0.0018, 0.0109
3	DSmT + PCR5	0.6439, 0.3342, 0.0018, 0.0084, 0.0016, 0.0006, 0.0008, 0.0008, 0.0008, 0.0006, 0.0006, 0.0060
	The method in [32]	0.6171, 0.1968, 0.0009, 0.0109, 0.0027, 0.0021, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000
	The method in this paper	0.6118, 0.3655, 0.0021, 0.0078, 0.0015, 0.0006, 0.0008, 0.0008, 0.0008, 0.0006, 0.0006, 0.0069
4	DSmT + PCR5	0.6726, 0.3181, 0.0012, 0.0017, 0.0004, 0.0004, 0.0004, 0.0004, 0.0004, 0.0004, 0.0004, 0.0040
	The method in [32]	0.6152, 0.2047, 0.0008, 0.0018, 0.0004, 0.0003, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000
	The method in this paper	0.6264, 0.3626, 0.0015, 0.0017, 0.0004, 0.0004, 0.0004, 0.0004, 0.0004, 0.0004, 0.0004, 0.0048

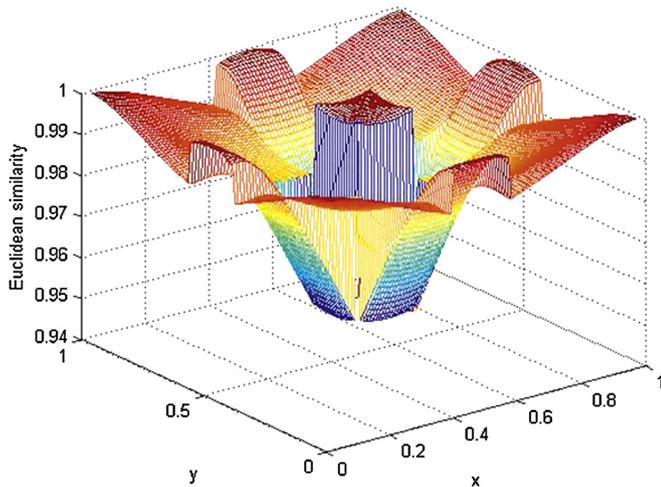


Fig. 2. Euclidean similarity of the method in this paper with DSmT + PCR5.

Through the fusion results each time, the different methods' convergences are analyzed in Table 4.

As shown in Table 4, the convergences of three methods are similar. The results of the methods can converge to the main focal elements after 3 times of evidences fusion. However, the results of the method in this paper have higher Euclidean similarity with DSmT + PCR5, and lose less information than the method in [32] each time.

4.5. Monte Carlo simulations in the case of non-empty multiple focal elements

If there are 2 evidence sources, assume that singleton focal elements and multiple focal elements have mass assignments in hyper-power sets, denoted by $P^\ominus = \{\theta_1, \theta_2, \dots, \theta_{20}, \theta_1 \cap \theta_5 \cap \theta_{10} \cap \theta_{20}\}$. Carry out 1000 times Monte Carlo simulation experiments. The simulation experiment process of each time consists of 3 steps. First, mass values of all focal elements of hyper-power sets in each evidence are assigned randomly. Then, the fusion results of 2 evidences are obtained by DSmT + PCR5 and the method in this paper separately. Thirdly, the Euclidean similarity of the method in this paper with DSmT + PCR5 and computing times of both methods are calculated and recorded. The Monte Carlo simulation results are shown in Fig. 3, Fig. 4 and Table 5.

As shown in Fig. 3, Fig. 4 and Table 5, in the case of non-empty multiple focal elements, the average Euclidean similarity of the method in this paper can reach 98.49% and Euclidean similarity changes little with different evidences. Computing times of the method in this paper almost reduce halfly than DSmT + PCR5.

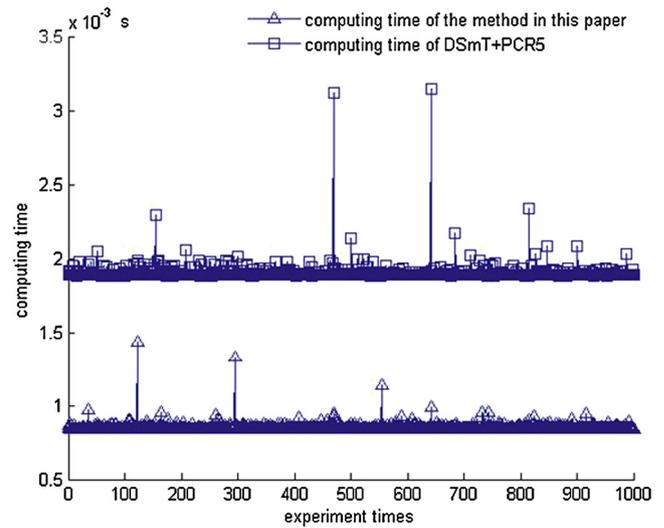


Fig. 3. Computing time comparison of the method in this paper with DSmT + PCR5.

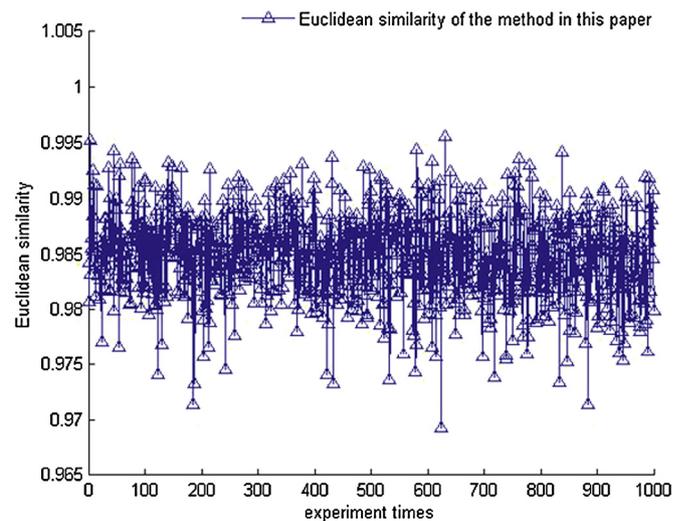


Fig. 4. Euclidean similarity of the method in this paper with DSmT + PCR5.

4.6. Monte Carlo simulations in the case of increasing focal elements number

Example 7. If there are 2 evidence sources, assume that only singleton focal elements have mass assignments in hyper-power sets, denoted by $C_k^\ominus = \{\theta_1, \theta_2, \dots, \theta_{10}\}$, $k = 1$ or 2. The focal elements of the hyper-power sets are increased 10 focal elements each time

Table 5
Fusion results comparison in the case of increasing focal elements number.

	Average Euclidean similarity	Max Euclidean similarity	Min Euclidean similarity	Average computing time (ms)	Max computing time (ms)	Min computing time (ms)
DSmT + PCR5				1.9	3.1	1.9
The method in this paper	0.9849	0.9956	0.9693	0.84911	1.4	0.83593

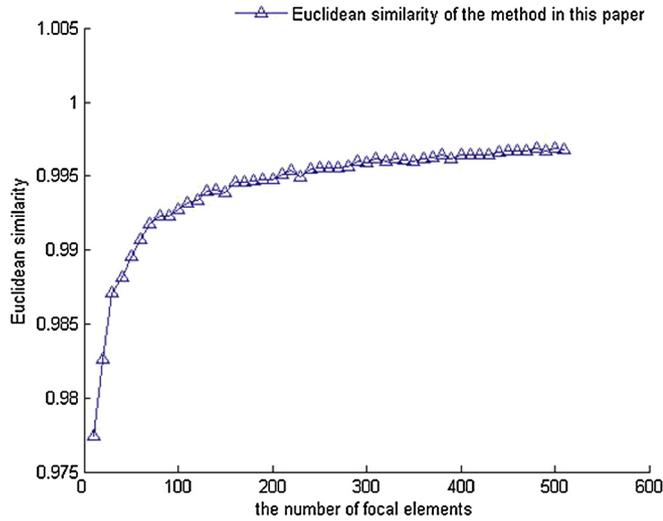


Fig. 5. Euclidean similarity of the method in this paper with DSmT + PCR5.

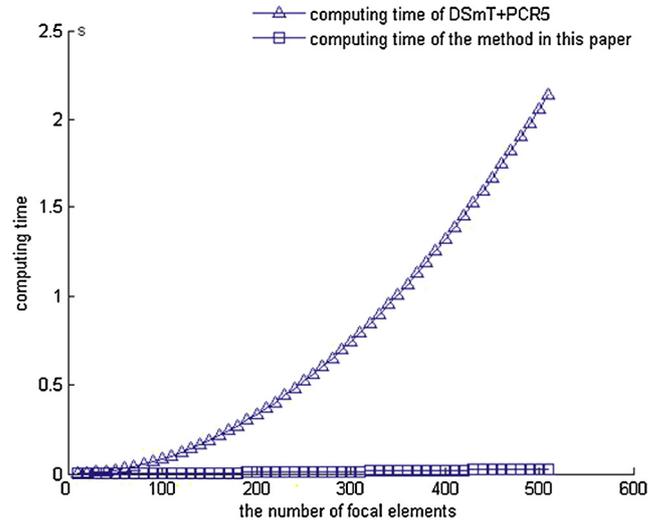


Fig. 6. Computing time comparison of the method in this paper with DSmT + PCR5.

until 500 and the mass values of every focal elements of hyper-power sets in each evidence are assigned randomly each time. 1000 times Monte Carlo simulation experiments in each hyper-power set are carried out and the average Euclidean similarities of the method in this paper with DSmT + PCR5 are calculated and recorded as shown in Fig. 5. The average computing time comparison of the method in this paper with DSmT + PCR5 in each hyper-power sets is shown in Fig. 6.

The fusion results comparison in the case of increasing focal elements number is shown in Table 6. (As the increasing number of focal elements, the mass assignment of average cluster center is decreasing sharply. For reducing computation complexity, neglect the influence of the different classification of clusters and apply the standard of $\frac{2}{n}$ as one step cluster method, n denotes the focal elements number of hyper-power sets.)

As shown in Fig. 5, Fig. 6 and Table 6,

1) In the case of increasing focal elements number, computing time of the method in this paper decreases significantly, and the computation complexity of the method almost appears linear growth instead of exponential growth, which proves that the method in this paper has a high application in the case of complex fusion problems with big number of focal elements.

2) The accuracy of the method in this paper is increasing with the growth of focal elements number of hyper-power sets as the

errors item becomes much smaller. The minimum average Euclidean Similarity is 0.9974 in the case of the minimum number of hyper-power sets. When the number of hyper-power sets increases over 50, the average Euclidean similarity exceeds 99%, which proves that the method in this paper can effectively support correct and quick decision in the case of large data.

5. Conclusions

A new Evidence Clustering DSmT Approximate Reasoning Method is proposed based on convex function analysis in this paper. This new method consists of three steps. First, the belief masses of focal elements in each evidence are clustered by the Evidence Clustering method. Second, the un-normalized approximate fusion results are obtained using the DSmT approximate convex function formula, which is acquired based on the mathematical analysis of PCR5 rule in DSmT. Finally, the normalization step is applied. The method reduces computation complexity of PCR5 based information fusion rule within DSmT framework significantly which constrains the wide application and development of DSmT and remains high accuracy. Simulation results in different cases show that the method in this paper can process evidences fusion problems effectively and efficiently, especially, in the case of large data

Table 6
Fusion results comparison in the case of increasing focal elements number.

	The focal elements number of hyper-power sets is increasing from 10 to 510																																																				
Average computing times of DSmT + PCR5 (s)	0.0008	0.0032	0.0073	0.0132	0.0208	0.0297	0.0406	0.0530	0.0670	0.0829	0.1017	0.1219	0.1405	0.1636	0.1867	0.2133	0.2431	0.2685	0.3009	0.3319	0.3698	0.4013	0.4427	0.4816	0.5212	0.5588	0.6011	0.6462	0.6959	0.7413	0.7909	0.8425	0.8962	0.9505	1.0082	1.0659	1.1263	1.1882	1.2516	1.3166	1.3828	1.4509	1.5225	1.5926	1.6660	1.7409	1.8174	1.8946	1.9765	2.0574	2.1397		
Average computing times of the method in this paper (s)	0.0003	0.0006	0.0010	0.0013	0.0017	0.0020	0.0024	0.0028	0.0032	0.0036	0.0040	0.0044	0.0048	0.0053	0.0057	0.0062	0.0067	0.0071	0.0076	0.0081	0.0087	0.0092	0.0098	0.0103	0.0108	0.0114	0.0118	0.0124	0.0130	0.0136	0.0142	0.0148	0.0154	0.0160	0.0167	0.0173	0.0180	0.0186	0.0193	0.0200	0.0207	0.0214	0.0222	0.0229	0.0236	0.0244	0.0251	0.0259	0.0265	0.0275	0.0282		
Average Euclidean similarities with results of DSmT + PCR5	0.9774	0.9826	0.9871	0.9881	0.9895	0.9907	0.9917	0.9922	0.9922	0.9927	0.9931	0.9933	0.9939	0.9940	0.9938	0.9945	0.9945	0.9946	0.9947	0.9947	0.9950	0.9953	0.9949	0.9954	0.9955	0.9955	0.9955	0.9956	0.9959	0.9958	0.9961	0.9959	0.9961	0.9960	0.9959	0.9961	0.9962	0.9964	0.9961	0.9964	0.9964	0.9964	0.9964	0.9965	0.9966	0.9966	0.9966	0.9966	0.9966	0.9968	0.9966	0.9968	0.9967

and complex fusion problems, the method can get highly accurate results and need low computation complexity.

However, our proposed method is not associative and the combination order really has influence on the results, which is the same problem to PCR5. This problem is really needed to be resolved especially for the efficient fusion of multiple (more than 2) sources of evidences and more research would be carried out to resolve this associative question of our method in the near future. Our method would also be applied to the realistic and applied data or database in future.

Acknowledgments

The authors are grateful to Prof. Xinde Li, Prof. De-cun Zhang and Prof. Ming-jiu Gai for discussions. We extremely appreciate Editor-in-chief Ercan Engin Kuruoglu and Journal Manager Allwyn Richard for their efforts to help us improve and present this paper. We are also very grateful to all the anonymous reviewers for their critical and constructive review of the manuscript. This study was co-supported by the National Natural Science Foundation of China (Nos. 61102166 and 61471379) and Program for New Century Excellent Talents in University (Nos. NCET-11-0872).

Appendix A. Supplementary material

Supplementary material related to this article can be found online at <http://dx.doi.org/10.1016/j.dsp.2015.06.009>.

References

- [1] Alan J. Terry, Munir Zaman, John Illingworth, Sensor fusion by a novel algorithm for time delay estimation, *Digit. Signal Process.* 22 (3) (2012) 439–452.
- [2] Yilmaz Kalkan, Buyurman Baykal, Multiple target localization & data association for frequency-only widely separated MIMO radar, *Digit. Signal Process.* 25 (2014) 51–61.
- [3] Donka Angelova, Lyudmila Mihaylova, Joint target tracking and classification with particle filtering and mixture Kalman filtering using kinematic radar information, *Digit. Signal Process.* 16 (2) (2006) 180–204.
- [4] Lauro Snidaro, Ingrid Visentini, Karna Bryan, Fusing uncertain knowledge and evidence for maritime situational awareness via Markov Logic Networks, *Inf. Fusion* 21 (2015) 159–172.
- [5] Yi Yang, Deqiang Han, Chongzhao Han, Discounted combination of unreliable evidence using degree of disagreement, *Int. J. Approx. Reason.* 54 (8) (2013) 1197–1216.
- [6] Ammar Cherchar, Messaoud Thameri, Adel Belouchrani, Performance improvement of direction finding algorithms in non-homogeneous environment through data fusion, *Digit. Signal Process.* 41 (2015) 41–47.
- [7] X. Li, J. Dezert, F. Smarandache, X. Dai, Combination of qualitative information based on 2-Tuple modelings in DSMT, *J. Comput. Sci. Technol.* 24 (4) (2009) 786–798.
- [8] X. Li, J. Dezert, F. Smarandache, Fusion of imprecise qualitative information, *Appl. Intell.* 33 (3) (2010) 340–351.
- [9] F. Smarandache, J. Dezert, *Advances and Applications of DSMT for Information Fusion*, vol. 1, American Research Press, Rehoboth, USA, 2004; F. Smarandache, J. Dezert, *Advances and Applications of DSMT for Information Fusion*, vol. 2, American Research Press, Rehoboth, USA, 2006; F. Smarandache, J. Dezert, *Advances and Applications of DSMT for Information Fusion*, vol. 3, American Research Press, Rehoboth, USA, 2009; F. Smarandache, J. Dezert, *Advances and Applications of DSMT for Information Fusion*, vol. 4, American Research Press, Rehoboth, USA, 2015.
- [10] G. Shafer, *A Mathematical Theory of Evidence*, Princeton Univ. Press, Princeton, NJ, 1976.
- [11] G. Shafer, Perspectives on the theory and practice of belief functions, *Int. J. Approx. Reason.* 4 (1990) 323–362.
- [12] R. Singh, M. Vatsa, A. Noore, Integrated multilevel image fusion and match score fusion of visible and infrared face images for robust face recognition, *Pattern Recognit.* 41 (3) (2008) 880–893.
- [13] X. Li, X. Huang, X. Dai, Z. Meng, Study on environment perception of mobile robots using DSMT-based fusion machine, *J. Huazhong Univ. Sci. Technol.* 12 (37) (2009) 64–67, Chinese.
- [14] Z. Liu, J. Dezert, Q. Pan, Grégoire Mercier, Combination of sources of evidence with different discounting factors based on a new dissimilarity measure, *Decis. Support Syst.* 52 (1) (2011) 133–141.
- [15] T. Denoeux, N. El Zoghby, V. Cherfaoui, A. Jouglet, Optimal object association in the Dempster–Shafer framework, *IEEE Trans. Cybern.* 44 (22) (2014) 2521–2531.
- [16] F. Faux, F. Luthon, Theory of evidence for face detection and tracking, *Int. J. Approx. Reason.* 53 (5) (2012) 728–746.
- [17] Z. Liu, J. Dezert, G. Mercier, Q. Pan, Dynamic evidential reasoning for change detection in remote sensing images, *IEEE Trans. Geosci. Remote Sens.* 50 (5) (2012) 1955–1967.
- [18] X. Li, X. Huang, J. Dezert, et al., A successful application of DSMT in sonar grid map building and comparison with DST-based approach, *Int. J. Innov. Comput. Inf. Control* 3 (3) (2007) 539–551.
- [19] Z. Liu, Q. Pan, J. Dezert, G. Mercier, Credal classification rule for uncertain data based on belief functions, *Pattern Recognit.* 47 (7) (2014) 2532–2541.
- [20] Z. Liu, Q. Pan, G. Mercier, J. Dezert, A new incomplete pattern classification method based on evidential reasoning, *IEEE Trans. Cybern.* 45 (4) (2015) 635–646.
- [21] C. Lian, S. Ruan, T. Denoeux, An evidential classifier based on feature selection and two-step classification strategy, *Pattern Recognit.* 48 (2015) 2318–2327.
- [22] Z. Liu, Q. Pan, J. Dezert, G. Mercier, Credal c-means clustering method based on belief functions, *Knowl.-Based Syst.* 74 (2015) 119–132.
- [23] T. Denoeux, Maximum likelihood estimation from uncertain data in the belief function framework, *IEEE Trans. Knowl. Data Eng.* 25 (1) (2013) 119–130.
- [24] U. Bergsten, J. Schubert, Dempster's rule for evidence ordered in a complete directed acyclic graph, *Int. J. Approx. Reason.* 9 (1) (1993) 37–73.
- [25] B. Tessem, Approximations for efficient computation in the theory of evidence, *Artif. Intell.* 61 (1993) 315–329.
- [26] T. Denoeux, A.B. Yaghlane, Approximating the combination of belief functions using the fast Möbius transform in a coarsened frame, *Int. J. Approx. Reason.* 31 (2002) 77–101.
- [27] P. Djiknavorian, D. Grenier, Reducing DSMT hybrid rule complexity through optimization of the calculation algorithm, in: *Advances and Application of DSMT for Information Fusion*, vol. 2, Amer. Res. Press, Rehoboth, 2006, pp. 365–429, chap. 15.
- [28] A. Martin, Implementing general belief function framework with a practical codification for low complexity, in: *Advances and Application of DSMT for Information Fusion*, vol. 3, Amer. Res. Press, Rehoboth, 2009, pp. 217–273, chap. 7.
- [29] N. Abbas, Y. Chibani, H. Nemmour, Handwritten digit recognition based on a DSMT-SVM parallel combination, in: *Proc. 13th International Conference on Frontiers in Handwriting Recognition, ICFHR, Bari, Italy, September 18–20, 2012*, pp. 241–246.
- [30] N. Abbas, Y. Chibani, Z. Belhadi, M. Hedi, A DSMT based combination scheme for multi-class classification, in: *Proc. 16th International Conference on Information Fusion, ICIF, Istanbul, Turkey, July 9–12, 2013*, pp. 1950–1957.
- [31] X. Li, J. Dezert, F. Smarandache, X. Huang, Evidence supporting measure of similarity for reducing the complexity in information fusion, *Inf. Sci.* 181 (2011) 1818–1835.
- [32] X. Li, X. Wu, J. Sun, Z. Meng, An approximate reasoning method in Dezert–Smarandache theory, *J. Electron. (China)* 26 (6) (2009) 738–745.
- [33] X. Li, J. Dezert, X. Huang, Z. Meng, X. Wu, A fast approximate reasoning method in hierarchical DSMT(A), *Acta Electron. Sin.* 11 (2010) 2566–2572, (in Chinese).
- [34] X. Li, W. Yang, X. Wu, J. Dezert, A fast approximate reasoning method in hierarchical DSMT(B), *Acta Electron. Sin.* 3 (A) (2011) 31–36, (in Chinese).
- [35] F. Smarandache, Algebraic Generalization of Venn Diagram, *Multispace and Multistruature*, North-European Scientific Publishers, Finland, 2010, pp. 84–86.
- [36] J. Dezert, F. Smarandache, Presentation of DSMT, in: *Advances and Application of DSMT for Information Fusion*, vol. 1, Amer. Res. Press, Rehoboth, 2004, pp. 3–31, chap. 1.
- [37] J. Dezert, F. Smarandache, DSMT: a new paradigm shift for information fusion, in: *Cogis06 International Conference, Paris, France, 2006*, pp. 15–17.
- [38] J. Dezert, F. Smarandache, Partial ordering on hyper-power sets, in: *Advances and Application of DSMT for Information Fusion*, vol. 1, Amer. Res. Press, Rehoboth, 2004, pp. 49–60, chap. 3.
- [39] F. Smarandache, J. Dezert, Information fusion based on new proportional conflict redistribution rules, in: *Proceedings of Fusion 2005 Conf., Philadelphia, July 26–29, 2005*, pp. 907–914.
- [40] J. Dezert, F. Smarandache, An introduction to DSMT, in: *Advances and Application of DSMT for Information Fusion*, vol. 3, Amer. Res. Press, Rehoboth, 2009, pp. 3–68, chap. 1.
- [41] Anne-Laure Jousselme, et al., A new distance between two bodies of evidence, *Inf. Fusion* 2 (2) (2001) 91–101.

Qiang Guo received his B.E. degree in information engineering and his M.S. degree in information and communication engineering from Naval Aeronautical and Astronautical University, Yantai, Shandong Province, China, in 2008 and 2010, respectively. Currently, he is pursuing a Ph.D. degree in the Research Institute of Information Fusion of the Naval Aeronautical and Astronautical University. His research interests are information fusion, evidential reasoning, and signal processing.

You He is currently the Academician of Chinese Academy of Engineering, Head, Professor, Doctor Tutor of the Research Institute of Information Fusion of Naval Aeronautical and Astronautical University, Yantai, Shandong Province, China, as well as Part-time Professor of Tsinghua University, Beijing, China. He also serves as IET fellow, a managing director of Chinese Aviation Society, a fellow member of Chinese Institute of Electronics and a member in a large range of professional and social organizations. Moreover, he was the founder of the National Conference on Information Fusion, and served as General Chairman in 2009. He received his B.S. degree in electronic engineering and the M.S. degree in computer science from Huazhong University of Science and Technology at Wuhan, P.R. China, in 1982 and 1988, respectively. From October 1991 to November 1992, he was with the institute of Communication at Technical University of Braunschweig, Germany, where he carried out automatic radar detection theory and CFAR algorithm studies. From November 1992 to December 1994, he was an Associate Professor in the Department of Electronic Engineering at Technical University of Yantai, China. Then, he received the Ph.D. degree in communication and information systems from Tsinghua University, in 1997. His research interests are information fusion, signal processing, and so on.

Xin Guan is Professor and Master tutor in the Department of Electronics and Information Engineering of Naval Aeronautical and Astronautical University, Yantai, Shandong Province, China. She also serves as a vice chairman of the youth working committee in Chinese Aviation Society and a senior member of Chinese Aviation Society. She received her B.E. degree in communication engineering from Liaoning University, Shenyang,

Liaoning Province, China, in 1999, her M.S. degree in signal and information processing and her Ph.D. degree in communication and information system from Naval Aeronautical and Astronautical University, Yantai, Shandong Province, China, in 2002 and 2006, respectively. Her research interests are evidence theory, signal processing and target recognition.

Li Deng received his B.E. degree and his M.S. degree in automatic equipments and ammunition engineering and his Ph.D. degree in system and application engineering from Naval Aeronautical and Astronautical University, Yantai, Shandong Province, China, in 2007, 2010 and 2014, respectively. Currently, he is the Lecturer of the Naval Aeronautical and Astronautical University. His research interests are information processing and armament science.

Lina Pan is Professor in Department of Basic Science of the Naval Aeronautical and Astronautical University. She received her B.S. degree in solid state physics from Jilin University, Jilin, China in 1984. Her research interests are physics science, signal processing and optical sensor applications.

Tao Jian is Associate Professor in Research Institute of Information Fusion of Naval Aeronautical and Astronautical University, Yantai, Shandong Province, China. He received his B.E. degree in radar engineering, his M.S. degree and his Ph.D. degree in signal and information processing from Naval Aeronautical and Astronautical University, Yantai, Shandong Province, China, in 2003, 2006 and 2011, respectively. His research interests are radar signal and information processing.