A novel divergence measure in Dempster–Shafer evidence theory based on pignistic probability transform and its application in multi-sensor data fusion

Shijun Xu, Yi Hou, Xinpu Deng, Peibo Chen, Kewei Ouyang and Ye Zhang

Abstract
Dempster–Shafer (D–S) evidence theory is more and more extensively applied in multi-sensor data fusion. However, it is still an open issue that how to effectively combine highly conflicting evidence in D–S evidence theory. In this article, a novel divergence measure, called pignistic probability transformation divergence, is proposed to measure the difference between evidences. The proposed pignistic probability transformation divergence can reflect the interaction between single-element and multi-element subsets by introducing the pignistic probability transformation, and satisfies the properties of boundedness, non-degeneracy, and symmetry. Moreover, the pignistic probability transformation divergence can degenerate as Jensen–Shannon divergence when mass function and the probability distribution are consistent. Based on the pignistic probability transformation divergence, a new multi-sensor data fusion method is presented. The proposed method takes advantage of pignistic probability transformation divergence to measure the discrepancy between evidences in order to obtain the credibility weights, and belief entropy to measure the uncertainty of the evidences in order to obtain the information volume weights, which can fully mine the potential information between evidences. Then, the credibility weights and the information volume weights are integrated to generate an appropriate weighted average evidence before using Dempster’s combination rule. The results of two application cases illustrate that the proposed method outperforms other related methods for combining highly conflicting evidences.

Keywords
Dempster–Shafer evidence theory, divergence, multi-sensor data fusion, conflicting evidences

Date received: 24 April 2021; accepted: 16 June 2021
Handling Editor: Dr Peio Lopez Iturri

Introduction
Multi-sensor data fusion is an information modeling process in which data from multiple sensors are comprehensively analyzed to realize decision-making. As we all know, the information collected by a single sensor cannot be used to describe a certain system in multiple levels and perspectives, so the results are not convincing. In contrast, multi-sensor data fusion can process data from multiple sensors comprehensively to obtain more reliable results than a single sensor. Thus, it is widely applied in massive applications in the real world.
However, it is still an open issue that how to effectively combine the uncertain, inconsistent, or even conflicted data collected from different sensors. Many theories have been presented to solve this problem, including Dempster–Shafer (D–S) evidence theory, fuzzy set theory, rough sets, Z-numbers, D-numbers, R-number, and so on.

The essence of D–S evidence theory is a generalization of probability theory. Compared with the traditional probability theory, it can more effectively express the random uncertainty. In addition, it also provides a powerful Dempster’s combination rule which satisfies the excellent characteristics of associativity and commutativity, and can realize the fusion of evidences without prior information. Due to its flexibility and effectiveness on handling uncertainty, D–S evidence theory is widely applied in various fields of multi-sensor data fusion, such as target recognition, fault diagnosis, decision-making, and risk analysis. Whereas, Dempster’s combination rule leads to counter-intuitive conclusions when there is a conflict between evidences. Thus far, many methods have been developed in order to overcome the defect, which are mainly divided into two types. One is to modify Dempster’s combination rule. It is believed that the counter-intuitive conclusions are made as a result of some defects of Dempster’s combination rule, that is, the normalization process causes the conflicting information to be completely discarded. So, the key to modifying Dempster’s combination rule is the redistribution of conflicts. However, the modified combination rule cannot keep the excellent mathematical properties of Dempster’s combination rule, like commutativity and associativity. The other is to modify the original evidence. It is believed that Dempster’s combination rule itself is not wrong, that conflicts are caused by the unstable evidence sources. The improvement method is first to preprocess the conflicting evidence in order to obtain the weighted average evidences. Then, the weighted average evidences are fused using Dempster’s combination rule to get reasonable results. In this article, we focus on the second method. Han et al. introduce the concept of evidence support based on the Jousselme distance function and take a weighted average of all the evidences. Xiao generalizes the traditional Jousselme distance to the complex evidence distance to measure the conflicts of complex the basic probability assignment (BPA) functions, and uses it as a weighted factor to revise the original evidence. Xiao utilizes belief Jensen–Shannon (BJS) divergence to measure the distance between evidences to generate the modified evidences. Whereas, BJS divergence measures the difference between evidences by treating multi-element subsets as single-element subset, which cannot reflect the interaction between single-element and multi-element subset. Although the reinforced belief (RB) divergence proposed by Xiao overcomes the shortcomings of BJS divergence, the multi-sensor data fusion method based on the RB divergence is still some room for improvement to achieve more high belief degree for the correct target. Besides, the new divergence proposed by Wang et al. also addresses the above deficiencies, but the time complexity of the multi-sensor data fusion method based on the new divergence is high.

In this article, we propose a novel divergence measure, called pignistic probability transformation (PPT) divergence, to measure the difference between evidences. The proposed PPT divergence can reflect the interaction between single-element and multi-element subsets by introducing the PPT, and satisfies the properties of boundedness, non-degeneracy, and symmetry. Moreover, the PPT divergence can degenerate as Jensen–Shannon (JS) divergence when the mass function and the probability distribution are consistent. Based on the PPT divergence and the Deng entropy, a new multi-sensor data fusion method is presented. This method takes advantage of PPT divergence to measure the discrepancy between evidences in order to obtain the credibility weights, and the Deng entropy to measure the uncertainty of the evidences in order to obtain the information volume weights, which can fully mine the potential information between evidences. Then, the credibility weights and information volume weights are integrated to generate an appropriate weighted average evidence before using Dempster’s combination rule. Two application cases are provided to illustrate the superiority of our method.

The rest of the article is organized as follows. Section “Preliminaries” introduces some relevant basic theoretical knowledge. Section “The PPT divergence measure” proposes the PPT divergence in D–S evidence theory, and the comparative analysis is conducted. In section “The multi-sensor data fusion method based on the PPT divergence,” the multi-sensor data fusion method is presented based on the PPT divergence and the Deng entropy. Section “Application” presents the two application cases of the proposed fusion method, and analyzes its results. Finally, section “Conclusion” concludes this article.

Preliminaries

In this section, some preliminaries are briefly introduced, including D–S evidence theory, the Deng entropy, pignistic probability transform, and several typical divergences for the purpose of understanding the descriptions in the rest of this article.

D–S evidence theory

Definition 1 (frame of discernment). Assuming that $\Theta$ is a set of mutually exclusive and exhaustive elements $\{F_i : i = 1, 2, 3, 4, \ldots, N\}$, and it can be defined as
\(
\Theta = \{F_1, F_2, F_3, F_4, \ldots, F_N\}
\)  
(1)

where \(\Theta\) is called the frame of discernment (FOD), and \(F_i\) is named single-element proposition or subset. We define \(2^\Theta\) as a power set which contains \(2^N\) elements and can be described as:

\[
2^\Theta = \{\emptyset, F_1, \ldots, F_N, \{F_1, F_2\}, \ldots, \Theta\}
\]  
(2)

where \(\emptyset\) is an empty set in equation (2).

**Definition 2 (the BPA).** In the FOD \(\Theta\), the BPA function \(m\) is also called mass function and is defined as a mapping of the power set \(2^\Theta\) to \([0,1]\):

\[
m: 2^\Theta \rightarrow [0,1]
\]

which satisfies

\[
m(\emptyset) = 0
\]

\[
0 \leq m(A) \leq 1
\]

\[
\sum_{A \subseteq \Theta} m(A) = 1
\]

where mass function \(m(A)\) represents the degree of support to \(A\), and \(A\) is called the focal element or subset. The mass function \(m(\emptyset)\) is equal to 0 in classical D–S evidence theory.

**Definition 3 (belief function).** Let \(A\) be a hypothesis in the FOD \(\Theta\). A belief function \(Bel: 2^\Theta \rightarrow [0,1]\) is defined as

\[
Bel(A) = \sum_{B \subseteq \Theta, B \subseteq A} m(B)
\]

(3)

**Definition 4 (plausibility function).** Let \(A\) be a hypothesis in the FOD \(\Theta\). A plausibility function \(Pl: 2^\Theta \rightarrow [0,1]\) is defined as

\[
Pl(A) = \sum_{B \subseteq \Theta, B \cap A \neq \emptyset} m(B)
\]

(4)

**Definition 5 (Dempster’s combination rule).** Assume \(m_1\) and \(m_2\) are the two independent BPAs on \(2^\Theta\). Dempster’s combination rule, represented in the form \(m = m_1 \oplus m_2\), is defined as follows:

\[
m(A) = (m_1 \oplus m_2)(A)
\]

(5)

\[
= \frac{1}{1 - K} \sum_{B \subseteq A} m_1(B)m_2(C)
\]

in which

\[
K = \sum_{B \subseteq A} m_1(B)m_2(C)
\]

(6)

where \(\oplus\) represents Dempster’s combination rule. \(K\) is called the conflict coefficient, and it has values between 0 and 1. The larger \(K\) is, the more conflict between two evidences is.

**Entropy**

**Definition 6 (the Deng entropy).** The Shannon entropy has a great contribution to the measurement of uncertainty, but it has some limitations when there is the BPA. Because it measures uncertainty based on probability. In order to solve this problem, Deng proposes the Deng entropy in the framework of D–S evidence theory. It is a generalization of the Shannon entropy, and is defined as

\[
H_d = - \sum_{A \subseteq \Theta} m(A) \log_2 \frac{m(A)}{2|A| - 1}
\]

(7)

where \(m\) is a mass function defined on the FOD \(\Theta\), \(A\) is the focal element of \(m\), and \(|A|\) is the cardinality of \(A\).

**Probability transform**

**Definition 7 (pignistic probability transform).** Suppose \(m\) is a BPA, the pignistic probability transform is defined as follows:

\[
Bet_Pm(A) = \sum_{B \subseteq \Theta} \frac{|A \cap B|}{|B|} \frac{m(B)}{1 - m(\emptyset)}
\]

(8)

where \(|A \cap B|\) is the cardinality of the intersection of \(A\) and \(B\). The mass function \(m(\emptyset)\) is equal to 0 in a closed world (i.e. in classical D–S evidence theory). So, it can be simplified to the following form

\[
Bet_Pm(A) = \sum_{B \subseteq \Theta} \frac{|A \cap B|}{|B|} m(B)
\]

(9)

The essence of pignistic probability transform is to convert a BPA function into a probability distribution. It allocates the belief to each element equally. When \(A\) is single-element subset, equation (9) can be simplified as

\[
Bet_Pm(A) = \sum_{B \subseteq \Theta} \frac{m(B)}{|B|}
\]

(10)

**Divergence measures**

**Definition 8 (JS divergence).** Given two probability distributions \(P = \{p_1, p_2, \ldots, p_n\}\) and \(Q = \{q_1, q_2, \ldots, q_n\}\), with \(\sum_i p_i = \sum_i q_i = 1\). The JS divergence between \(P\) and \(Q\) is denoted as

xu et al.
\[ JS(P, Q) = \frac{1}{2} \left[ \sum_i P_i \log \frac{P_i}{\frac{1}{2}P_i + \frac{1}{2}Q_i} + \sum_i Q_i \log \frac{Q_i}{\frac{1}{2}P_i + \frac{1}{2}Q_i} \right] \] (11)

**Definition 9 (BJS divergence).** Given two independent BPAs \( m_1 \) and \( m_2 \) defined on \( \Theta \), the BJS divergence between \( m_1 \) and \( m_2 \) is denoted as:

\[ BJS(m_1, m_2) = \frac{1}{2} \left[ \sum_{A \subseteq \Theta} m_1(A) \log \frac{m_1(A)}{\frac{1}{2}m_1(A) + \frac{1}{2}m_2(A)} + \sum_{A \subseteq \Theta} m_2(A) \log \frac{m_2(A)}{\frac{1}{2}m_1(A) + \frac{1}{2}m_2(A)} \right] \] (12)

The essence of BJS divergence is to replace the probabilities distributions in JS divergence with BPAs, so that it can be applied to measure the discrepancy between evidences.

**Definition 10 (RBS divergence).** Let \( H \) be the FOD, which contains \( h \) mutually exclusive and collectively exhaustive events. Let \( m_1 \) and \( m_2 \) be the two belief functions in \( H \). The RBS divergence between the belief functions \( m_1 \) and \( m_2 \) is defined as:

\[ RB(m_1, m_2) = \sqrt{\frac{\mathcal{B}(m_1, m_1) + \mathcal{B}(m_2, m_2) - 2\mathcal{B}(m_1, m_2)}{2}} \] (13)

where

\[ \mathcal{B}(m_1, m_2) = ~ 
\]

\[ = \frac{1}{2} \sum_{i=1}^{h} \sum_{j=1}^{h} m_1(A_i) \log \frac{m_1(A_i)}{\frac{1}{2}m_1(A_i) + \frac{1}{2}m_2(A_i)} |A_i \cap A_j| + \frac{1}{2} \sum_{i=1}^{h} \sum_{j=1}^{h} m_2(A_j) \log \frac{m_2(A_j)}{\frac{1}{2}m_1(A_i) + \frac{1}{2}m_2(A_j)} |A_i \cap A_j| \]

**Definition 11 (Wang et al’s divergence).** Given two independent BPAs \( m_1 \) and \( m_2 \) defined on \( \Theta \), the new divergence proposed by Wang et al. between \( m_1 \) and \( m_2 \) is denoted as:

\[ D(m_1, m_2) = \frac{1}{2} \left[ \sum_{\theta \in \Theta} PBl_m(\theta) \log \frac{PBl_m(\theta)}{\frac{1}{2}PBl_m(\theta) + \frac{1}{2}PBl_m(\theta)} + \sum_{\theta \in \Theta} PBl_m(\theta) \log \frac{PBl_m(\theta)}{\frac{1}{2}PBl_m(\theta) + \frac{1}{2}PBl_m(\theta)} \right] \] (14)

where

\[ PBl_m(\theta) = \frac{Bel(\theta) + Pl(\theta)}{\sum_{\theta' \in \Theta} Bel(\theta') + Pl(\theta')} \]

**The PPT divergence measure**

Although BJS divergence can measure the difference between evidences and achieve accurate fusing results, the multi-element subsets are treated as single-element subsets for divergence calculation, which ignores the influence of multi-element subsets on divergence measurement. A counter example is depicted in Example 1.

**Example 1.** Given three BPAs \( m_1, m_2, \) and \( m_3 \) defined on \( \Theta = \{A, B\} \), they are shown as follows:

\[ m_1 : m_1(A) = 0.70, m_1(B) = 0.15, m_1(A, B) = 0.15 \]
\[ m_2 : m_2(A) = 0.15, m_2(B) = 0.70, m_2(A, B) = 0.15 \]
\[ m_3 : m_3(A) = 0.15, m_3(B) = 0.15, m_3(A, B) = 0.70 \]

The divergence is calculated with BJS divergence as follows:

\[ BJS(m_1, m_2) = 0.3863 \]
\[ BJS(m_1, m_3) = 0.3863 \]
\[ BJS(m_2, m_3) = 0.3863 \]

As we know, object \( A \) has the greatest reliability for \( m_1 \), object \( B \) has the greatest reliability for \( m_2 \), and object \( A, B \) has the greatest reliability for \( m_3 \). Obviously, the difference between \( m_1 \) and \( m_2 \) is greater than that of \( m_1 \) and \( m_3 \) or \( m_2 \) and \( m_3 \). Whereas, the results calculated by BJS divergence illustrate that there is no difference between \( m_1, m_2, \) and \( m_3 \), which is contrary to common sense. Therefore, to solve above problem, we propose a novel divergence measure.

**The definition of the PPT divergence**

Let \( m_1 \) and \( m_2 \) be the BPAs in the FOD \( \Theta \). The novel divergence between \( m_1 \) and \( m_2 \) is expressed as:

\[ PPT(m_1, m_2) = \frac{1}{2} \left[ I \left( \frac{BetP_{m_1} + BetP_{m_2}}{2} \right) + I \left( \frac{BetP_{m_1} + BetP_{m_2}}{2} \right) \right] \] (15)

where \( BetP_m(A) = \sum_{B \subseteq \Theta} \frac{(m(B) \cap |B|)}{2} \), and note that \( A \) is a single-element subset. \( I(m_1, m_2) \) represents the Kullback–Leibler (KL) divergence. Obviously, \( PPT(m_1, m_2) \) can also be described in the following form.
The pignistic probability transform is described as keeping the belief of the single-element subset unchanged and evenly allocating the belief of the multi-element subset to the included single-element subset. Essentially, the function Bet$_P$ transforms the BPA into a probability distribution. The proposed divergence, named PPT divergence, fully reflects the interaction between single-element and multi-element subsets, thereby effectively avoiding the situation that multi-element subsets are regarded as single-element subsets in calculating the divergence between evidences. Furthermore, when the BPA consists of only singleton subsets, the PPT divergence is degenerated as JS divergence.

The properties of the PPT divergence

There are some properties of the PPT divergence shown as follows:

1. boundedness: $0 \leq \text{PPT}(m_1, m_2) \leq 1$;
2. non-degeneracy: PPT$(m_1, m_2) = 0$ if and only if $m_1 = m_2$;
3. symmetry: PPT$(m_1, m_2) = \text{PPT}(m_2, m_1)$.

**Boundedness. Proof.** For equation (10), we have

$$0 \leq \text{Bet}_P(A) = \sum_{B \subseteq A} \frac{m(B)}{|B|} = m(A) + \sum_{B \subseteq A} m(B) = 1$$

Based on the Shannon inequality, \(^{38}\) we have $I(P, Q) \geq 0$, where $P, Q$ are the two probability distributions. In equation (15), the Bet$_P$ transforms BPA into a probability distribution. Consequently, Bet$_P^1$, Bet$_P^2$, and Bet$_P^1 + \text{Bet}_P^2 / 2$ are all the probability distributions. In the same way, $I(\text{Bet}_P^1, \text{Bet}_P^2) + \text{Bet}_P^1 / 2) \geq 0$ and $I(\text{Bet}_P^2, \text{Bet}_P^1 + \text{Bet}_P^1 / 2) \geq 0$ hold, so PPT$(m_1, m_2) \geq 0$

For equation (15), we have

$$I \left( \frac{\text{Bet}_P^1 + \text{Bet}_P^2}{2}, \frac{\text{Bet}_P^1 + \text{Bet}_P^2}{2} \right) \leq 1$$

$$I \left( \frac{\text{Bet}_P^1 + \text{Bet}_P^2}{2}, \frac{\text{Bet}_P^1 + \text{Bet}_P^2}{2} \right) \leq 1$$

Obviously
For equation (16), we have

\[ \text{Non-degeneracy. Proof.} \] For equation (16), we have

\[
PPT(m_1, m_2) \leq 1
\]

Finally, we can conclude that the value of the PPT divergence is \([0, 1]\).

**Symmetry. Proof.** For equation (15), we have

\[
PPT(m_1, m_2) = \frac{1}{2} \left[ \sum_{A_i \in \Theta} \text{BetP}_{m_1}(A_i) \log \frac{\text{BetP}_{m_1}(A_i)}{\frac{1}{2} \text{BetP}_{m_1}(A_i) + \frac{1}{2} \text{BetP}_{m_2}(A_i)} \right]
\]

This equation holds only if

\[ \text{BetP}_{m_1}(A_i) = \text{BetP}_{m_2}(A_i) \]

Thus, we can conclude that the non-degeneracy of PPT divergence holds.

\[
PPT(m_1, m_2) = \frac{1}{2} \left[ \left( \text{BetP}_{m_1} + \text{BetP}_{m_2} \right) + \left( \text{BetP}_{m_1} + \text{BetP}_{m_2} \right) \right]
\]

Obviously

\[ m_1 = m_2 \]

Thus, we can conclude that the symmetry of PPT divergence holds.

\[
PPT(m_2, m_1) = \frac{1}{2} \left[ \left( \text{BetP}_{m_2} + \text{BetP}_{m_1} \right) + \left( \text{BetP}_{m_2} + \text{BetP}_{m_1} \right) \right]
\]

Obviously

\[ PPT(m_1, m_2) = PPT(m_2, m_1) \]

Thus, we can conclude that the symmetry of PPT divergence holds.

**Comparative analysis**

In this section, numerical examples (Examples 1–4 from Wang et al.\(^{29}\)) are used to illustrate the validity of PPT divergence and compare it with BJS divergence.

The divergence value, recalled Example 1, is calculated with the PPT divergence as follows

\[ \text{BetP}_{m_1}(A) = 0.70 + \frac{0.15}{2} = 0.775 \]

\[ \text{BetP}_{m_1}(B) = 0.15 + \frac{0.15}{2} = 0.225 \]

\[ \text{BetP}_{m_2}(A) = 0.15 + \frac{0.15}{2} = 0.225 \]

\[ \text{BetP}_{m_2}(B) = 0.70 + \frac{0.15}{2} = 0.775 \]

\[ \text{BetP}_{m_1}(A) = 0.15 + \frac{0.15}{2} = 0.500 \]

\[ \text{BetP}_{m_1}(B) = 0.70 + \frac{0.15}{2} = 0.500 \]

\[ PPT(m_1, m_2) = \frac{1}{2} \left[ (0.775 \times \log \frac{0.775}{0.775 + 0.225}) + \frac{0.225 \times \log \frac{0.225}{0.775 + 0.225}}{2} \right] \]

\[ + \left( 0.225 \times \log \frac{0.225}{0.775 + 0.225} \right) + 0.775 \times \log \frac{0.775}{0.775 + 0.225} \]

\[ = 0.3194 \]

\[ PPT(m_2, m_3) = \frac{1}{2} \left[ (0.225 \times \log \frac{0.225}{0.225 + 0.5}) + \frac{0.775 \times \log \frac{0.775}{0.775 + 0.5}}{2} \right] \]

\[ + \left( 0.5 \times \log \frac{0.5}{0.775 + 0.225} \right) + 0.5 \times \log \frac{0.5}{0.775 + 0.225} \]

\[ = 0.1992 \]

\[ PPT(m_1, m_3) = \frac{1}{2} \left[ (0.775 \times \log \frac{0.775}{0.775 + 0.225}) + \frac{0.225 \times \log \frac{0.225}{0.225 + 0.5}}{2} \right] \]

\[ + \left( 0.5 \times \log \frac{0.5}{0.775 + 0.225} \right) + 0.5 \times \log \frac{0.5}{0.775 + 0.225} \]

\[ = 0.1992 \]

The results show that the divergence value \(PPT(m_1, m_2)\) is larger than that of \(PPT(m_1, m_3)\) and \(PPT(m_2, m_3)\), which is in line with human intuition.
Example 2. Suppose that there are two BPAs $m_1$ and $m_2$ in the FOD which is $\Theta = \{A, B, C, D, E, F, G, H\}$. We change $\alpha$ step by step from 0 to 1. In each step, $\alpha$ increases $\Delta = 0.1$. The cardinality of subset $T$ changes from 1 to 8, as shown in Table 1.

$$m_1 : m_1(B) = \alpha, \quad m_1(T) = 1 - \alpha$$
$$m_2 : m_2(B) = 0.9, \quad m_2(T) = 0.1$$

![Graphs](image)

**Table 1.** The variable subset $T$.

| $|T|$ | $T$          |
|-----|--------------|
| 1   | $A$          |
| 2   | $A, B$      |
| 3   | $A, B, C$   |
| 4   | $A, B, C, D$|
| 5   | $A, B, C, D, E$ |
| 6   | $A, B, C, D, E, F$ |
| 7   | $A, B, C, D, E, F, G$ |
| 8   | $A, B, C, D, E, F, G, H$ |

**Figure 1.** The change of divergence value in Example 2. (a) The divergence value varying with $|T|$ and $\alpha$. (b) The divergence value varying with $|T|$. (c) The divergence value varying with $\alpha$.

**Figure 2.** Comparison of the BJ$S$ and PPT divergence in Example 3.

The PPT divergence value varying with $|T|$ and $\alpha$ is shown in Figure 1. We can conclude from Figure 1(a)
that the divergence value fluctuates from 0 to 1, no matter how $|T|$ and $\alpha$ change. The boundedness of the PPT divergence is verified and its value is $[0,1]$. When $|T| = 1$, namely, the evidences $m_1$ and $m_2$ consist of singleton subsets, there is no intersection between the subsets. Thus, the divergence value is larger than that of other $|T|$ values in Figure 1(b). As the $|T|$ increases from 2 to 8, the divergence value increases accordingly. This is intuitive, because the cardinality extension of a multi-element subset must result in an increase in uncertainty. As shown in Figure 1(c), when $|T|$ is fixed, the divergence value decreases monotonically as $\alpha$ increases from 0 to 0.9. This is also consistent with intuitive, because the belief distribution of the two BPAs is getting more and more similar. When $\alpha = 0.9$, namely, $m_1 = m_2$, the divergence value is always equal to 0, no matter what the value of $|T|$ is. It proves once again that the non-degeneracy of PPT divergence holds.

**Example 3.** Suppose that there are two BPAs $m_1$ and $m_2$ in the FOD which is $\Theta = \{A, B, C, D, E, F, G, H\}$. The cardinality of subset $T$ changes from 2 to 8, as shown in Table 1

$$m_1 : m_1(B) = 0.1, \quad m_1(T) = 0.9$$
$$m_2 : m_2(B) = 0.9, \quad m_2(T) = 0.1$$

The divergence values varying with $T$ are shown in Figure 2. Intuitively, with the increase in cardinality of multi-element subsets, the divergence value between two BPAs increases accordingly. However, the BJS divergence value always remains the same. The reason is that the BJS divergence treats the multi-element subset as single-element subset for calculation, which cannot reflect the influence of multi-element subsets on divergence measurement. On the contrary, the PPT divergence fully considers the interaction between single-element and multi-element subsets and obtains rational results, which illustrates the feasibility of the PPT divergence.

**Example 4.** Suppose that there are two BPAs $m_1$ and $m_2$ in the FOD which is $\Theta = \{A, B\}$. We change $\alpha$ step by step from 0 to 1. In each step, $\alpha$ increases $\Delta = 0.02$.

$$m_1 : m_1(A) = 1 - \alpha, \quad m_1(B) = \alpha$$
$$m_2 : m_2(A) = 0.1, \quad m_2(B) = 0.9$$

For every change of $\alpha$, the PPT divergence and the BJS divergence achieve the same value as shown in Figure 3. This is reasonable, because $\text{BetP}_m(A) = \sum_{B \in \Theta} m(B)/|B|$ can be simplified to $\text{BetP}_m(A) = m(A)$ when two BPAs only consist of singleton subsets, namely, the PPT divergence degenerates into the BJS divergence. In this case, the BJS divergence assigns the mass function’s hypothesis to the single element and the BPA is turned into probability. In fact, the BJS divergence also degenerates to JS divergence.

We can conclude from the above examples that the results obtained by the PPT divergence are more reasonable than those obtained by the BJS divergence in measuring the divergence between BPAs.

**The multi-sensor data fusion method based on the PPT divergence**

To combine highly conflicting evidence effectively, a new multi-sensor data fusion method is proposed based on the PPT divergence and the Deng entropy. In this proposed method, the PPT divergence is used to measure the discrepancy between evidences to obtain the credibility weight of evidence which is inversely proportional to the divergence value, and the Deng entropy is used to measure the uncertainty of evidence to obtain the information volume weight of evidence. The proposed method comprehensively determines the final weight of evidence from two aspects: difference and
uncertainty, which fully excavates the potential information between evidences, so that it can deal with conflicting evidence more effectively. The flowchart of proposed method is shown in Figure 4.

**Obtain the credibility weight of evidence**

Step 1-1: assume there are \( k \) independent BPAs on the same FOD that contains \( n \) elements: \( \Theta = \{A_1, A_2, \ldots, A_n\} \). The divergence value between two BPAs is generated with equation (16). Then, a divergence matrix \( D_{k \times k} \) is constructed as follows

\[
D_{k \times k} = \begin{bmatrix}
0 & PPT(m_1, m_2) & \cdots & PPT(m_1, m_k) \\
PPT(m_2, m_1) & 0 & \cdots & PPT(m_2, m_k) \\
\vdots & \vdots & \ddots & \vdots \\
PPT(m_k, m_1) & PPT(m_k, m_2) & \cdots & 0
\end{bmatrix}
\]  

(17)

Step 1-2: the support degree \( S \) of evidence \( m_i \) is calculated as

\[
S(m_i) = \frac{1}{\sum_{j=1}^{k} PPT(m_i, m_j)}
\]  

(18)

Note that, to avoid that the support degree \( S \) of evidence \( m_i \) is meaningless, we need to replace \( \sum_{j=1}^{k} PPT(m_i, m_j) \) with \( e^{\sum_{j=1}^{k} PPT(m_i, m_j)} \), if and only if there are \( k \) identical pieces of evidence.

Step 1-3: the credibility weights \( W_d \) of evidence \( m_i \) are defined as

\[
W_d(m_i) = \frac{S(m_i)}{\sum_{i=1}^{k} S(m_i)}
\]  

(19)

**Obtain the information volume weight of evidence**

Step 2-1: the Deng entropy \( H_d \) of evidence \( m_i \) is generated with equation (7)

\[
H_d(m_i) = -\sum_{A \in \Theta} m_i(A) \log_2 \frac{m_i(A)}{2^{d_i} - 1}, \quad i = 1, 2, \ldots, k
\]  

(20)

Step 2-2: to avoid assigning zero weight to the evidence whose belief entropy is zero, the information volume is defined as

\[
IV(m_i) = e^{H_d(m_i)}, \quad i = 1, 2, \ldots, k
\]  

(21)

Step 2-3: the information volume weights \( W_{iv} \) of evidence \( m_i \) is calculated as
Table 2. The collected sensor reports modeled as BPA in the target recognition problem.

|      | A  | B  | C  | [A, C] |
|------|----|----|----|--------|
| m₁   | 0.40| 0.28| 0.30| 0.02   |
| m₂   | 0.01| 0.90| 0.08| 0.01   |
| m₃   | 0.63| 0.06| 0.01| 0.30   |
| m₄   | 0.60| 0.09| 0.01| 0.30   |
| m₅   | 0.60| 0.09| 0.01| 0.30   |

BPA: basic probability assignment.

\[
W_i(m_i) = \frac{IV(m_i)}{\sum_{i=1}^{k} IV(m_i)}, \quad i = 1, 2, \ldots, k
\] (22)

Generate and combine the weighted average evidence

Step 3-1: the final weight \( w \) of evidence \( m_i \) is calculated as

\[
w(m_i) = \frac{W_c(m_i) \times W_i(m_i)}{\sum_{i=1}^{k} W_c(m_i) \times W_i(m_i)}, \quad i = 1, 2, \ldots, k
\] (23)

Step 3-2: the weighted average evidence \( \tilde{m} \) is calculated as

\[
\tilde{m}(A) = \sum_{i=1}^{k} w(m_i) \times m_i(A)
\] (24)

Step 3-3: Dempster’s combination rule is used \( k - 1 \) times to combine the weighted average evidence according to equation (5). Then, the final combination result is calculated as

\[
m(A) = (\tilde{m} \oplus \tilde{m} \oplus \cdot \cdot \cdot \oplus \tilde{m})_{k-1}(A)
\] (25)

Application

In this section, the proposed multi-sensor data fusion method is applied in two application cases of target recognition and fault diagnosis.

Application in target recognition

Problem statement. Consider a multi-sensor target recognition problem associated with the sensor reports that are collected from five different types of sensors \( \{S_1, S_2, S_3, S_4, S_5\} \), each of which is allocated to a different position in order to monitor the objectives. The FOD which consists of three types target is given by \( \Theta = \{F_1, F_2, F_3\} \). These sensors collect the target information and generate reports, which are modeled as BPA, denoted by \( m_1(\cdot), m_2(\cdot), m_3(\cdot), m_4(\cdot) \) and \( m_5(\cdot) \) in Table 2. Obviously, the evidence \( m_3 \) highly conflicts with other evidences. The application is cited from Xiao.²⁸

Combine evidences by the proposed multi-sensor data fusion method

Step 1-1: establish the divergence matrix \( D_{5 \times 5} \) as follows

\[
D_{5 \times 5} = \begin{bmatrix}
0 & 0.2347 & 0.0802 & 0.0632 & 0.0632 \\
0.2347 & 0 & 0.4648 & 0.4265 & 0.4265 \\
0.0802 & 0.4648 & 0 & 0.0020 & 0.0020 \\
0.0632 & 0.4265 & 0.0020 & 0 & 0.0001 \\
0.0632 & 0.4265 & 0.0020 & 0.0001 & 0
\end{bmatrix}
\]

Step 1-2: calculate the support degree of evidence as follows

\[
S(m_1) = 2.2659 \\
S(m_2) = 0.6441 \\
S(m_3) = 1.8228 \\
S(m_4) = 2.0338 \\
S(m_5) = 2.0338
\]

Step 1-3: calculate the credibility weight \( W_d \) of evidence as follows

\[
W_d(m_1) = 0.2575 \\
W_d(m_2) = 0.0732 \\
W_d(m_3) = 0.2071 \\
W_d(m_4) = 0.2311 \\
W_d(m_5) = 0.2311
\]

Step 2-1: compute the belief entropy of evidence \( m_i \) as follows

\[
H_d(m_1) = 1.7089 \\
H_d(m_2) = 0.5771 \\
H_d(m_3) = 1.7268 \\
H_d(m_4) = 1.8181 \\
H_d(m_5) = 1.8181
\]

Step 2-2: the information volume \( IV \) of evidence \( m_i \) is calculated as follows

\[
IV(m_i) = \sum_{A} w(m_a) 	imes H_d(m_a)
\]
IV m
dðÞ = 5.5231
IV m
dðÞ = 1.7808
IV m
dðÞ = 5.6225
IV m
dðÞ = 6.1605
IV m
dðÞ = 6.1605

Step 2-3: the information volume weight \( W_{iv} \) of evidence \( m_i \) is calculated as follows

\[
\begin{align*}
W_{iv}(m_1) &= 0.2188 \\
W_{iv}(m_2) &= 0.0705 \\
W_{iv}(m_3) &= 0.2227 \\
W_{iv}(m_4) &= 0.2440 \\
W_{iv}(m_5) &= 0.2440
\end{align*}
\]

Step 3-1: the final weight \( w \) of evidence \( m_i \) is calculated as follows

\[
\begin{align*}
w(m_1) &= 0.2556 \\
w(m_2) &= 0.0234 \\
w(m_3) &= 0.2092 \\
w(m_4) &= 0.2559 \\
w(m_5) &= 0.2559
\end{align*}
\]

Step 3-2: the weighted average evidence \( \tilde{m} \) is calculated as follows

\[
\tilde{m}(A) = 0.5413 \\
\tilde{m}(B) = 0.1513 \\
\tilde{m}(C) = 0.0858 \\
\tilde{m}(A, C) = 0.2216
\]

Step 3-3: use Dempster’s combination rule to fuse the weighted average evidence four times, and the final combination results are shown in Table 3 and Figure 5.

**Discussion**

The result of combining conflicting evidence with different methods is shown in Table 3. The comparisons of the BPAs generated by different methods are shown in Figure 5.

As shown in Table 3, although all of the methods can recognize the objective \( A \) as the true target, the proposed method achieves a much higher belief value (0.9892) than Xiao’s\(^{27}\) method and Xiao’s\(^{28}\) method, which illustrates that the proposed method solves the shortcomings of Xiao’s\(^{27}\) method and has a greater improvement in the belief value for the correct target than Xiao’s\(^{28}\) method. The reason is that it not only reflects the interaction between single-element and multi-element subsets by introducing the PPT, but also uses the credibility weights and the information volume weights to determine the final weights of evidences. In real applications, a slight increase of the belief value for the correct target is significant to improve the performance of target recognition system. Comparatively, the proposed method and Wang et al.’s\(^{29}\) method similarly achieve the highest belief degree of the correct target.

According to the time complexity calculation method presented in Li and Xiao,\(^{39}\) the proposed method calculates the divergence value between two evidences for \( k(k - 1) \) times. Since \( Bet_{\tilde{m}}(A_i) = \sum_{A_i \subseteq B} m(B)/|B| (i = 1, 2, \ldots, n) \), we need to process

![Figure 5. The comparison of the BPAs generated by different methods for target recognition.](image-url)
The complexity of calculation of divergence matrix is \(O(k^2 \times n \times 2^{n-1})\). The time complexity of combining the weighted average evidence is \(O(k \times 2^{n-1})\). The overall time complexity of the proposed method is \(O(k^2 \times n \times 2^{n-1})\), where \(k\) and \(n\) represent the number of evidences and elements in the FOD, respectively. Wang et al.’s method\(^{29}\) calculates the divergence value between two evidences for \(k - 1\) times. Since

\[
PBl_m(A_i) = \sum_{B \subseteq A_i} m(B) + \sum_{B \cap A_i \neq \emptyset} m(B) / \sum_{A \subseteq \Theta} \left( \sum_{B \subseteq A} m(B) + \sum_{B \cap A \neq \emptyset} m(B) \right) (i = 1, 2, \ldots, n),
\]

we need to process \(n^2 \times 2^{n-1}\) elements to obtain \(D(m_1, m_2)\). The time complexity of calculation of divergence matrix is \(O(k^2 \times n \times 2^{n-1})\). The weighted average evidence will be combined for \(k - 1\) times. Consequently, the time complexity of combining the weighted average evidence is \(O(k \times 2^{n-1})\). The overall time complexity of the proposed method is \(O(k^2 \times n \times 2^{n-1})\), where \(k\) and \(n\) represent the number of evidences and elements in the FOD, respectively. Wang et al.’s method\(^{29}\) calculates the divergence value between two evidences for \(k - 1\) times. Since

\[
PBl_m(A_i) = \sum_{B \subseteq A_i} m(B) + \sum_{B \cap A_i \neq \emptyset} m(B) / \sum_{A \subseteq \Theta} \left( \sum_{B \subseteq A} m(B) + \sum_{B \cap A \neq \emptyset} m(B) \right) (i = 1, 2, \ldots, n),
\]

we need to process \(n^2 \times 2^{n-1}\) elements to obtain \(D(m_1, m_2)\). The time complexity of calculation of divergence matrix is \(O(k^2 \times n \times 2^{n-1})\). The weighted average evidence will be combined for \(k - 1\) times. Thus, the time complexity of combining the weighted average evidence is \(O(k \times 2^{n-1})\). The overall time complexity of Wang et al.’s method is \(O(k^2 \times n^2 \times 2^{n-1})\). Obviously, the time complexity of the proposed method is lower than that of Wang et al.’s method. Therefore, the proposed method outperforms other four methods in target recognition application.

### Application in fault diagnosis

**Problem statement.** Given the FOD \(\Theta = \{F_1, F_2, F_3\}\), where \(F_1, F_2,\) and \(F_3\) represent the three kinds of faults in the automobile system: low oil pressure, air leakage in the intake system, and solenoid valve jam. The set of sensors which are positioned on different places for collecting the fault information in the automobile system is \(S = \{S_1, S_2, S_3, S_4, S_5\}\). \(m_i\) is the evidence given by sensor \(S_i(i = 1, 2, 3, 4, 5)\). The collected sensor reports which are modeled as BPAs are provided in Table 4. According to the collected sensor reports, \(S_5\) is damaged due to speed overload, so the evidence \(m_5\) highly conflicts with other evidences. The application is cited from Xiao.\(^{28}\)

### Combine evidences by the proposed multi-sensor data fusion method

**Step 1-1:** establish the divergence matrix \(D_{5 \times 5}\) as follows

\[
D_{5 \times 5} = \begin{bmatrix}
0 & 0.0062 & 0.0222 & 0.0151 & 0.4494 \\
0.0062 & 0 & 0.0140 & 0.0022 & 0.4410 \\
0.0222 & 0.0140 & 0 & 0.0261 & 0.4315 \\
0.0151 & 0.0022 & 0.0261 & 0 & 0.4436 \\
0.4494 & 0.4410 & 0.4315 & 0.4436 & 0 \\
\end{bmatrix}
\]

**Step 1-2:** compute the support degree \(S_i\) of evidence as follows

\[
S(m_1) = 2.1146 \\
S(m_2) = 2.1590 \\
S(m_3) = 2.1105 \\
S(m_4) = 2.0535 \\
S(m_5) = 0.5665
\]

**Step 1-3:** calculate the credibility weight \(W_d\) of evidence as follows

\[
W_d(m_1) = 0.2348 \\
W_d(m_2) = 0.2398 \\
W_d(m_3) = 0.2344 \\
W_d(m_4) = 0.2281 \\
W_d(m_5) = 0.0629
\]

**Step 2-1:** compute the belief entropy of evidence \(m_i\) as follows

\[
H_d(m_1) = 1.7183 \\
H_d(m_2) = 1.7235 \\
H_d(m_3) = 1.8404 \\
H_d(m_4) = 1.5532 \\
H_d(m_5) = 0.7219
\]

**Step 2-2:** the information volume \(IV\) of evidence \(m_i\) is calculated as follows

\[
IV(m_1) = 5.5748 \\
IV(m_2) = 5.6041 \\
IV(m_3) = 6.2989 \\
IV(m_4) = 4.7267 \\
IV(m_5) = 2.0584
\]
Step 2-3: the information volume weight $W_{iv}$ of evidence $m_i$ is calculated as follows

\[
W_{iv}(m_1) = 0.2298
\]

\[
W_{iv}(m_2) = 0.2310
\]

\[
W_{iv}(m_3) = 0.2596
\]

\[
W_{iv}(m_4) = 0.1948
\]

\[
W_{iv}(m_5) = 0.0848
\]

Step 3-1: the final weight $w$ of evidence $m_i$ is calculated as follows

\[
w(m_1) = 0.2453
\]

\[
w(m_2) = 0.2517
\]

\[
w(m_3) = 0.2767
\]

\[
w(m_4) = 0.2020
\]

\[
w(m_5) = 0.0243
\]

Step 3-2: the weighted average evidence $\tilde{m}$ is calculated as follows

\[
\tilde{m}(F_1) = 0.6793
\]

\[
\tilde{m}(F_2) = 0.0709
\]

\[
\tilde{m}(F_3) = 0.0295
\]

\[
\tilde{m}(\Theta) = 0.2203
\]

Step 3-3: use Dempster’s combination rule to fuse the weighted average evidence four times, and the final combination results are shown in Table 5 and Figure 6.

**Table 5.** The fusing results by different methods in the fault diagnosis.

| Method      | $F_1$ | $F_2$ | $F_3$ | $\Theta$ | Target |
|-------------|-------|-------|-------|----------|--------|
| Dempster13  | 0     | 0.3443| 0.6557| 0        | $F_3$  |
| Xiao27      | 0.9946| 0.0030| 0.0015| 0.0009   | $F_1$  |
| Xiao28      | 0.9941| 0.0026| 0.0024| 0.0009   | $F_1$  |
| Wang et al.29 | 0.9957| 0.0026| 0.0008| 0.0009   | $F_1$  |
| Proposed method | 0.9957| 0.0026| 0.0008| 0.0009   | $F_1$  |

**Figure 6.** The comparison of the BPAs generated by different methods for fault diagnosis.

As shown in Table 5 and Figure 6, facing the conflicting sensor report $m_5$, Dempster’s method fails to combine the highly conflicting evidence, which generates counter-intuitive result and treats the fault type $F_1$ as the true fault type, even if the other four evidences support the fault type $F_1$. Whereas, the proposed method, Xiao’s27 method, Xiao’s28 method, and Wang et al.’s29 method can handle the highly conflicting evidence and recognizes the fault type $F_1$ correctly. Furthermore, compared with Xiao’s27 method, the proposed method is more effective in managing conflicting evidence, and has the higher belief value (0.9957) of the correct fault type as shown in Table 5. The main reason is that the PPT divergence reflects the interaction between single-element and multi-element subsets by introducing PPT. Meanwhile, the proposed method has more room for improvement in the belief value of the correct target than Xiao’s28 method. The reason is that the proposed method not only makes use of the PPT divergence to obtain the credibility weight, but also considers the uncertainty of the evidences by adopting the Deng entropy to obtain the information volume weight. Then, these two kinds of weights are used to determine the final weight comprehensively. The proposed method and Wang et al.’s29 method similarly achieve the highest belief value for the correct fault types, but the time complexity of the proposed method is lower than that of Wang et al.’s29 method. The above analysis fully demonstrates the effectiveness and superiority of the proposed method in fault diagnosis application.

**Discussion**

As shown in Table 5 and Figure 6, facing the conflicting sensor report $m_5$, Dempster’s method fails to combine the highly conflicting evidence, which generates counter-intuitive result and treats the fault type $F_1$ as the true fault type, even if the other four evidences support the fault type $F_1$. Whereas, the proposed method, Xiao’s27 method, Xiao’s28 method, and Wang et al.’s29 method can handle the highly conflicting evidence and recognizes the fault type $F_1$ correctly. Furthermore, compared with Xiao’s27 method, the proposed method is more effective in managing conflicting evidence, and has the higher belief value (0.9957) of the correct fault type as shown in Table 5. The main reason is that the PPT divergence reflects the interaction between single-element and multi-element subsets by introducing PPT. Meanwhile, the proposed method has more room for improvement in the belief value of the correct target than Xiao’s28 method. The reason is that the proposed method not only makes use of the PPT divergence to obtain the credibility weight, but also considers the uncertainty of the evidences by adopting the Deng entropy to obtain the information volume weight. Then, these two kinds of weights are used to determine the final weight comprehensively. The proposed method and Wang et al.’s29 method similarly achieve the highest belief value for the correct fault types, but the time complexity of the proposed method is lower than that of Wang et al.’s29 method. The above analysis fully demonstrates the effectiveness and superiority of the proposed method in fault diagnosis application.

**Conclusion**

In this article, a novel divergence measure, called PPT divergence, is proposed to measure the difference between evidences. The proposed PPT divergence can reflect the interaction between single-element and multi-element subsets by introducing the PPT, and overcome the shortcomings of the BJS divergence which treats multi-element subset as single-element subset for calculation. Furthermore, the PPT divergence satisfies the properties of boundedness, non-degeneracy, and symmetry, and can degenerate as JS divergence when the mass function and the probability distribution are consistent. Comparative analysis shows that the PPT divergence can obtain more reasonable results than the BJS divergence in measuring the
discrepancy between evidences. Based on the PPT divergence, a new multi-sensor data fusion method is presented. The proposed method takes advantage of PPT divergence to measure the discrepancy between evidences in order to obtain the credibility weights, and the Deng entropy to measure the uncertainty of the evidences in order to obtain the information volume weights, which can fully mine the potential information between evidences. Then, the credibility weights and the information volume weights are integrated to generate the weighted average evidence before using Dempster’s combination rule. Through two application cases analysis, compared with Xiao’s and Wang et al.’s multi-sensor data fusion method, our proposed method has higher belief degree for the correct target and lower time complexity, respectively. In general, the proposed method outperforms other four methods for combining highly conflicting evidences.

Declaration of conflicting interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National Natural Science Foundation of China (grant nos 61903373 and 61921001).

ORCID iD
Shijun Xu https://orcid.org/0000-0003-4379-4587

References
1. Deng Y. Information volume of mass function. Int J Comput Commun Control 2020; 15(6): 3983.
2. Liu F, Gao X, Zhao J, et al. Generalized belief entropy and its application in identifying conflict evidence. IEEE Access 2019; 7(1): 126625–126633.
3. Zadeh LA. Fuzzy sets as a basis for a theory of possibility. Fuzzy Set Syst 1999; 100(1): 9–34.
4. Baykasoglu A and Golecik I. Development of a novel multiple-attribute decision making model via fuzzy cognitive maps and hierarchical fuzzy TOPSIS. Inform Sci 2015; 301: 75–98.
5. Ah A, As A and Hn A. Distributed approach for computing rough set approximations of big incomplete information systems. Inform Sci 2021; 547: 427–449.
6. Zhang P, Li T, Wang G, et al. Multi-source information fusion based on rough set theory: a review. Inform Sci 2021; 68: 85–117.
7. Jiang W, Huang C and Deng X. A new probability transformation method based on a correlation coefficient of belief functions. Int J Intell Syst 2019; 34(6): 1337–1347.
8. Huang J, Xu DH, Liu HC, et al. A new model for failure mode and effect analysis integrating linguistic Z-numbers and projection method. IEEE T Fuzzy Syst 2021; 29(3): 530–538.
9. Xia J, Feng YQ, Liu LN, et al. On entropy function and reliability indicator for D numbers. Appl Intell 2019; 49(9): 3248–3266.
10. Mo HM. A SWOT method to evaluate safety risks in life cycle of wind turbine extended by D number theory. J Intell Fuzzy Syst 2021; 40(3): 4439–4452.
11. Seiti H, Hafezalkotob A and Martinez L. R-Sets, comprehensive fuzzy sets risk modeling for risk-based information fusion and decision-making. IEEE T Fuzzy Syst 2021; 29(2): 385–399.
12. Mou X, Zhang Q-Z, Liu H-C, et al. Knowledge representation and acquisition using R-numbers Petri nets considering conflict opinions. Expert Syst 2021; 38(3): 604–617.
13. Dempster AP. Upper and lower probabilities generated by a random closed interval. Ann Math Statist 1968; 39(3): 957–966.
14. Wang J and Liu FX. Temporal evidence combination method for multi-sensor target recognition based on DS theory and IFS. J Syst Eng Eng 2017; 28(6): 1114–1125.
15. Liu M and Chen S. SAR target configuration recognition based on the Dempster–Shafer theory and sparse representation using a new classification criterion. Int J Remote Sens 2019; 40(11–12): 4604–4622.
16. Chen L, Diao L and Sang J. A novel weighted evidence combination rule based on improved entropy function with a diagnosis application. Int J Distrib Syst N 2019; 15(1): 1–13.
17. Xiao F, Cao Z and Jolfaei A. A novel conflict measurement in decision-making and its application in fault diagnosis. IEEE T Fuzzy Syst 2021; 29(1): 186–197.
18. Deneux T and Shenoy PP. An interval-valued utility theory for decision making with Dempster-Shafer belief functions. Int J Approx Reason 2020; 124: 194–216.
19. Zhang R, Xu Z and Gou X. An integrated method for multi-criteria decision-making based on the best-worst method and Dempster-Shaifer evidence theory under double hierarchy hesitant fuzzy linguistic environment. Appl Intell 2021; 51(2): 713–735.
20. Pan Y, Zhang L, Wu X, et al. Multi-classifier information fusion in risk analysis. Inform Sci 2020; 60: 121–136.
21. Pan Y, Zhang L, Li Z, et al. Improved fuzzy Bayesian network-based risk analysis with interval-valued fuzzy sets and D–S evidence theory. IEEE T Fuzzy Syst 2020; 28(9): 2063–2077.
22. Song Y and Deng Y. A new method to measure the divergence in evidential sensor data fusion. Int J Distrib Syst N 2019; 15(4): 1–8.
23. Smet P. The combination of evidence in the transferable belief model. IEEE T Pattern Anal 1990; 12(5): 447–458.
24. Lefevre E and Elouedi Z. How to preserve the conflict as an alarm in the combination of belief functions? Decis Support Syst 2013; 56(12): 326–333.
25. Yong D, WenKang S, ZhenFu Z, et al. Combining belief functions based on distance of evidence. Decis Support Syst 2004; 38(3): 489–493.
26. Xiao F. CED: a distance for complex mass functions. *IEEE T Neur Net Lear* 2020; 32: 1525–1535.
27. Xiao F. Multi-sensor data fusion based on the belief divergence measure of evidences and the belief entropy. *Inform Fusion* 2019; 46: 23–32.
28. Xiao F. A new divergence measure for belief functions in D–S evidence theory for multisensor data fusion. *Inform Sci* 2020; 514: 462–483.
29. Wang H, Deng X, Jiang W, et al. A new belief divergence measure for Dempster–Shafer theory based on belief and plausibility function and its application in multi-source data fusion. *Eng Appl Artif Intel* 2021; 97: 104030.
30. Lin J. Divergence measures based on the Shannon entropy. *IEEE T Inform Theory* 1991; 37(1): 145–151.
31. Zhang H and Deng Y. Entropy measure for orderable sets. *Inform Sci* 2021; 561: 141–151.
32. Honghui Xu and Yong Deng. Dependent evidence combination based on decision-making trial and evaluation laboratory method. *Int J Intell Syst* 2019; 34(7): 1555–1571.
33. He Z and Wen J. An evidential Markov decision making model. *Inform Sci* 2018; 467: 357–372.
34. Shannon CE. A mathematical theory of communication. *Bell Syst Tech J* 2001; 5(3): 3–55
35. Deng Y. Deng entropy. *Chaos Soliton Fract* 2016; 91: 549–553.
36. Smets P and Kennes R. The transferable belief model. *Artif Intell* 1994; 66(2): 191–234.
37. Kullback S. *Information Theory and Statistics*. Mineola, NY: Dover Publication, 1997
38. Campbell LL. On measures of information and their characterizations. *SIAM Review* 1977; 19(3): 577–579.
39. Li H and Xiao F. A method for combining conflicting evidences with improved distance function and Tsallis entropy. *Int J Intell Syst* 2020; 35(11): 1814–1830.