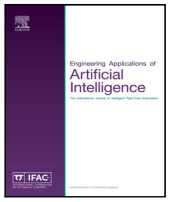
Engineering Applications of Artificial Intelligence 108 (2022) 104584

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai



A generalized divergence of information volume and its applications

Xiaozhuan Gao [a,d], Lipeng Pan [a], Yong Deng [a,b,c,d,*]

[a] Institute of Fundamental and Frontier Science, University of Electronic Science and Technology of China, Chengdu, 610054, China
[b] School of Education, Shaanxi Normal University, Xi'an, 710062, China
[c] School of Knowledge Science, Japan Advanced Institute of Science and Technology, Nomi, Ishikawa 923-1211, Japan
[d] Department of Management, Technology, and Economics, ETH Zurich, Zurich, Switzerland

ARTICLE INFO

Keywords:
Dempster–Shafer evidence theory
Uncertainty
Jensen–Shannon divergence
Information volume of mass function
Information fusion

ABSTRACT

Dempster–Shafer evidence theory provides a powerful method for the expression and fusion of uncertain information. When handling the high conflict information, traditional Dempster combination rule can produce counterintuitive results. Hence, the reasonable conflict measure is essential in information fusion. Inspired by this view, the paper propose the new method to measure conflict between bodies of evidence. Firstly, we define a new information volume of mass function for the perspective of information discord and non-specificity. Second, we propose a generalized divergence based on information volume of mass function, denoted as Jensen–Shannon divergence of information volume (IJS). IJS can effectively measure the conflict between bodies of evidence. IJS reflects the conflict between bodies of evidence in terms of the differences between the support of propositions and the elements. That is, compared to the current approach, IJS not only fully considers the differences between the support degree of propositions, but also the differences of elements in propositions from the perspective of information non-specificity. When the mass function degenerates to a probabilistic distribution, IJS also degenerates to the classical Jensen–Shannon divergence. Meanwhile, IJS also satisfies the axioms of distance measure, such as non-negativity, symmetry and etc. Further, we proved these axioms based on mathematical derivation, and some numerical examples are applied to explain axioms and advantages. Based on the proposed divergence measure, we propose a multi-source information fusion method in the real world, and several data sets can be used to show that the proposed fusion method is superior to current method under the framework of evidence theory.

1. Introduction

The uncertain information can be seen everywhere in real society, how to handle uncertain information is one of the key and hot issues in recent years (Yager, 2018; Marchau et al., 2019; Li et al., 2021; Fu et al., 2020). Various theories for modelling and processing uncertain information have been proposed, such as probability theory (Jaynes, 2003), fuzzy set (Zadeh, 1996), non-standard fuzzy set (Atanassov, 1986; Yager and Abbasov, 2013; Pan et al., 2021), rough set (Pawlak, 1982) and entropy measure (Deng, 2021). These theories have their own advantages and are widely used in medical diagnosis (Cao et al., 2019; Tian et al., 2021), group decision-making (Capuano et al., 2017; Gao et al., 2021a; Khan et al., 2021), target classification (de Souza et al., 2019; Vylegzhanin et al., 2019), cluster analysis (Liu et al., 2021a; Miyamoto, 2012; Dang et al., 2019; Ding et al., 2021; Jiang et al., 2021) and other fields (Dzitac et al., 2017; Yasmin et al., 2020; Gao et al., 2021b; Kamacı, 2021; Deng et al., 2021a). To solve more complex problems, these theories are further extended (Xiao, 2021a). Taking probability as an example, it is generalized to Dempster–Shafer evidence theory (Dempster, 1966; Shafer, 1976). In contrast to probability theory, the support of propositions is no longer a point estimate, rather than interval estimate by using the power sets whose meaning has been explained (Song and Deng, 2021). Meanwhile, Dempster–Shafer evidence theory has larger fault tolerance. Therefore, evidence theory is widely used in the fields of information fusion (Wang et al., 2021; Xiao, 2021b), reliability analysis (Chen and Zhang, 2021; Wang and Zhang, 2021), evidential reasoning (Fan et al., 2018; Ng and Law, 2020; Liao et al., 2020), pattern recognition (Xiao, 2020) and etc (Huang et al., 2021; Zhou et al., 2020; Zhao and Deng, 2021).

However, the conflict measure of evidence theory has limited its further applications due it cannot handle the high conflict information (Zadeh, 1986; Murphy, 2000; Roquel et al., 2014). The degree of conflict between bodies of evidence determines whether the Dempster combination rule can effectively combine information. In order to measure conflicts rationally, a series of methods have been proposed. Jousselme et al. proposed the classical evidence distance based on the Marcian matrix and the Jaccard coefficient, which initially considers

* Corresponding author at: Institute of Fundamental and Frontier Science, University of Electronic Science and Technology of China, Chengdu, 610054, China.
E-mail address: dengentropy@uestc.edu.cn (Y. Deng).

https://doi.org/10.1016/j.engappai.2021.104584
Received 19 August 2021; Received in revised form 7 November 2021; Accepted 30 November 2021
Available online 16 December 2021

Table 1
The various conflict measures in Example 1.1.

mass functions	k (Dempster, 1966; Shafer, 1976)	d_{BPA} (Jousselme et al., 2001)	k_r (Jiang, 2018)	BJS (Xiao, 2019)
m_1, m_2	0	0.7746	0.4	1
m_3, m_4	0	0.7746	0.4	1

the non-specificity between the mass functions (Jousselme et al., 2001). Jiang proposed correlation coefficients which take into account the support of the proposition and the differences in the elements of the proposition from the perspective of the cosine similarity (Jiang, 2018). Xiao proposed the belief Jensen–Shannon divergence based on the Shannon entropy (Xiao, 2019), and based on this, Xiao proposed an reinforced Jensen–Shannon divergence measure which takes into account the differences in elements between propositions. These methods can effectively measure conflict between bodies of evidence in a number of cases. In some cases it is not valid, e.g. Example 1.1.

Example 1.1. Suppose that there is a frame of discernment $\mathbb{E} = \{E_1, E_2, E_3, E_4, E_5\}$, four mass functions are described as follows:

$m_1 : m_1(\{E_1, E_2\}) = 1$; $m_2 : m_2(\{E_1, E_2, E_3, E_4, E_5\}) = 1$

$m_3 : m_3(\{E_1, E_2, E_3\}) = 1$; $m_4 : m_4(\{E_2, E_3, E_4, E_5\}) = 1$.

The conflict measure results of m_1 and m_2, m_3 and m_4 are shown in Table 1. The table shows that the difference between m_1 and m_2 is equal to the difference between m_3 and m_4. Clearly, those results are counter-intuitive in terms of the support of the proposition. The reason is that these methods do not fully consider the non-specificity of mass functions. Therefore, how to reasonably measure the conflict between bodies of evidence based on non-specificity of mass functions remains an open issue, which motivates this work.

In this article, we define a new information volume of mass function. It fully reflects the discord and non-specificity of mass function. Then, we present the generalized Jensen–Shannon divergence based on the information volume of mass function (IJS). *IJS* can adequately reflect differences in propositional support and differences in the elements of the proposition. i.e. it can effectively measure differences between bodies of evidence based on the discord and non-specificity of mass function. It also satisfies the axiom of distance measure, such as nonnegativity, nonnegativity, and etc. Meanwhile, it is compatible. When the mass function degenerates to the probability distribution, *IJS* also degenerates to the classical Jensen–Shannon divergence. The axioms and advantages of *IJS* are discussed by a number of numerical examples. Additionally, based on *IJS*, we propose a multi-source information fusion method in a data set environment. The results of the data set experiments indicate that the proposed method has a high target recognition rate compared to other methods.

The main works of this article are summarized as follows:

- We propose a generalized Jensen–Shannon divergence based on the information volume of mass function, which satisfies some axioms of distance measure. In addition, the proposed divergence is superior to other methods when measuring the difference between bodies of evidence.
- We also propose a multi-source information fusion method in the framework of evidence theory, and the results of the data set experiments show that the proposed fusion method has a higher target recognition rate by comparing with other methods.

The remainders of this work are organized as follows. Section 2 introduces the preliminaries of evidence theory and the information volume of the mass function. In Section 3, we propose a new information volume of the mass function, based on it, a generalized Jensen–Shannon divergence is proposed. Meanwhile, we prove its properties. Section 4 discusses the properties and advantages of the proposed divergence by some numerical examples. Section 5 presents a new multi-source information fusion method in the framework of evidence theory. We summarize this work in Section 6.

2. Preliminaries

In this section, we introduce some basic knowledge of evidence theory and the information volume of the mass function.

2.1. Dempster–Shafer Evidence theory

The evidence theory is proposed by Dempster and further developed by Shafer. It was first applied to expert systems, also applied to pattern recognition (Liu et al., 2020), artificial intelligence and system decision making (Chen et al., 2021; Xu et al., 2018). The relevant knowledge of the evidence theory is presented below (Dempster, 1966; Shafer, 1976).

Definition 2.1 (*Frame of Discernment*)**.** For a set E, which includes mutually exclusive, positive definite and finite events, denoted frame of discernment, is described as follows.

$$E = \{E_1, E_2, \ldots, E_N\} \tag{1}$$

Its power set can be composed by 2^N elements, which is as follows:

$$2^E = \{\varnothing, \{E_1\}, \{E_2\}, \ldots, \{E_N\}, \{E_1, E_2\}, \ldots, E\} \tag{2}$$

The $\varnothing$ is a empty set, and $\{\cdot\}$ denotes the element in the proposition.

Definition 2.2 (*Mass Function*)**.** When the function $m : 2^E \to [0, 1]$ satisfies the following two conditions

$$m(\varnothing) = 0, \text{ and } \sum_{F_j \in 2^E} m(F_i) = 1 \tag{3}$$

where $m(F_i)$ is denoted as mass function, represents the support degree of the evidence information for F_j.

Definition 2.3 (*Belief Function*)**.** Existing a mass function m, if $Bel : 2^{|E|} \to [0, 1]$ satisfies that

$$Bel(A) = \sum_{B \subseteq A} m(B), A \in 2^E$$

where $Bel(A)$ is the belief measure of proposition A, also called belief function.

Definition 2.4 (*Plausibility Function*)**.** For a mass function m, if $Pl : 2^{|E|} \to [0, 1]$ satisfies follow conditions

$$Pl(A) = \sum_{A \cap B} m(B)$$

where $Pl(A)$ is the plausible measure, denoted as plausibility function.

Definition 2.5 (*Dempster Combination Rule*)**.** For two mass functions m_1 and m_2 in frame of discernment E, the $m_1 \bigoplus m_2$ denotes the orthogonal sum of m_1 and m_2, which is named the Dempster combination rule. It has the following mathematical form.

$$m(F_k) = \begin{cases} \frac{\sum_{i,j:F_i \cap F_j = F_k} m_1(F_i) m_2(F_j)}{1-k} & F_k \neq \varnothing \\ 0 & F_k = \varnothing \end{cases} \tag{4}$$

where $k = \sum_{i,j:F_i \cap F_j = \varnothing} m_1(F_i) m_1(F_j)$, denoted as conflict coefficient, and represents the degree of conflict between bodies of evidence. Due Dempster's combination rule can get the counterintuitive results in some conditions, there are some studies about conflict (Liu et al., 2021b; Ali et al., 2012; Song et al., 2020; Xiong et al., 2021).

2.2. Information volume of the mass function

In Deng (2020), Deng proposed the information volume of mass function based on Deng entropy, which can measure the more bigger uncertainty of mass function than Deng entropy. Deng entropy has attracted more and more researcher attention (Tuğal and Karcı, 2019; Kazemi et al., 2021; Buono and Longobardi, 2020). The information volume of mass function is calculated according to the following steps.

- Step 1: Input a mass function m^1.
- Step 2: Multi-subset propositions are split according to the following equation

$$m^i\left(A_j\right)=m^{i-1}\left(B_j\right)\frac{2^{\left|A_j\right|}-1}{\sum_{A_j\subseteq B_j}\left(2^{\left|A_j\right|}-1\right)} \tag{5}$$

  where $i=2,\ldots,r$, it is the number of iterations, and $\left|A_j\right|$ represents the cardinality of proposition A_j.
- Step 3: Calculating the uncertainty of m^r based on Deng entropy.

The more details about information volume of the mass function in Deng (2020). There are some new studies about information volume, such as fuzzy membership function (Deng and Deng, 2021), fractal dimension (Gao et al., 2021c).

3. A generalized divergence for information volume of mass function

Divergence plays an essential role in the uncertain information, which has been attracted many people attention and been used in information fusion (Han et al., 2014), decision making (Gou et al., 2017; Khalaj et al., 2020) and other fields (Xiao, 2021c; Versaci and Morabito, 2021; Rani et al., 2020; Song et al., 2019). In this section, we present a new information volume of mass function, based on it, the generalized divergence is proposed.

3.1. A new information volume of the mass function

Deng proposed an information volume of the mass function based on the maximum Deng entropy. In some cases, the distribution of the maximum Deng entropy does not reflect the true situation of the proposition, so in this sub-section, we propose a new information volume for the mass function. The new information volume of mass function is calculated as follows.

- Step 1: Input a mass function m^1.
- Step 2: Multi-subset propositions are split according to the following equation

$$m^i\left(A_j\right)=m^{i-1}\left(B_j\right)\frac{e^{m^1\left(A_j\right)}}{\sum_{A_j\subseteq B_j}e^{m^1\left(A_j\right)}} \tag{6}$$

  where $i=2,\ldots,r$
- Step 3: Calculating the uncertainty of m^r based on Deng entropy.

Obviously, $m^i\left(A_j\right)$ meets the following conditions:
(1) When m^1 is a probability distribution, $m^1=m^i$
(2) $Bel\left(A_j\right)\le m^1\left(A_j\right)+\cdots+m^i\left(A_j\right)+\cdots+m^r\left(A_j\right)\le Pl\left(A_j\right)$, and $Bel\left(A_j\right)$ denotes the belief function for proposition A_j. The $Pl\left(A_j\right)$ denotes the plausibility function for proposition A_j.

3.2. A generalized divergence based on information volume of mass function

The current conflict measure is not applicable to some special cases, such as Example 1.1. Therefore, based on the proposed information volume of the mass function, we define a new divergence which can be used to measure conflict.

Definition 3.1 (*Divergence Based on Information Volume of Mass Function*)**.**

Suppose that there are two mass function m_1^1 and m_2^1 in frame of discernment E, and the divergence measure between them is described

$$IJS^r\left(m_1,m_2\right)=\frac{1}{2}\sum_{i=1}^{r}\left[D_{KL}\left(m_1^i,\frac{m_1^i+m_2^i}{2}\right)+D_{KL}\left(m_2^i,\frac{m_1^i+m_2^i}{2}\right)\right] \tag{7}$$

where $D_{KL}\left(m_1^i,m_2^i\right)=\sum_{j=1}^{n}m_1^i\left(A_j\right)log_2\frac{m_1^i\left(A_j\right)}{m_2^i\left(A_j\right)}$. Based on the properties of logarithmic functions, $IJS\left(m_1,m_2\right)$ is re-expressed as follows.

$$IJS^r\left(m_1,m_2\right)=\sum_{i=1}^{r}\left[H\left(\frac{m_1^i+m_2^i}{2}\right)-\frac{1}{2}H\left(m_1^i\right)-\frac{1}{2}H\left(m_2^i\right)\right] \tag{8}$$

where $H\left(m_1^i\right)=-\sum_{j=1}^{n}m_1^i\left(A_j\right)log_2m_1^i\left(A_j\right)$. IJS can measure differences in propositional support, and also differences in elements of propositions. The larger the value of $IJS^r\left(m_1,m_2\right)$, the larger the differences between bodies of evidence, and the smaller the value of $IJS^r\left(m_1,m_2\right)$, the smaller the differences between bodies of evidence. When the mass function degenerates to a probability distribution, the IJS also degenerates to the classical Jensen–Shannon divergence. In addition, IJS has the following properties.

Property 1: Nonnegativity $IJS^r\left(m_1,m_2\right)\ge 0$;

Property 2: Symmetry $IJS^r\left(m_1,m_2\right)=IJS^r\left(m_2,m_1\right)$;

Property 3: Nondegeneracy $IJS^r\left(m_1,m_2\right)=0$ if and only if $m_1=m_2$;

Property 4: Boundary $0\le IJS^r\left(m_1,m_2\right)\le 1$;

Property 5: Triangular inequality $\sqrt{IJS^r\left(m_1,m_2\right)}\le\sqrt{IJS^r\left(m_1,m_3\right)}+\sqrt{IJS^r\left(m_2,m_3\right)}$.

Then, these four properties are demonstrated as follows.

Proof 1: For $IJS^r\left(m_1,m_2\right)$, it can be expressed as follows

$$IJS^r\left(m_1,m_2\right)=\frac{1}{2}\left[\sum_{i=1}^{r}\sum_{j=1}^{n}m_1^i\left(A_j\right)log_2\frac{m_1^i\left(A_j\right)}{\frac{m_1^i\left(A_j\right)+m_2^i\left(A_j\right)}{2}}+\sum_{i=1}^{r}\sum_{j=1}^{n}m_2^i\left(A_j\right)log_2\frac{m_2^i\left(A_j\right)}{\frac{m_1^i\left(A_j\right)+m_2^i\left(A_j\right)}{2}}\right]$$

Consider a convex function φ, which satisfies the satisfies Jensen inequality

$\mathbb{E}\left[\varphi\left(X\right)\right]\ge\varphi\left(\mathbb{E}\left[X\right]\right)$

It is obvious that the logarithmic function is a convex function. Therefore

$$\begin{aligned}IJS^r\left(m_1,m_2\right)&=\frac{1}{2}\left[\sum_{i=1}^{r}\sum_{j=1}^{n}m_1^i\left(A_j\right)log_2\frac{m_1^i\left(A_j\right)}{\frac{m_1^i\left(A_j\right)+m_2^i\left(A_j\right)}{2}}+\sum_{i=1}^{r}\sum_{j=1}^{n}m_2^i\left(A_j\right)log_2\frac{m_2^i\left(A_j\right)}{\frac{m_1^i\left(A_j\right)+m_2^i\left(A_j\right)}{2}}\right]\\&\ge\frac{1}{2}\left[log_2\sum_{i=1}^{r}\sum_{j=1}^{n}m_1^i\left(A_j\right)\frac{m_1^i\left(A_j\right)}{\frac{m_1^i\left(A_j\right)+m_2^i\left(A_j\right)}{2}}+log_2\sum_{i=1}^{r}\sum_{j=1}^{n}m_2^i\left(A_j\right)\frac{m_2^i\left(A_j\right)}{\frac{m_1^i\left(A_j\right)+m_2^i\left(A_j\right)}{2}}\right]\\&=-\frac{1}{2}\left[log_2\sum_{i=1}^{r}\sum_{j=1}^{n}m_1^i\left(A_j\right)\frac{\frac{m_1^i\left(A_j\right)+m_2^i\left(A_j\right)}{2}}{m_1^i\left(A_j\right)}\right.\end{aligned}$$

$$+ log_2 \sum_{i=1}^{r}\sum_{j=1}^{n} m_2^i(A_j) \frac{\frac{m_1^i(A_j)+m_2^i(A_j)}{2}}{m_2^i(A_j)}\Bigg]$$
$$= -\frac{1}{2}\left[log_2\left(\frac{2}{2}\right) + log_2\left(\frac{2}{2}\right)\right]$$
$$= 0$$

To sum up, for any two mass functions m_1 and m_2, we obtain $IJS^r(m_1, m_2) \geq 0$ □

Thus, the symmetry of generalized divergence measure is proven.

Proof 2: Obviously,

$$IJS^r(m_1, m_2) = \sum_{i=1}^{r} H\left(\frac{m_1^i + m_2^i}{2}\right) - \frac{1}{2}H(m_1^i) - \frac{1}{2}H(m_2^i)$$
$$= \sum_{i=1}^{r} H\left(\frac{m_2^i + m_1^i}{2}\right) - \frac{1}{2}H(m_2^i) - \frac{1}{2}H(m_1^i)$$
$$= IJS^r(m_2, m_1) \quad \square$$

Thus, the nonnegativity of generalized divergence measure is proven.

Proof 3: Consider $m_1 = m_2$, then

$$IJS^r(m_1, m_2) = \sum_{i=1}^{r} H\left(\frac{m_1^i + m_2^i}{2}\right) - \frac{1}{2}H(m_1^i) - \frac{1}{2}H(m_2^i)$$
$$= \sum_{i=1}^{r} H(m_1^i) - \frac{1}{2}H(m_1^i) - \frac{1}{2}H(m_1^i)$$
$$= \sum_{i=1}^{r} H(m_2^i) - \frac{1}{2}H(m_2^i) - \frac{1}{2}H(m_2^i)$$
$$= 0$$

When $IJS^r(m_1, m_2) = 0$, next

$$\left[\sum_{i=1}^{r}\sum_{j=1}^{n} m_1^i(A_j) log_2 \frac{m_1^i(A_j)}{\frac{m_1^i(A_j)+m_2^i(A_j)}{2}} + \sum_{i=1}^{r}\sum_{j=1}^{n} m_2^i(A_j) log_2 \frac{m_2^i(A_j)}{\frac{m_1^i(A_j)+m_2^i(A_j)}{2}}\right] = 0$$

we know that

$$\sum_{i=1}^{r}\sum_{j=1}^{n} m_1^i(A_j) log_2 \frac{m_1^i(A_j)}{\frac{m_1^i(A_j)+m_2^i(A_j)}{2}} = 0$$
$$\sum_{i=1}^{r}\sum_{j=1}^{n} m_2^i(A_j) log_2 \frac{m_2^i(A_j)}{\frac{m_1^i(A_j)+m_2^i(A_j)}{2}} = 0$$

Hence, it can be obtained that

$m_1 = m_2$. □

Hence, the nondegeneracy of generalized divergence measure is proven.

Proof 4: We know from Property 1 that $IJS^r(m_1, m_2) \geq 0$. Obviously, the $IJS^r(m_1, m_2) = 0$ if and only if $m_1 = m_2$. Next, we explain the upper bound of $IJS^r(m_1, m_2)$.

Suppose there is a pattern classification problem with $P = \{p_1, p_2\}$, and a priori mass function $m(p_1) = \alpha_1$, $m(p_2) = \alpha_2$. $m(A_j|p_1) = m_1^i(A_j), (i = 1, \ldots, r)$ and $m(A_j|p_2) = m_2^i(A_j), (i = 1, \ldots, r)$ are the corresponding conditional mass functions in E. Then, the mass function of the error is expressed as follows (Wong and You, 1985; Lin, 1991).

$$M_e(m_1, m_2) = \sum_{i=1}^{r}\sum_{A_j \subseteq E} min\{\alpha_1 m_1^i(A_j), \alpha_2 m_2^i(A_j)\}$$

From the (Wong and You, 1985; Lin, 1991) we know that

$M_e(m_1, m_2) \leq \frac{1}{2}H(P|E)$

Based on conditional entropy (Lin, 1991), it is known that

$H(P|E) = H(P+) + H(E|C) - H(E)$

Obviously,

$$H(P) = -(\alpha_1 log_2 \alpha_1 + \alpha_2 log_2 \alpha_2)$$

$$H(E|P) = \sum_{i=1}^{r} \alpha_1 H(m_1^i) + \alpha_2 H(m_2^i)$$

$$H(E) = \sum_{i=1}^{r} H(\alpha_1 m_1^i + \alpha_1 m_2^i)$$

where $H(m_1^i) = -\sum_{A_j \subseteq E} m_1^i(A_j) log_2 m_1^i(A_j)$, thus,

$$M_e(m_1, m_2) \leq \frac{1}{2}\left[H(P) + \sum_{i=1}^{r} \alpha_1 H(m_1^i) + \alpha_2 H(m_2^i) - \sum_{i=1}^{r} H(\alpha_1 m_1^i + \alpha_1 m_1^i)\right]$$

$$M_e(m_1, m_2) \leq \frac{1}{2}\left[H(P) - IJS^r(m_1, m_2)\right]$$

$$IJS^r(m_1, m_2) \leq H(P) - 2M_e(m_1, m_2)$$

$$IJS^r(m_1, m_2) \leq H(P) - 2M_e(m_1, m_2) \leq H(P)$$

When $\alpha_1 = \alpha_2 = \frac{1}{2}$, then

$$IJS^r(m_1, m_2) \leq H(P) - 2M_e(m_1, m_2) \leq H(P) = 1$$

Obviously, the upper bound of $IJS^r(m_1, m_2)$ is 1, that is to say, $M_e(m_1, m_2) = 0$. It is also known from $M_e(m_1, m_2) = \sum_{i=1}^{r}\sum_{A_j \subseteq E} min\{\alpha_1 m_1^i(A_j), \alpha_2 m_2^i(A_j)\}$ that the elements of a proposition with support larger than 0 are completely free of intersections in m_1 and m_2. At this point, the difference between m_1 and m_2 is maximal. The value of $IJS^r(m_1, m_2)$ has reached the upper bound, i.e., $IJS^r(m_1, m_2) = 1$. □

Hence, the boundary of generalized divergence measure is proven.

Proof 5: For property 5, we verified it based on the idea of disproof. Assuming $\sqrt{IJS^r(m_1, m_2)} > \sqrt{IJS^r(m_1, m_3)} + \sqrt{IJS^r(m_2, m_3)}$, it contains the following three cases.

$$\begin{cases} m_1 = m_2 = m_3 & ① \\ m_1 = m_2 \neq m_3 \ or \ m_1 = m_3 \neq m_2 \ or \ m_2 = m_3 \neq m_1 & ② \\ m_1 \neq m_2 \neq m_3 & ③ \end{cases}$$

For Case ①, when $m_1 = m_2 = m_3$,

$$\sqrt{IJS^r(m_1, m_2)} = \sqrt{IJS^r(m_1, m_3)} + \sqrt{IJS^r(m_2, m_3)}$$

$0 = 0 + 0$

Obviously, this result is contradictory to the assume.

$$\sqrt{IJS^r(m_1, m_2)} = \sqrt{IJS^r(m_1, m_3)} + \sqrt{IJS^r(m_2, m_3)}$$

For Case ②, when $m_1 = m_2 \neq m_3$, then

$$0 > \sqrt{IJS^r(m_1, m_3)} + \sqrt{IJS^r(m_2, m_3)}$$

$$0 > \sqrt{IJS^r(m_1, m_3)} \ or \ 0 > \sqrt{IJS^r(m_2, m_3)}$$

It violates Property 1. The assumption is not reasonable in Case ②. Besides, supposing $\sqrt{IJS^r(m_1, m_3)} + \sqrt{IJS^r(m_2, m_3)} = \sqrt{IJS^r}_{1323}$, $\sqrt{IJS^r(m_1, m_2)} + \sqrt{IJS^r(m_2, m_3)} = \sqrt{IJS^r}_{1223}$.

For Case ①, when $m_1 \neq m_2 \neq m_3$, then

$$\sqrt{IJS^r(m_1, m_2)} > \sqrt{IJS^r}_{1323}, \quad \sqrt{IJS^r(m_1, m_3)} > \sqrt{IJS^r}_{1223}$$

Adding the above two inequalities together,

$$\sqrt{IJS^r(m_1, m_2)} + \sqrt{IJS^r(m_1, m_3)} > \sqrt{IJS^r}_{1323} + \sqrt{IJS^r}_{1223}$$

$$0 > \sqrt{IJS^r(m_2, m_3)}$$

Similar, this result violates Property 1. In all three cases, the assumptions are unreasonable. That is, the following formula is satisfied

$$\sqrt{IJS^r\left(m_1,m_2\right)} \le \sqrt{IJS^r\left(m_1,m_3\right)} + \sqrt{IJS^r\left(m_2,m_3\right)}.$$

The proof of property 5 completes. □

4. Numerical examples

In this section, some numerical examples are used to explain and verify the properties and characteristics of the proposed divergence.

Example 4.1. Suppose there is a frame of discernment $E = \{E_1, E_2, E_3\}$, m_1 and m_2 are two mass functions of the following form.

$m_1\left(E_1\right) = 0.1$, $m_1\left(E_2\right) = 0.1$, $m_1\left(E_3\right) = 0.2$, $m_1\left(E_1,E_2\right) = 0.1$, $m_1\left(E_1,E_3\right) = 0.1$, $m_1\left(E_2,E_3\right) = 0.2$, $m_1\left(E_1,E_2,E_3\right) = 0.2$;

$m_2\left(E_1\right) = 0.2$, $m_2\left(E_2\right) = 0.2$, $m_2\left(E_3\right) = 0.1$, $m_2\left(E_1,E_2\right) = 0.1$, $m_2\left(E_1,E_3\right) = 0.1$, $m_2\left(E_2,E_3\right) = 0.2$, $m_2\left(E_1,E_2,E_3\right) = 0.1$.

In Example 4.1, m_1 and m_2 are two different mass functions. Specifically, $m_1\left(E_1\right) = 0.1$ and $m_2\left(E_1\right) = 0.2$, $m_1\left(E_2\right) = 0.1$ and $m_2\left(E_2\right) = 0.2$, $m_1\left(E_3\right) = 0.2$ and $m_2\left(E_3\right) = 0.1$, however, the support degree of the rest propositions is similar. Based on Eq. (7), the difference between m_1 and m_2 is as follows.

$IJS^2\left(m_1,m_2\right) = 0.0490, IJS^3\left(m_1,m_2\right) = 0.0498, IJS^5\left(m_1,m_2\right) = 0.0499$

For different values of r, the value of $IJS^r\left(m_1,m_2\right)$ is always larger than 0, which preliminary explains Property 1. Next, the generalized divergence measure of m_2 and m_1 is computed as follows:

$IJS^1\left(m_2,m_1\right) = 0.0490, IJS^3\left(m_2,m_1\right) = 0.0498, IJS^5\left(m_2,m_1\right) = 0.0499$

Thus, the following formula is obtained

$IJS^r\left(m_1,m_2\right) = IJS^r\left(m_2,m_1\right)$

Hence, the Example 4.1 is also verified the symmetry of generalized divergence measure.

Example 4.2. There are two mass functions in the frame of discernment $E = \{E_1, E_2, E_3\}$ that are described as follows.

$m_1\left(E_1\right) = 0.1$, $m_1\left(E_1,E_3\right) = 0.25$, $m_1\left(E_2,E_3\right) = 0.35$, $m_1\left(E_1,E_2,E_3\right) = 0.3$;

$m_2\left(E_1\right) = 0.1$, $m_2\left(E_1,E_3\right) = 0.25$, $m_2\left(E_2,E_3\right) = 0.35$, $m_2\left(E_1,E_2,E_3\right) = 0.3$.

Obviously, m_1 and m_2 have the same belief value, where $m_1\left(E_1\right) = m_2\left(E_1\right) = 0.1$, $m_1\left(E_1,E_3\right) = m_2\left(E_1,E_3\right) = 0.25$, $m_1\left(E_2,E_3\right) = m_2\left(E_2,E_3\right) = 0.35$ and $m_1\left(E_1,E_2,E_3\right) = m_2\left(E_1,E_2,E_3\right) = 0.3$. Meanwhile, there are multiple subset propositions in m_1 and m_2. Under this case, the generalized divergence measure of m_2 and m_1 is calculated as follows

$IJS^1\left(m_1,m_2\right) = 0, IJS^3\left(m_1,m_2\right) = 0, IJS^5\left(m_1,m_2\right) = 0$

This result shows that when m_1 and m_2 have the same belief value, whether the proposition is single set or multiple subsets, the generalized divergence measure is also zero.

Example 4.3. For a frame of discernment $E = \{E_1, E_2\}$, the two mass functions in E have follow form:

$m_1\left(E_1\right) = \alpha$, $m_1\left(E_2\right) = \beta$, $m_1\left(E_1,E_2\right) = 1-\alpha-\beta$;

$m_2\left(E_1\right) = \beta$, $m_2\left(E_2\right) = \alpha$, $m_2\left(E_1,E_2\right) = 1-\alpha-\beta$.

In this example, the values of α and β are shown in Fig. 1(a). The divergence between m_1 and m_2 is shown in Fig. 1. As you can see from Fig. 1, no matter what happens to α and β, $IJS^r\left(m_1,m_2\right)$ is always larger than or equal to 0. When α is equal to β, divergence measure is equal to 0 due m1 and m2 are the same. Further, it can be seen that it varies between [0,1]. Meanwhile, because α is symmetric with β, $IJS^r\left(m_1,m_2\right)$ is also symmetric with respect to the plane $X+Y=0$. This example fully verifies the first four properties of the generalized divergence measure in a dynamic environment. In addition, for any two mass functions, the divergence between them increases as r increases, which leads to higher resolution. So it can measure differences between bodies of evidence with high similarity.

Example 4.4. There is a frame of discernment $E = \{E_1, E_2, E_3\}$, m_1, m_2 and m_3 are three mass functions, described as

$m_1\left(E_1\right) = 0.0052$, $m_1\left(E_2\right) = 0.1736$, $m_1\left(E_3\right) = 0.2176$, $m_1\left(E_1,E_2\right) = 0.2513$,

$m_1\left(E_1,E_3\right) = 0.0155$, $m_1\left(E_2,E_3\right) = 0.1166$, $m_1\left(E_1,E_2,E_3\right) = 0.1503$;

$m_2\left(E_1\right) = 0.0211$, $m_2\left(E_2\right) = 0.3439$, $m_2\left(E_3\right) = 0.0982$, $m_2\left(E_1,E_2\right) = 0.2070$,

$m_2\left(E_1,E_3\right) = 0.3368$, $m_2\left(E_2,E_3\right) = 0.0667$, $m_2\left(E_1,E_2,E_3\right) = 0.0667$.

$m_3\left(E_1\right) = 0.0485$, $m_3\left(E_2\right) = 0.1642$, $m_3\left(E_3\right) = 0.0336$, $m_3\left(E_1,E_2\right) = 0.2276$,

$m_3\left(E_1,E_3\right) = 0.0037$, $m_3\left(E_2,E_3\right) = 0.2127$, $m_3\left(E_1,E_2,E_3\right) = 0.2948$.

In this example, the divergence measures between m_1, m_2 and m_3 are shown in Table 2. It can be seen from Table 2 that no matter how r changes, the proposed divergence always satisfies the triangle inequality in Example 4.4.

Example 4.5. For a frame of discernment $E = \{E_1, E_2\}$, the two mass functions are described as follows:

$m_1\left(E_1\right) = \alpha$, $m_1\left(E_2\right) = 0.55-\alpha$, $m_1\left(E_1,E_2\right) = 0.45$;

$m_2\left(E_1\right) = 0.33$, $m_2\left(E_2\right) = 0.33$, $m_2\left(E_1,E_2\right) = 0.34$.

The divergence between m_1 and m_2 is shown in Fig. 2. It can be seen from Fig. 2 that with the change of α, the value of $IJS^r\left(m_1,m_2\right)$ decreases and then increases, which is intuitive. When $\alpha = 0.33$, the value of $IJS^r\left(m_1,m_2\right)$ is the minimum, which is 0. In this case, m_1 and m_2 are the same. In addition, as r increases, $IJS^r\left(m_1,m_2\right)$ is getting bigger, which indicates that the resolution of $IJS^r\left(m_1,m_2\right)$ improves as the non-specificity of the information increases.

Furthermore, we calculate the divergence in Example 1.1 based on the proposed method, and the results are shown in Table 3. Table 3 shows that by comparing with other methods, the proposed divergence can effectively measure the difference between the mass functions. it gives a intuitive results, which indicate that the proposed method is superior to other methods.

5. Method and applications

In this section, we propose a new multi-source information fusion method based on proposed divergence measure under the framework of evidence theory. By using some real world data, the performance of the proposed fusion method can be verified though Monte Carlo Cross Validation and K-fold Cross Validation.

5.1. A new multi-source information fusion method

In this sub-section, based on generalized divergence measure, we propose a new multi-source information fusion method in data set environment, and briefly discuss the time complexity of this method.

Problem description: Suppose there is a data set containing N_c classes, each class contains N_a attributes, and each attribute has N_d data points. Therefore, based on the idea of power sets, the mass function of each attribute would contain $2^{N_c}-1$ propositions. Based on this, the proposed method is described as follows (or see Fig. 3).

- Step 1: Select $N\ (0 < N < N_d)$ data from each attribute of each class as the training set, and the remaining data sets as the test set.
- Step 2: Generates the optimal mass function of the target (Take the qth attribute of the pth class as a example).
  - Step 2.1: Calculate the variance of the corresponding proposition in each attribute, respectively, STD_{q_j} $(j = 1, \ldots, 2^{N_c}-1)$.
  - Step 2.2: Calculate the variance except for the ith data and for each of the attributes, denoted as $STD^{\neq i,}_{p-q_j}$ $(j = 1, \ldots, 2^{N_c}-1)$.

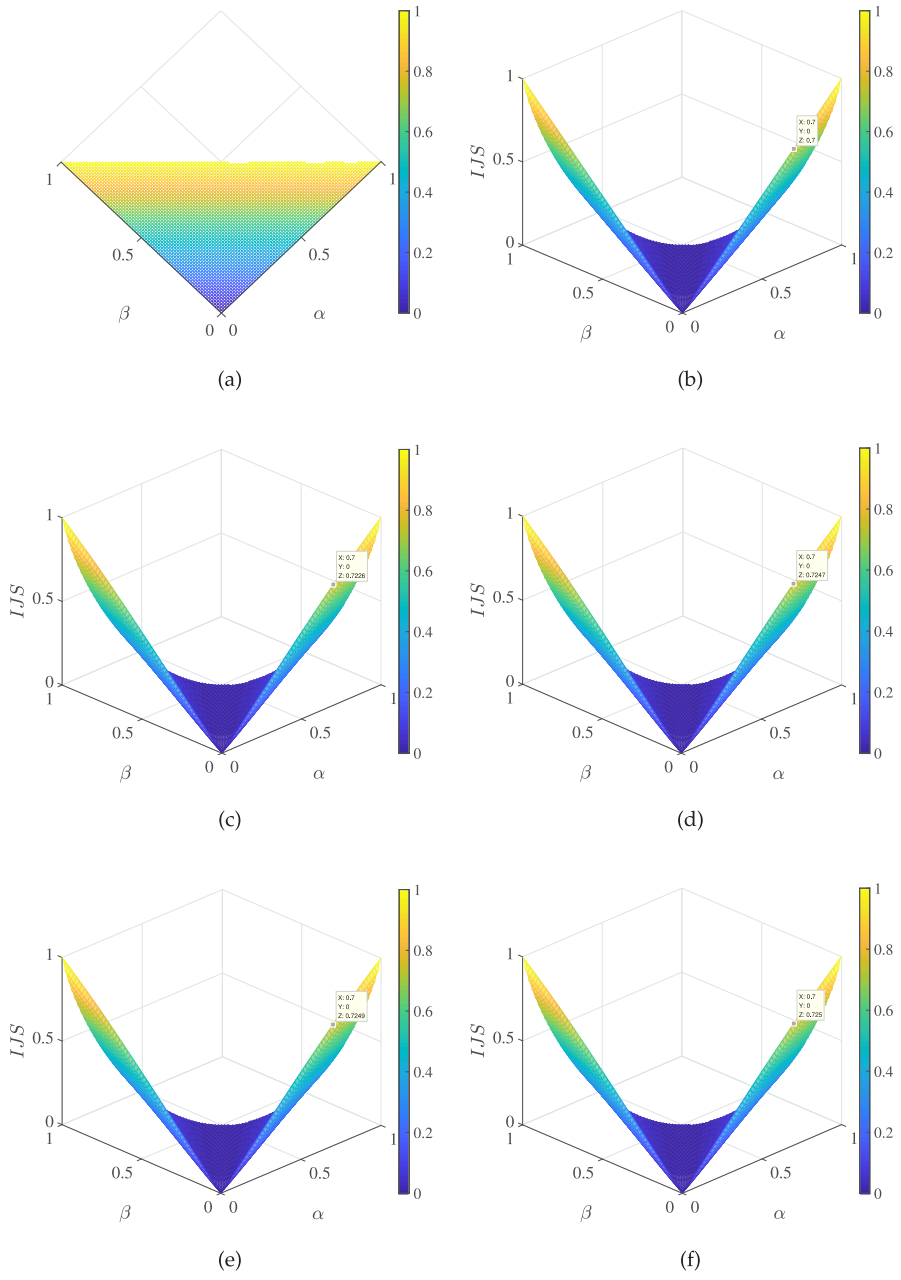


Fig. 1. The relationship between α and β and the divergence of m_1 and m_2 in different cases. (a): $0 \leq \alpha + \beta \leq 1$; (b): The divergence of m_1 and m_2 when $r = 1$; (c): The divergence of m_1 and m_2 when $r = 3$; (d): The divergence of m_1 and m_2 when $r = 5$; (e): The divergence of m_1 and m_2 when $r = 7$; (f): The divergence of m_1 and m_2 when $r = 9$.

Table 2
The divergence in Example 4.4.

i	$\sqrt{IJS^r(m_1, m_2)}$	$\sqrt{IJS^r(m_1, m_3)}$	$\sqrt{IJS^r(m_2, m_3)}$	$\sqrt{IJS^r(m_1, m_3)} + \sqrt{IJS^r(m_2, m_3)}$
$i = 1$	0.4355	0.3158	0.5294	0.8451
$i = 2$	0.4374	0.3174	0.5311	0.8485
$i = 3$	0.4376	0.3177	0.5313	0.8490

Table 3
The various conflict measures in Example 1.1.

mass functions	k (Dempster, 1966; Shafer, 1976)	d_{BPA} (Jousselme et al., 2001)	k_r (Jiang, 2018)	BJS (Xiao, 2019)	IJS
m_1, m_2	0	0.7746	0.4	1	0.3123
m_3, m_4	0	0.7746	0.4	1	0.0716

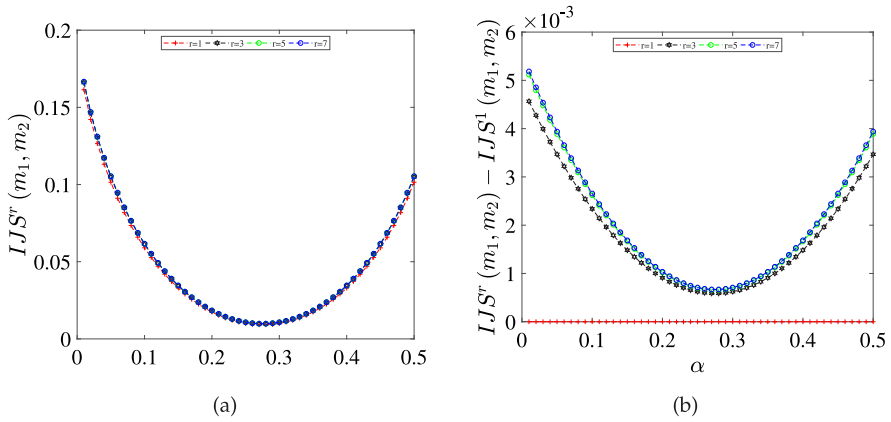


Fig. 2. (a): The divergence of m_1 and m_2 when α varies; (b): The divergence of m_1 and m_2 when r varies.

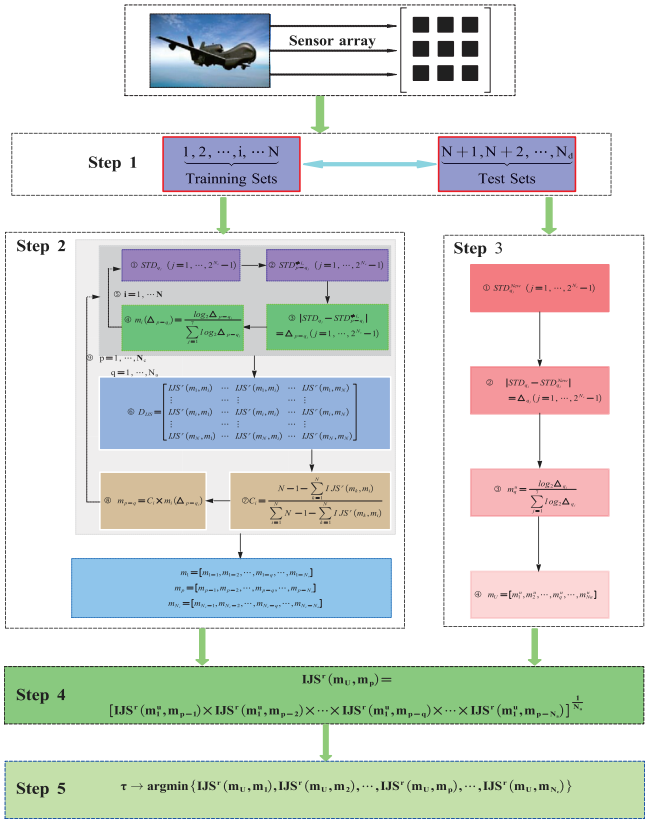


Fig. 3. The flow of the fusion method.

– Step 2.3: Calculate the difference between STD_{q_j} $(j = 1, \ldots, 2^{N_c} - 1)$ and $STD^{\neq i,}_{p-q_j}$ $\left(j = 1, \ldots, 2^{N_c} - 1\right)$.

$$\left|STD_{q_j} - STD^{\neq i,}_{p-q_j}\right| = \triangle_{p-q_j} \left(j = 1, \ldots, 2^{N_c} - 1\right) \tag{9}$$

– Step 2.4: Generates the mass function of the ith data (Gao and Deng, 2021).

$$m_i\left(\triangle_{p-q_j}\right) = \frac{log_2 \triangle_{p-q_j}}{\sum_{j=1}^{7} log_2 \triangle_{p-q_j}} \tag{10}$$

– Step 2.5: Repeat N times Step 2.2–Step 2.4, and obtain N mass functions of the training set.

– Step 2.6: Calculate the divergence matrix between the N mass functions.

$$D_{IJS} = \begin{bmatrix} IJS^r\left(m_1, m_i\right) & \cdots & IJS^r\left(m_1, m_i\right) & \cdots & IJS^r\left(m_1, m_N\right) \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ IJS^r\left(m_i, m_1\right) & \cdots & IJS^r\left(m_i, m_i\right) & \cdots & IJS^r\left(m_i, m_N\right) \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ IJS^r\left(m_N, m_1\right) & \cdots & IJS^r\left(m_N, m_i\right) & \cdots & IJS^r\left(m_N, m_N\right) \end{bmatrix} \tag{11}$$

– Step 2.7: Calculate the credibility of a body of evidence

$$C_i = \frac{N - 1 - \sum_{k=1}^{N} IJS^r\left(m_k, m_i\right)}{\sum_{i=1}^{N} N - 1 - \sum_{k=1}^{N} IJS^r\left(m_k, m_i\right)} \tag{12}$$

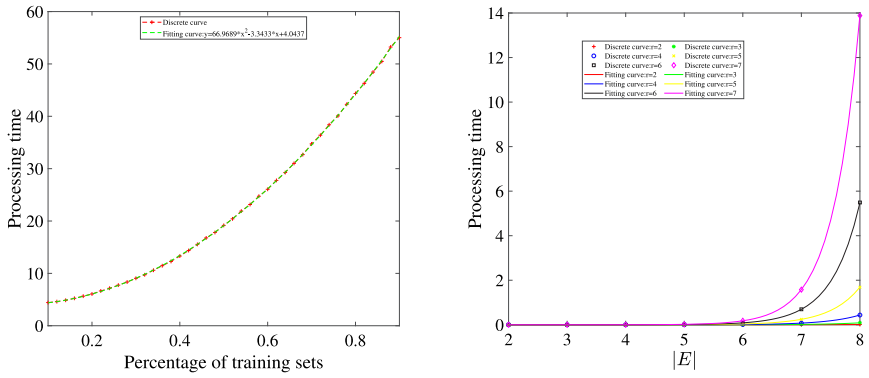


Fig. 4. The time complexity analysis. (a): The time complexity analysis of the proposed methodology; (b): The time complexity analysis of information volume.

– Step 2.8: Generate a new mass function based on evidence credibility, i.e., the optimal mass function.

$$m_{p-q} = C_i \times m_i \left(\triangle_{p-q_j} \right) \tag{13}$$

– Step 2.9: Generate the rest mass functions by repeating Step 2.1–Step 2.8 based on the data of other attributes. For the sake of description, noted as follows.

$$m_1 = \left[m_{1-1}, m_{1-2}, \ldots, m_{1-q}, \ldots, m_{1-N_a} \right]$$
$$m_p = \left[m_{p-1}, m_{p-2}, \ldots, m_{p-q}, \ldots, m_{p-N_a} \right]$$
$$m_{N_c} = \left[m_{N_c-1}, m_{N_c-2}, \ldots, m_{N_c-q}, \ldots, m_{N_c-N_a} \right]$$

- Step 3: Generates a mass function for an unknown target (Take the qth attribute as a example).

– Step 3.1: Calculate the new variance after adding the first test point, respectively, $STD_{q_j}^{New}$ $\left(j = 1, \ldots, 2^{N_c} - 1 \right)$.

– Step 3.2: Calculate the difference between STD_{q_j} and $STD_{q_j}^{\neq i,}$.

$$\left| STD_{q_j} - STD_{q_j}^{New} \right| = \triangle_{q_j} \left(j = 1, \ldots, 2^{N_c} - 1 \right) \tag{14}$$

– Step 3.3: Generates the mass function of unknown target.

$$m_q^u = \frac{log_2 \triangle_{q_j}}{\sum_{j=1}^{7} log_2 \triangle_{q_j}} \tag{15}$$

– Step 3.4: Repeat Step 3.1–Step 3.3 to generate the rest mass function based on the data of the remaining three attributes, denoted as follows.

$$m_U = \left[m_1^u, m_2^u, \ldots, m_q^u, \ldots, m_{Na}^u \right]$$

- Step 4: Calculate the divergence between the optimal mass functions and the mass functions of unknown target. $IJS^r \left(m_U, m_p \right) =$

$$\left[IJS^r \left(m_1^u, m_{p-1} \right) \times IJS^r \left(m_1^u, m_{p-2} \right) \times \cdots \right.$$
$$\left. IJS^r \left(m_1^u, m_{p-q} \right) \times \cdots \times IJS^r \left(m_1^u, m_{p-N_a} \right) \right]^{\frac{1}{N_a}} \tag{16}$$

- Step 5: Recognize unknown target.

$$\tau \rightarrow argmin \left\{ IJS^r \left(m_U, m_1 \right), IJS^r \left(m_U, m_2 \right), \right.$$
$$\left. \ldots, IJS^r \left(m_U, m_p \right), \ldots, IJS^r \left(m_U, m_{N_c} \right) \right\} \tag{17}$$

which indicates unknown target is recognized as the class with the minimum divergence.

In order to better explanation, the pseudo-code of the proposed method is shown in Algorithm 1.

Algorithm 1: Multi-source information fusion method

Input: Data set (N training sets, the rest are test sets).
Output: Recognition result τ

```
for p=1;p ≤ N_c do
    for q=1;q ≤ N_a do
        for i=1;i ≤ N do
            Calculate the mass function of the ith data by Step 2.1–Step 2.4;
        end
        Construct the divergence matrix is based on Equation 11;
        Calculate the credibility of each evidence based on Equation 12;
        Calculate optimal mass function of the q-th attribute of the p-th class based on Equation 13;
    end
end
for k=1;k ≤ N_d − N do
    for q=1;q ≤ N_a do
        Generate the mass function of the q-th attribute of the unknown target based on Step 3;
    end
end
The divergence between the unknown target and the known class is calculated based on Equations 16;
The unknown targets are recognized by Equation 17.
```

Next, we analyse the time complexity of this method. In this experiment, $N_c = 3$, $N_a = 4$ and $N_d = 50$. We randomly select the training set from 5 to 45, and the time complexity of this method is shown in Fig. 4. Fig. 4(a) shows the computational complexity of proposed fusion method. It can be seen from the figure that as the number of training sets increases, the running time of the code also increases. The computational complexity of the fusion method is $O\left(n^2\right)$. In addition, we also analysed the computational complexity of information volume in the proposed fusion method, as shown in Fig. 4(b). It can be seen that the computational complexity of information volume is affected by the size of the framework and the number of iterations. The more the number of iterations, the longer the run time. The larger the framework, the longer the running time. The computational complexity is exponential. Therefore, on the premise of obtaining the appropriate divergence measure, the number of iterations should be appropriately reduced, which in turn can reduce the running time of the proposed fusion method.

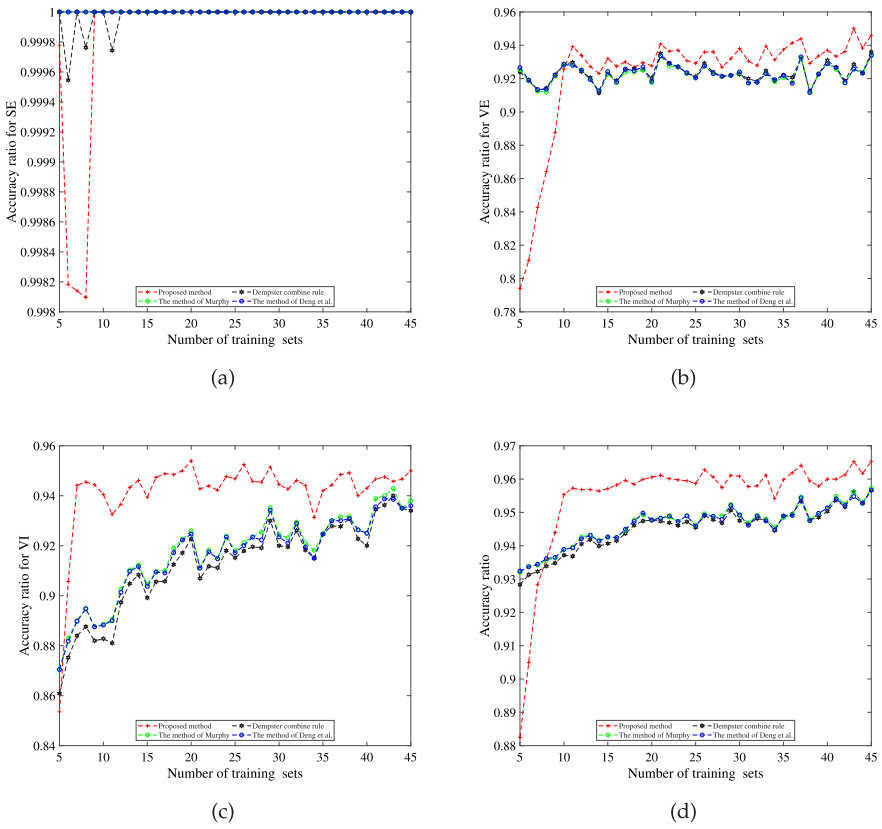


Fig. 5. Target recognition rate of different methods under the evidence framework when the number of training sets changes based on Monte Carlo validation. (a): The recognition rate of Setosa; (b): The recognition rate of Versicolour; (c): The recognition rate of Virginica; (d): Average recognition rate of the three types of flowers.

Table 4
The recognition results of different methods based on the five-fold cross-validation experiment.

Method	Setosa	Versicolour	Virginica	Average
Dempster combination rule (Dempster, 1966; Shafer, 1976)	1	0.915	0.895	0.9366
The method of Murphy (2000)	1	0.915	0.895	0.9366
The method of Deng et al. (2021b)	1	0.915	0.895	0.9366
Proposed method	1	0.93	0.96	0.9633

5.2. Application of iris recognition

In this subsection, we apply the proposed method to iris recognition to explain its performance by Monte Carlo validation and K-fold cross-validation. Iris data set comes from UCI database (http://archive.ics.uci.edu/ml/datasets/Iris), including three classes, namely Setosa, Versicolour and Virginica. Every class contains four attributes, namely SepalLength, SepalWidth, PetalLength and PetalWidth. Each attribute contains 50 data.

In Monte Carlo validation experiments, we randomly select the data set from 5 to 45 and run it 100 times. When $r = 3$, the experimental results are shown in Fig. 5. Fig. 5 shows the results of the four methods for the recognition of Iris. As can be seen in Fig. 5(a), the recognition rate of the proposed method for Setosa is low relative to several methods when the training set is small. As the number of training sets increases, the performance of the proposed method and other methods is the same for Setosa. For Versicolour, the recognition rate of the proposed method is higher than that of other methods when the number of training sets increases, as shown in Fig. 5(b). In addition, the proposed method also has a high recognition rate relative to Dempster combine rule, the method of Murphy and the method of Deng et al. when the unknown target is Virginica, as shown in Fig. 5(c). Further, we calculate the average recognition rate of the three flowers based on four methods as shown in Fig. 5(d). It can be seen from the figure that the proposed method has better performance than the other methods in the framework of evidence theory.

Meanwhile, we apply the proposed fusion method to the five-fold cross-validation experiment. One sample is selected in turn as the training set, and the remaining four samples are selected as the test set. The recognition results five-fold cross-validation experiment are shown in Table 4. It can be seen from the table that the recognition results of Setosa by the four methods are the same, and the recognition rates are all 1. This is because the attributes of Setosa are relatively easy to distinguish and can therefore be fully recognized. For Versicolour and Virginica, their attributes are not easily distinguishable. Compared with the other three methods, the proposed method has the highest recognition rate, which are 0.93 and 0.96 respectively. This result indicates that the proposed fusion method also has good performance in K-fold cross validation experiments.

5.3. Application of fault diagnosis

In this subsection, we apply the proposed method to the industrial field, taking the fault diagnosis of motor as an example. Wen et al. studied fault diagnosis using a multi-functional flexible rotor test platform in Ref. Wen et al. (2012). Rotor faults mainly include three kinds, respectively, rotor imbalance (RI), rotor misalignment (RM) and pedestal looseness (PL). Rotor unbalance is the phenomenon that centroid and centroid do not coincide in the process of rotor manufacturing. Rotor

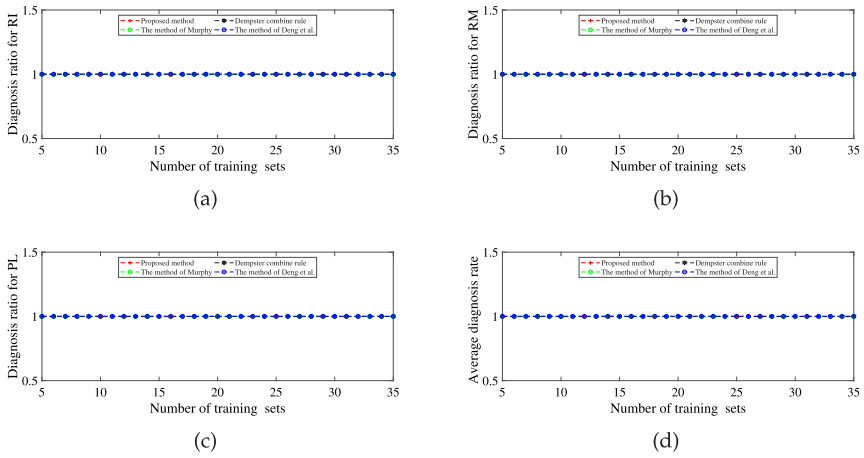


Fig. 6. The diagnosis results of different methods under the evidence framework when the number of training sets changes. (a): The diagnosis rate of RI ; (b): The diagnosis rate of RM; (c): The diagnosis rate of PL; (d): Average diagnosis rate of the three types of fault.

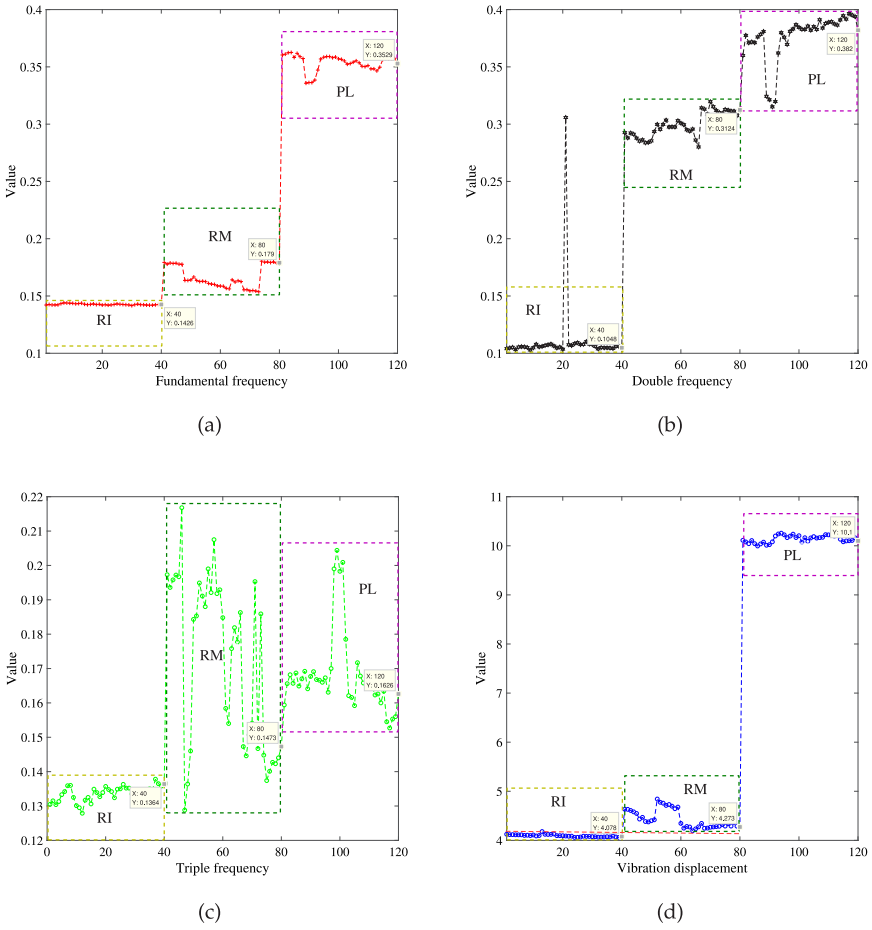


Fig. 7. Data of different faults in the same attribute. (a): Fundamental frequency ; (b): Double frequency; (c): Triple frequency; (d): Vibration displacement.

misalignment is the axis of each rotor is no longer the same straight line. Pedestal looseness is the loosening between the bearing and the base, which is generally caused by low quality or long-term vibration. Each fault contains four characteristics, fundamental frequency, double frequency, triple frequency and vibration displacement. Each characteristic contains 40 pieces of data.

Based on the idea of Monte Carlo validation, we randomly selected data from 5 to 35 at a time as the training set and the rest as the test set. Running 100 times Monte Carlo validation experiment. The result is shown in Fig. 6. It can be seen from the figure that the four methods have the maximum diagnosis rate for three kinds of faults, that is to say, the diagnosis rate of the four methods for three kinds of faults is 1. In the analysis of the fault diagnosis data set, it is found that the same attribute varies greatly among different faults. This resulted in the four methods having the highest diagnostic rates. To better explain, the four attributes based on the three failures are shown in Fig. 7. Based on this situation, we also design 4-fold cross-validation experiment. One sample is selected as the training set and the other three samples as the test set. Through the proposed method, the diagnosis results of rotor faults are shown in Table 5. As shown in Table 5, For the three

Table 5
The diagnosis results of different methods based on the five-fold cross-validation experiment.

Method	RI	RM	PL	Average
Dempster combination rule (Dempster, 1966; Shafer, 1976)	1	1	1	1
The method of Murphy (2000)	1	1	1	1
The method of Deng et al. (2021b)	1	1	1	1
Proposed method	1	1	1	1

faults, the proposed method has the same diagnostic rate as the other three methods, which is 1. This result shows that the performance of the proposed method is as good as that of the current method under the condition of large difference of attributes, which also indicates that proposed fusion method also has a certain potential in industrial applications.

6. Conclusion

As we all known, the tradition Dempster combination rule has counterintuitive results when handling the high conflict information. Inspired by this, the paper propose the generalized divergence of mass function to measure conflict. Firstly, we define a new information volume of the mass function from the point of view of information non-specificity. Then, we propose a generalized divergence based on the information volume of mass function to measure the conflict between bodies of evidence. The proposed divergence satisfies some axioms of divergence measures, such as nonnegativity, nonnegativity, and etc. Meanwhile, the proposed divergence can measure the difference between bodies of evidence more reasonably than the current methods. In addition, when the mass function degenerates to probability distribution, the proposed divergence also degenerates to classical divergence. Some numerical examples explain the properties and characteristics of the proposed divergence. Further, we propose a new multi-source information fusion method in the environments of data sets. Experimental results of Monte Carlo Cross Validation and K-fold Cross Validation indicate that the proposed method has a higher recognition rate for unknown targets compared to the other methods under the framework of evidence theory, which also shows that this method has some potential in industrial applications.

CRediT authorship contribution statement

Xiaozhuan Gao: Conceptualization, Methodology, Writing – original draft, Writing – review & editing. **Lipeng Pan:** Validation, Writing – review & editing. **Yong Deng:** Validation, Resources, Writing – review & editing, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The work is partially supported by National Natural Science Foundation of China (Grant No. 61973332), JSPS Invitational Fellowships for Research in Japan (Short-term).

References

Ali, T., Dutta, P., Boruah, H., 2012. A new combination rule for conflict problem of Dempster-Shafer evidence theory. Int. J. Energy Inf. Commun. 3 (1), 35–40.

Atanassov, K.T., 1986. Intuitionistic fuzzy sets. vol. 20, (1), Springer, pp. 87–96,

Buono, F., Longobardi, M., 2020. A dual measure of uncertainty: The deng extropy. Entropy 22 (5), 582.

Cao, Z., Chuang, C.-H., King, J.-K., Lin, C.-T., 2019. Multi-channel EEG recordings during a sustained-attention driving task. Sci. Data 6, http://dx.doi.org/10.1038/s41597--019--0027--4.

Capuano, N., Chiclana, F., Fujita, H., Herrera-Viedma, E., Loia, V., 2017. Fuzzy group decision making with incomplete information guided by social influence. IEEE Trans. Fuzzy Syst. 26 (3), 1704–1718.

Chen, L., Deng, Y., Cheong, K.H., 2021. Probability transformation of mass function: A weighted network method based on the ordered visibility graph. Eng. Appl. Artif. Intell. (ISSN: 0952-1976) 105, 104438.

Chen, W., Zhang, L., 2021. Resilience assessment of regional areas against earthquakes using multi-source information fusion. Reliab. Eng. Syst. Saf. 107833.

Dang, T.H., Mai, D.S., Ngo, L.T., 2019. Multiple kernel collaborative fuzzy clustering algorithm with weighted super-pixels for satellite image land-cover classification. Eng. Appl. Artif. Intell. 85, 85–98.

Dempster, A., 1966. Upper and lower probabilities induced by a multivalued mapping(upper and lower probabilities induced by multivalued mapping).

Deng, Y., 2020. Information volume of mass function. Int. J. Comput. Commun. Control 15 (6), 3983. http://dx.doi.org/10.15837/ijccc.2020.6.3983.

Deng, Y., 2021. Uncertainty measure in evidence theory. Sci. China Inf. Sci. 64, http://dx.doi.org/10.1007/s11432--020--3006--9.

Deng, X., Cui, Y., Jiang, W., 2021a. An ECR-pcr rule for fusion of evidences defined on a non-exclusive framework of discernment. Chin. J. Aeronaut. http://dx.doi.org/10.1016/j.cja.2021.06.004.

Deng, J., Deng, Y., 2021. Information volume of fuzzy membership function. Int. J. Comput. Commun. Control 16 (1), 4106. http://dx.doi.org/10.15837/ijccc.2021.1.4106.

Deng, J., Deng, Y., Cheong, K.H., 2021b. Combining conflicting evidence based on pearson correlation coefficient and weighted graph. Int. J. Intell. Syst. http://dx.doi.org/10.1002/int.22593.

Ding, W., Chakraborty, S., Mali, K., Chatterjee, S., Nayak, J., Das, A.K., Banerjee, S., 2021. An unsupervised fuzzy clustering approach for early screening of COVID-19 from radiological images. IEEE Trans. Fuzzy Syst. http://dx.doi.org/10.1109/TFUZZ.2021.3097806.

Dzitac, I., Filip, F.G., Manolescu, M.-J., 2017. Fuzzy logic is not fuzzy: World-renowned computer scientist lotfi a. zadeh. Int. J. Comput. Commun. Control 12 (6), 748–789.

Fan, C.-l., Song, Y., Lei, L., Wang, X., Bai, S., 2018. Evidence reasoning for temporal uncertain information based on relative reliability evaluation. Expert Syst. Appl. 113, 264–276.

Fu, C., Hou, B., Chang, W., Feng, N., Yang, S., 2020. Comparison of evidential reasoning algorithm with linear combination in decision making. Int. J. Fuzzy Syst. 22 (2), 686–711.

Gao, X., Deng, Y., 2021. Generating method of pythagorean fuzzy sets from the negation of probability. Eng. Appl. Artif. Intell. 105, http://dx.doi.org/10.1016/j.engappai.2021.104403.

Gao, X., Pan, L., Deng, Y., 2021a. Quantum pythagorean fuzzy evidence theory (qpfet): A negation of quantum mass function view. IEEE Trans. Fuzzy Syst. http://dx.doi.org/10.1109/TFUZZ.2021.3057993.

Gao, X., Su, X., Qian, H., Pan, X., 2021b. Dependence assessment in human reliability analysis under uncertain and dynamic situations. Nuclear Engineering and Technology http://dx.doi.org/10.1016/j.net.2021.09.045.

Gao, Q., Wen, T., Deng, Y., 2021c. Information volume fractal dimension. Fractals http://dx.doi.org/10.1142/S0218348X21502637.

Gou, X., Xu, Z., Liao, H., 2017. Hesitant fuzzy linguistic entropy and cross-entropy measures and alternative queuing method for multiple criteria decision making. Inform. Sci. 388, 225–246.

Han, D., Dezert, J., Yang, Y., 2014. Evaluations of evidence combination rules in terms of statistical sensitivity and divergence. In: 17th International Conference on Information Fusion (FUSION). IEEE, pp. 1–7.

Huang, C., Mi, X., Kang, B., 2021. Basic probability assignment to probability distribution function based on the Shapley value approach. Int. J. Intell. Syst. 36 (8), 4210–4236.

Jaynes, E.T., 2003. Probability Theory: The Logic of Science. Cambridge University Press.

Jiang, W., 2018. A correlation coefficient for belief functions. Internat. J. Approx. Reason. 103, 94–106.

Jiang, W., Huang, K., Geng, J., Deng, X., 2021. Multi-scale metric learning for few-shot learning. IEEE Trans. Circuits Syst. Video Technol. 31 (3), 1091–1102. http://dx.doi.org/10.1109/TCSVT.2020.2995754.

Jousselme, A.-L., Grenier, D., Bossé, E., 2001. A new distance between two bodies of evidence. Inf. Fusion 2 (2), 91–101.

Kamacı, H., 2021. Linguistic single-valued neutrosophic soft sets with applications in game theory. Int. J. Intell. Syst. http://dx.doi.org/10.1002/int.22445.

Kazemi, M.R., Tahmasebi, S., Buono, F., Longobardi, M., 2021. Fractional deng entropy and extropy and some applications. Entropy 23 (5), 623.
Khalaj, M., Tavakkoli-Moghaddam, R., Khalaj, F., Siadat, A., 2020. New definition of the cross entropy based on the Dempster-Shafer theory and its application in a decision-making process. Comm. Statist. Theory Methods 49 (4), 909–923.
Khan, Q., Mahmood, T., Ullah, K., 2021. Applications of improved spherical fuzzy dombi aggregation operators in decision support system. Soft Comput. 25, 9097–9119.
Li, X., Liao, H., Wen, Z., 2021. A consensus model to manage the non-cooperative behaviors of individuals in uncertain group decision making problems during the COVID-19 outbreak. Appl. Soft Comput. 99, 106879.
Liao, H., Ren, Z., Fang, R., 2020. A deng-entropy-based evidential reasoning approach for multi-expert multi-criterion decision-making with uncertainty. Int. J. Comput. Intell. Syst. 13 (1), 1281–1294.
Lin, J., 1991. Divergence measures based on the Shannon entropy. IEEE Trans. Inform. Theory 37 (1), 145–151.
Liu, Z., Liu, Y., Dezert, J., Cuzzolin, F., 2020. Evidence combination based on credal belief redistribution for pattern classification. IEEE Trans. Fuzzy Syst. 28 (4), 618–631.
Liu, Z.-G., Qiu, G., Mercier, G., Pan, Q., 2021a. A transfer classification method for heterogeneous data based on evidence theory. IEEE Trans. Syst. Man Cybern. Syst. 51 (8), 5129–5141. http://dx.doi.org/10.1109/TSMC.2019.2945808.
Liu, Z., Zhang, X., Niu, J., Dezert, J., 2021b. Combination of classifiers with different frames of discernment based on belief functions. IEEE Trans. Fuzzy Syst. 29 (7), 1764–1774. http://dx.doi.org/10.1109/TFUZZ.2020.2985332.
Marchau, V.A., Walker, W.E., Bloemen, P.J., Popper, S.W., 2019. Decision Making Under Deep Uncertainty: From Theory to Practice. Springer Nature.
Miyamoto, S., 2012. Fuzzy Sets in Information Retrieval and Cluster Analysis. vol. 4, Springer Science & Business Media.
Murphy, C.K., 2000. Combining belief functions when evidence conflicts. Decis. Support Syst. 29 (1), 1–9.
Ng, C., Law, K.M., 2020. Investigating consumer preferences on product designs by analyzing opinions from social networks using evidential reasoning. Comput. Ind. Eng. 139, 106180.
Pan, L., Gao, X., Deng, Y., Cheong, K.H., 2021. The constrained pythagorean fuzzy sets and its similarity measure. IEEE Trans. Fuzzy Syst. http://dx.doi.org/10.1109/TFUZZ.2021.3052559.
Pawlak, Z., 1982. Rough sets. Int. J. Comput. Inf. Sci. 11 (5), 341–356.
Rani, P., Mishra, A.R., Mardani, A., Cavallaro, F., Alrasheedi, M., Alrashidi, A., 2020. A novel approach to extended fuzzy TOPSIS based on new divergence measures for renewable energy sources selection. J. Cleaner Prod. 257, 120352.
Roquel, A., Le Hégarat-Mascle, S., Bloch, I., Vincke, B., 2014. Decomposition of conflict as a distribution on hypotheses in the framework on belief functions. Internat. J. Approx. Reason. 55 (5), 1129–1146.
Shafer, G., 1976. A Mathematical Theory of Evidence. vol. 42, Princeton University Press.
Song, Y., Deng, Y., 2021. Entropic explanation of power set. Int. J. Comput. Commun. Control 16 (4), 4413.
Song, Y., Fu, Q., Wang, Y.-F., Wang, X., 2019. Divergence-based cross entropy and uncertainty measures of atanassov's intuitionistic fuzzy sets with their application in decision making. Appl. Soft Comput. 84, 105703.
Song, Y., Zhu, J., Lei, L., Wang, X., 2020. Self-adaptive combination method for temporal evidence based on negotiation strategy. Sci. China Inf. Sci. 63 (11), 1–13.
de Souza, R.W.R., De Oliveira, J.a.V.C., Passos, L.A., Ding, W., Papa, J.a.P., de Albuquerque, V.H.C., 2019. A novel approach for optimum-path forest classification using fuzzy logic. IEEE Trans. Fuzzy Syst. 28 (12), 3076–3086.
Tian, Y., Mi, X., Cui, H., Zhang, P., Kang, B., 2021. Using Z-number to measure the reliability of new information fusion method and its application in pattern recognition. Appl. Soft Comput. 111, 107658.
Tuğal, İ., Karcı, A., 2019. Comparisons of karcı and Shannon entropies and their effects on centrality of social networks. Physica A 523, 352–363.
Versaci, M., Morabito, F.C., 2021. Image edge detection: A new approach based on fuzzy entropy and fuzzy divergence. Int. J. Fuzzy Syst. 1–19.
Vylegzhanin, I., Vovshin, B., Vylegzhanina, O., Pushkov, A., 2019. Fuzzy logic alghorithms for target classification in radar observations. In: 2019 International Conference on Engineering and Telecommunication (EnT). IEEE, pp. 1–3.
Wang, H., Deng, X., Jiang, W., Geng, J., 2021. 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. Intell. 97, 104030.
Wang, Y., Zhang, L., 2021. Feature-based evidential reasoning for probabilistic risk analysis and prediction. Eng. Appl. Artif. Intell. 102, 104237.
Wen, C., Xu, X., Jiang, H., Zhou, Z., 2012. A new dsmt combination rule in open frame of discernment and its application. Sci. China Inf. Sci. 55 (3), 551–557.
Wong, A.K., You, M., 1985. Entropy and distance of random graphs with application to structural pattern recognition. IEEE Trans. Pattern Anal. Mach. Intell. (5), 599–609.
Xiao, F., 2019. Multi-sensor data fusion based on the belief divergence measure of evidences and the belief entropy. Inf. Fusion 46, 23–32.
Xiao, F., 2020. Evidence combination based on prospect theory for multi-sensor data fusion. ISA Trans. 106, 253–261.
Xiao, F., 2021a. CaFtR: A Fuzzy complex event processing method. Int. J. Fuzzy Syst. http://dx.doi.org/10.1007/s40815--021--01118--6.
Xiao, F., 2021b. CEQD: A Complex mass function to predict interference effects. IEEE Trans. Cybern. http://dx.doi.org/10.1109/TCYB.2020.3040770.
Xiao, F., 2021c. A distance measure for intuitionistic fuzzy sets and its application to pattern classification problems. IEEE Trans. Syst. Man Cybern. Syst. 51 (6), 3980–3992.
Xiong, L., Su, X., Qian, H., 2021. Conflicting evidence combination from the perspective of networks. Inform. Sci. 580, 408–418. http://dx.doi.org/10.1016/j.ins.2021.08.088.
Xu, X., Xu, H., Wen, C., Li, J., Hou, P., Zhang, J., 2018. A belief rule-based evidence updating method for industrial alarm system design. Control Eng. Pract. 81, 73–84.
Yager, R.R., 2018. Decision making under measure-based granular uncertainty. Granul. Comput. 3 (4), 345–353.
Yager, R.R., Abbasov, A.M., 2013. Pythagorean membership grades, complex numbers, and decision making. Int. J. Intell. Syst. 28 (5), 436–452.
Yasmin, G., Das, A.K., Nayak, J., Pelusi, D., Ding, W., 2020. Graph based feature selection investigating boundary region of rough set for language identification. Expert Syst. Appl. 158, 113575.
Zadeh, L.A., 1986. A simple view of the Dempster-Shafer theory of evidence and its implication for the rule of combination. AI Mag. 7 (2), 85.
Zadeh, L.A., 1996. Fuzzy sets. In: Fuzzy Sets, Fuzzy Logic, and Fuzzy Systems: Selected Papers By Lotfi a Zadeh. World Scientific, pp. 394–432.
Zhao, J., Deng, Y., 2021. Complex network modeling of evidence theory. IEEE Trans. Fuzzy Syst. 29 (11), 3470–3480.
Zhou, M., Liu, X.-B., Chen, Y.-W., Qian, X.-F., Yang, J.-B., Wu, J., 2020. Assignment of attribute weights with belief distributions for madm under uncertainties. Knowl.-Based Syst. 189, 105110.