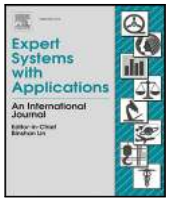
Expert Systems With Applications 238 (2024) 122297

Contents lists available at ScienceDirect

Expert Systems With Applications

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



A high order fractal-based Kullback–Leibler divergence with application in classification

Jie Zeng, Fuyuan Xiao *

School of Big Data and Software Engineering, Chongqing University, No. 55 South University Town Road, Shapingba District, Chongqing 401331, China

ARTICLE INFO

Keywords:
Dempster–Shafer evidence theory
High order KL divergence
Conflict management
Multi-source data fusion
Classification

ABSTRACT

Dempster–Shafer evidence theory (DSET) is extensively employed in multi-source data fusion applications. Nonetheless, when belief probability assignments (BPAs) exhibit considerable conflict, unexpected results can occur. To address this limitation, the high-order fractals are explored and a $\mathcal{K}$-order fractal-based Kullback–Leibler divergence ($\mathcal{K}$O-FKL) is introduced, which defines the $\mathcal{K}$-order as the optimal fractal epoch. This measure is employed to quantify the divergence between BPAs and demonstrates superior performance in assessing the conflict between two BPAs in numerical examples, compared to existing belief divergence methods. To utilize the $\mathcal{K}$O-FKL divergence measure to real-world problems, a novel $\mathcal{K}$O-FKL-based multi-source data fusion ($\mathcal{K}$O-FKL-MSDF) algorithm is designed. Through comparisons with well-known related methods, our proposed $\mathcal{K}$O-FKL-MSDF algorithm demonstrates superiority and enhanced robustness. Lastly, the $\mathcal{K}$O-FKL-MSDF algorithm is applied to real-world classification problems, underlining its high practical applicability.

1. Introduction

Multi-source data fusion is a technique that integrates information from various sources to generate a final decision target in real-world applications. A major challenge lies in managing highly conflicting multi-source data. Specifically, due to the ambiguity, inconsistency, or even inaccuracy of multi-source data, it is crucial to evaluate the confidence level for each data source. This allows for the downgrading or complete exclusion of unreliable data. Numerous well-established theories exist in the multi-source data fusion domain to model and address various types of uncertainties, including Z-number (Zhu, Liu et al., 2022), soft set (Fujita et al., 2020), rough set (Ye et al., 2021), fuzzy set (Garg & Rani, 2022) and multi-granularity (Miao et al., 2023). Moreover, these theories have found extensive applications across diverse fields, such as reliability analysis (Li, Huang et al., 2022), malicious information localization (Wang, Hou et al., 2022), emergency management (Che et al., 2022), pattern classification (Ko & Koo, 2023; Xiao, Cao et al., 2022; Xiao, Wen et al., 2022) and remote sensing image fusion (Kurban, 2022). Additionally, multi-source data fusion is particularly essential for addressing large-scale complex issues, including multi-attribute decision-making (Liu, Li et al., 2022), multidisciplinary design optimization (Meng, Yang et al., 2022), multi-agent learning (Chu et al., 2022; Wang, Mu et al., 2022), and other applications (Meng, Wang et al., 2022; Wang et al., 2023).

A fundamental theoretical basis for multi-source data fusion is the Dempster–Shafer evidence theory (DSET) (Dempster, 1967; Shafer, 1976). The primary benefits of D-S evidence theory include its ability to quantify belief values for individual targets as well as the unions of objects. Furthermore, D-S evidence theory allows for flexible and efficient uncertainty reasoning through the Dempster combination rule without requiring prior information. Due to its effectiveness and adaptability in modeling uncertainty, D-S evidence theory has been extensively applied in various information fusion domains (Deng, 2020b), such as group decision-making (Hua et al., 2022; Zhou et al., 2023, 2022), multi-criteria analysis (Fang et al., 2022), evidential reasoning (Ren et al., 2022; Wang, Zhou et al., 2022), credit risk assessment (Wang, Liu et al., 2022), portfolio construction in financial sector (Bisht & Kumar, 2023), knowledge management (Anjaria, 2022), and database retrieval (Yager et al., 2019).

Nevertheless, a significant challenge in Dempster's method lies in its ineffectiveness in handling highly conflicting BPAs, potentially leading to counterintuitive results (Xiong et al., 2021). Various approaches have been proposed to address this problem by measuring uncertainty and managing conflicting BPAs. For instance, Deng et al. use a distance matrix to estimate the support degrees among given BPAs (Deng et al., 2004). Building on this, some researchers explore this issue from a novel perspective of belief divergence (Fan & Xiao, 2022;

* Corresponding author.
E-mail addresses: zengjie@cqu.edu.cn (J. Zeng), xiaofuyuan@cqu.edu.cn (F. Xiao).

https://doi.org/10.1016/j.eswa.2023.122297
Received 14 April 2023; Received in revised form 8 August 2023; Accepted 20 October 2023
Available online 27 October 2023

Xiao, 2019, 2020). Divergence measures provide advantages in quantifying the differences and discrepancies between BPAs, facilitating the management of conflicting information and enhancing decision-making performance in uncertain contexts (Xiao, 2022a). Moreover, the efficacy of divergence-based multi-source data fusion in conflict measurement is verified. For example, divergence-based intuitionistic fuzzy sets are well-suited for decision-making (Verma & Álvarez-Miranda, 2023). Additionally, this approach has been successfully employed in EEG data analysis (Zhu, Xiao et al., 2022).

Among various uncertainty measurement techniques, belief divergence has emerged as a novel and promising direction in data fusion. In this study, our primary focus is on exploring belief divergence to address multi-source data fusion challenges. A comprehensive analysis of existing belief divergence methods reveals that, in some instances, these approaches may produce counterintuitive results, which will be discussed in detail in Section 2.4. Moreover, such counterintuitive divergence values could impact the outcomes of algorithm applications, including pattern recognition and classification. Therefore, developing a belief divergence model that effectively measures the discrepancy between BPAs has become an urgent necessity. To tackle this issue, this paper introduces an innovative belief divergence measurement method to overcome these challenges and enhance data fusion.

The fractal theory is renowned for its exceptional property of self-similarity, where macro-level integral objects consist of similar parts. Various fractal-based methods exist for uncertainty measurement, including information volume (Deng, 2020a; Zhou & Deng, 2022) and information dimension (Qiang et al., 2022). It has been observed that subsets in different BPAs might be partially supported, implying that fractal methods can evaluate the degree of overlap between events in uncertain information processing. Building on this idea, Deng introduces an improved belief structure satisfaction that considers the overlapping degree between events (Deng & Cui, 2021). In combination with divergence measures, Zeng proposes an FBD_{SKL} divergence measure (Zeng & Xiao, 2023) that integrates a pignistic fractal process (Smets, 2005) and divergence measurement to quantify conflict between BPAs. However, this measurement only takes into account a single fractal epoch, which may not be optimal for measuring conflicts between BPAs. Consequently, the proposed divergence-based measurement introduces a high-order fractal-based Kullback–Leibler (HO-FKL) divergence and defines an ideal fractal epoch $\mathcal{K}$. The primary contributions of this research can be summarized as follows:

- In the context of high-order fractal research, the optimal fractal epoch is proposed. As a result, a $\mathcal{K}$-order fractal-based Kullback–Leibler ($\mathcal{K}$O-FKL) divergence is introduced, which better measures the conflict between two BPAs under converging information differences.
- Based on the newly defined $\mathcal{K}$O-FKL, a novel $\mathcal{K}$O-FKL-based multi-source data fusion ($\mathcal{K}$O-FKL-MSDF) algorithm is proposed.
- The $\mathcal{K}$O-FKL-MSDF algorithm is subjected to an experiment and sensitivity analysis, as well as applied to classification problems involving five datasets with varying complexity levels, demonstrating high accuracy and strong robustness.

The remainder of this paper is organized as follows. Section 2 offers a brief overview of the fundamental concepts of evidence theory, and reviews existing related methods in the field. Section 3 introduces the $\mathcal{K}$-order fractal-based Kullback–Leibler divergence and thoroughly analyzes its characteristics through various examples. In Section 4, based on the proposed $\mathcal{K}$O-FKL, a novel multi-source data fusion algorithm ($\mathcal{K}$O-FKL-MSDF) is presented. Section 5 demonstrates the superiority and robustness of the $\mathcal{K}$O-FKL-MSDF algorithm through multi-source data fusion experiments for target recognition and a sensitivity analysis. In Section 6, the practical applicability of the proposed method is showcased using real-world datasets and compared with well-established existing approaches. Finally, Section 7 summarizes the research findings and future directions.

2. Preliminaries

2.1. Dempster–Shafer evidence theory

D-S evidence theory (Dempster, 1967; Shafer, 1976), as a generalization and an extension of Bayes probability theory, is a powerful and classical tool for handling ambiguous data. Consequently, it has been applied to various theories, including probability transformation (Chen et al., 2021), belief rule base (Chang et al., 2021), random permutation set (Chen & Deng, 2023; Chen et al., 2023a; Deng, 2022), evidence reasoning (Xu et al., 2020), complex evidence theory (Yang & Xiao, 2022), and generalized quantum evidence theory (Xiao, 2022b, 2023; Xiao & Pedrycz, 2022).

Definition 2.1 (*Frame of Discernment*)**.** Let Θ be a set of events that are mutually exclusive and collectively exhaustive. Thus, we have

$$\Theta = \{H_1, H_2, \dots, H_k, \dots, H_n\}, \tag{1}$$

which is defined as a frame of discernment.

Comprised of all the subsets in Θ, the power set of Θ, denoted as 2^Θ, is defined as

$$2^\Theta = \{\varnothing, \{H_1\}, \{H_2\}, \dots, \{H_n\}, \{H_1, H_2\}, \dots, \{H_1, H_2, \dots, H_k\}, \dots, \Theta\}, \tag{2}$$

where $\varnothing$ represents an empty set. It is evident that the number of subsets in 2^Θ is 2^n.

If $A \in 2^\Theta$, then A is referred to as a hypothesis.

Definition 2.2 (*Mass Function*)**.** In the frame of discernment Θ, a mass function from 2^Θ, denoted by m, is defined as:

$$m : 2^\Theta \to [0, 1], \tag{3}$$

which also satisfies the following conditions:

$$\sum_{A \in 2^\Theta} m(A) = 1 \ and \ m(\varnothing) = 0, \tag{4}$$

where $m(A)$ quantifies the belief value of hypothesis A. In D-S theory, a mass function m is also referred to as a BPA. If $m(A) > 0$, A is called a focal element. Because BPA effectively expresses uncertainty, it has been well-studied (Deng, 2020b; Li, Pelusi et al., 2022).

Definition 2.3 (*Belief Function and Plausibility Function*)**.** Let A be a hypothesis in the frame of discernment Θ. A belief function Bel is defined as the sum of the masses of all focal elements that contain A:

$$Bel(A) = \sum_{B \subseteq A} m(B). \tag{5}$$

Similarly, a plausibility function Pl is defined as the sum of the masses of all focal elements that intersect with A:

$$Pl(A) = \sum_{B \cap A \neq \varnothing} m(B). \tag{6}$$

It is essential to note that $Pl(A) \geqslant Bel(A)$, where $Bel(A)$ and $Pl(A)$ represent the lower and upper limits of the hypothesis A, respectively.

Definition 2.4 (*Dempster's Rule of Combination*)**.** Let $m1$ and $m2$ be two independent BPAs in the frame of discernment Θ. The combined BPA $m = m_1 \oplus m_2$ is defined as:

$$m(A) = \begin{cases} \frac{1}{1-K} \sum_{B \cap C = A} m_1(B) m_2(C), & A \neq \varnothing, \\ 0, & A = \varnothing, \end{cases} \tag{7}$$

where

$$K = \sum_{B \cap C = \varnothing} m_1(B) m_2(C), \tag{8}$$

and $B, C \in 2^{\Theta}$. Generally, the coefficient K is used to measure the level of conflict between m_1 and m_2.

Note that from Eqs. (7) and (8), the range of K is inferred: $0 \leq K < 1$.

2.2. Information entropy of BPA

In the realm of uncertainty measurement, entropy serves as a means to quantify the amount of information contained within a BPA. Shannon entropy (Shannon, 2001), a classic information entropy, is commonly employed to measure the uncertainty of probability distributions. In the context of measuring BPAs, Deng entropy and its extensions (Deng, 2020b) are widely adopted for quantifying BPA uncertainty. Furthermore, entropy has practical applications, such as measuring EEG (Cao et al., 2019), demonstrating its versatility and utility in various fields.

Definition 2.5 (*Shannon Entropy*)**.** Let P be a probability distribution defined in the frame of discernment $\Theta = \{H_1, H_2, \ldots, H_n\}$. Then, Shannon entropy is defined (Shannon, 2001):

$$H_s(P) = \sum_{\theta \in \Theta} -P(\theta) \log P(\theta), \tag{9}$$

where $\sum_{\theta \in \Theta} P(\theta) = 1$ and $0 < P(\theta) \leqslant 1$. Typically, Shannon entropy is measured in the log base of 2. In this condition, it is noted that $0 \leqslant H_s(P) \leqslant \log_2 n$.

Definition 2.6 (*Deng Entropy*)**.** Let m be BPA in the frame of discernment Θ, the Deng entropy is defined (Deng, 2020b):

$$H_{DE}(m) = -\sum_{A \in 2^{\Theta}} m(A) \log \left(\frac{m(A)}{2^{|A|} - 1} \right), \tag{10}$$

where $|A|$ is the cardinality of focal element A.

Specifically, when each focal element is a single-element, the BPA is reduced to a probability distribution. In this case, Deng entropy degenerates into Shannon entropy.

2.3. Belief divergence between two BPAs

Recently, numerous divergence-based methods have been proposed for measuring conflict between information (Chen et al., 2023b; Huang et al., 2023; Song et al., 2019). Xiao introduced a Belief Jensen–Shannon divergence (BJS divergence) (Xiao, 2019) to quantify the conflict between BPAs. Subsequently, several divergence measures have been proposed to assess the discrepancy and conflict among BPAs. Among these methods, $\mathfrak{B}$ divergence (Xiao, 2020) stands out as a representative approach for measuring the discrepancy between BPAs.

Definition 2.7 (*BJS Divergence Measure*)**.** Let m_1 and m_2 be two independent BPAs, and A is a hypothesis in the frame of discernment Θ. A BJS divergence measure between two BPAs m_1 and m_2 is defined as follows (Xiao, 2019):

$$BJS(m_1, m_2) = \frac{1}{2}[S\left(m_1, \frac{(m_1 + m_2)}{2}\right) + S\left(m_2, \frac{(m_1 + m_2)}{2}\right)], \tag{11}$$

where $S\left(m_1, m_2\right) = \sum_{i=1}^{2^n-1} m_1\left(A_i\right) \log \frac{m_1(A_i)}{m_2(A_i)}$ and $\sum_{i=1}^{2^n-1} m_j(A_i) = 1$ $(j = 1, 2)$.

So $BJS(m_1, m_2)$ divergence is also expanded, and expressed with the following formula:

$$\begin{aligned} BJS(m_1, m_2) =& \frac{1}{2} \sum_{i=1}^{2^n-1} m_1(A_i) \log(\frac{2m_1(A_i)}{m_1(A_i) + m_2(A_i)}) \\ &+ \frac{1}{2} \sum_{i=1}^{2^n-1} m_2(A_i) \log(\frac{2m_2(A_i)}{m_1(A_i) + m_2(A_i)}). \end{aligned} \tag{12}$$

Definition 2.8 ($\mathfrak{B}$ *Divergence Measure for Two BPAs*)**.** Let m_1 and m_2 be two BPAs in the frame of discernment Θ, in which there are n mutually exclusive hypotheses. In this frame of discernment Θ, let A_i and A_j be two hypotheses from m_1 and m_2 $(1 \leqslant i, j, \leqslant 2^n - 1,)$. The $\mathfrak{B}$ belief divergence measure between m_1 and m_2 is defined as (Xiao, 2020):

$$\begin{aligned} \mathfrak{B}(m_1, m_2) =& \frac{1}{2} \sum_{i=1}^{2^n-1} \sum_{j=1}^{2^n-1} m_1(A_i) \log \frac{m_1(A_i)}{\frac{1}{2}m_1(A_i) + \frac{1}{2}m_2(A_j)} \frac{|A_i \cap A_j|}{|A_j|} \\ &+ \frac{1}{2} \sum_{i=1}^{2^n-1} \sum_{j=1}^{2^n-1} m_2(A_j) \log \frac{m_2(A_j)}{\frac{1}{2}m_1(A_i) + \frac{1}{2}m_2(A_j)} \frac{|A_i \cap A_j|}{|A_i|}, \end{aligned} \tag{13}$$

where the $|A_i|$ represents the cardinality of subset of A_i, and $|A_i \cap A_j|$ equals the cardinality of intersection subset between A_i and A_j.

It is evident that when the focal elements consist only of singletons, the $\mathfrak{B}$ divergence degenerates into BJS divergence:

$$\begin{aligned} \mathfrak{B}(m_1, m_2) =& \frac{1}{2} \sum_{i=1}^{2^n-1} m_1(A_i) \log \frac{m_1(A_i)}{\frac{1}{2}m_1(A_i) + \frac{1}{2}m_2(A_i)} \\ &+ \frac{1}{2} \sum_{i=1}^{2^n-1} m_2(A_i) \log \frac{m_2(A_i)}{\frac{1}{2}m_1(A_i) + \frac{1}{2}m_2(A_i)}. \end{aligned} \tag{14}$$

The FBD_{SKL} method, proposed by Zeng, combines the principles of one iteration of fractals and has various desirable properties.

Definition 2.9 (FBD_{SKL} *Divergence Measure*)**.** Let m_1 and m_2 be two BPAs in the frame of discernment Θ. The symmetric fractal-based belief KL divergence $FBD_{SKL}(m_1, m_2)$ is defined as (Zeng & Xiao, 2023):

$$FBD_{SKL}(m_1, m_2) = \frac{1}{2} \sum_{i=1}^{2^n-1} \left[\begin{array}{l} m_{F_1}\left(H_i\right) \log \frac{m_{F_1}(H_i)}{\sqrt{m_{F_1}(H_i) \times m_{F_2}(H_i)}} \\ + m_{F_2}\left(H_i\right) \log \frac{m_{F_2}(H_i)}{\sqrt{m_{F_2}(H_i) \times m_{F_1}(H_i)}} \end{array} \right], \tag{15}$$

where $m_F\left(H_i\right)$ is called fractal-based basic probability assessment (FBBPA) and is defined on a pignistic probability transformation (Smets, 2005):

$$m_F\left(H_i\right) = \sum_{H_i \subseteq G_i} \frac{m\left(G_i\right)}{2^{|G_i|} - 1}, \tag{16}$$

where $H_i, G_i \subseteq \Theta$.

2.4. Numerical analysis of BJS and $\mathfrak{B}$ divergence measure methods

In this subsection, numerical examples of three divergence measures, namely BJS, $\mathfrak{B}$, and FBD_{SKL}, are provided. By analyzing the performance and limitations of these measures based on concrete examples, it is motivated to propose a new divergence method.

Example 1. Suppose there are two independent BPAs, m_1 and m_2, in the frame of discernment $\Theta = A, B$, where α is an independent variable that ranges from 0 to 1:

$$\begin{array}{ll} m_1: & m_1(\{A\}) = 1 - \alpha, \quad m_1(\{A, B\}) = \alpha; \\ m_2: & m_2(\{B\}) = 1. \end{array}$$

In Example 1, when $\alpha = 0$, it obtained that $\begin{cases} m_1(\{A\}) = 1 \\ m_2(\{B\}) = 1 \end{cases}$, indicating maximum conflicts between m_1 and m_2, and the divergence should also be maximum at this point. As α gradually increases, the belief value in $m_1(\{A, B\})$ increases, while the belief value in $m_1(\{A\})$ decreases. Thus, during this process, $Bel(B)$ in m_1 increases. Consider that it is observed that m_2 constantly supports event B. It is intuitively concluded that the conflict between m_1 and m_2 decreases. Finally, when $\alpha = 1$, it is obtained that $\begin{cases} m_1(\{A, B\}) = 1 \\ m_2(\{B\}) = 1 \end{cases}$, so there should still remain some uncertainty, but the conflict level between m_1 and m_2 has been greatly reduced compared with the initial status: $\begin{cases} m_1(\{A\}) = 1 \\ m_2(\{B\}) = 1 \end{cases}$.

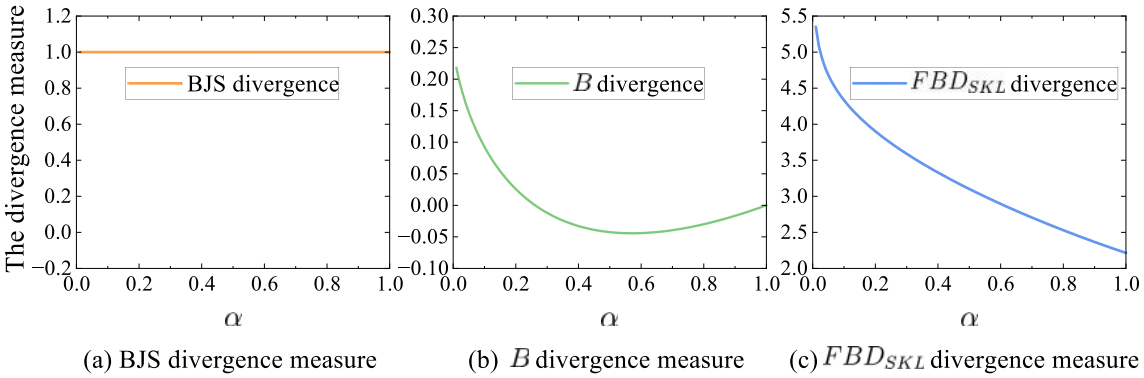


Fig. 1. Divergence measure with varying α in Example 1.

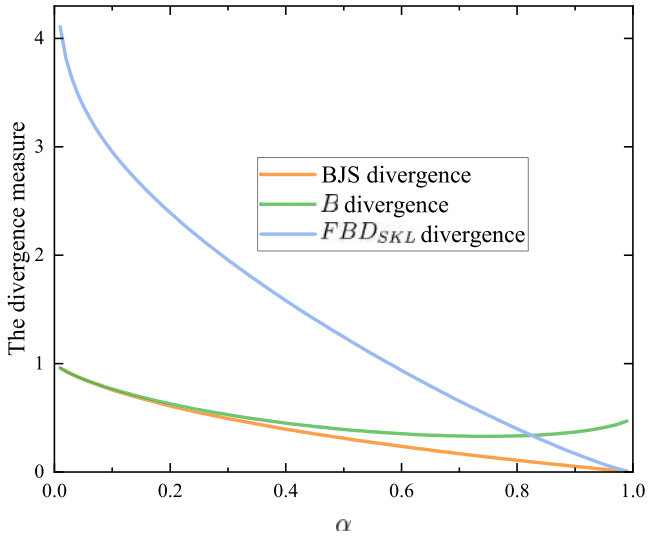


Fig. 2. Divergence measure with varying α in Example 2.

However, from Fig. 1, it is observed that the BJS divergence remains constant throughout. Moreover, the $\mathfrak{B}$ divergence reveals an unobvious inflection point. As for the FBD_{SKL} divergence, its curve shows a trend that is more in line with our expectations: On the one hand, as α changes from 0 to 1, the conflict between m_1 and m_2 is continuously decreasing. On the other hand, when $\alpha = 1$, some conflicts are still preserved.

Based on our analysis, it can be observed that the BJS divergence (from Eq. (12)) does not consider the similarity between subsets, as it only considers sets that are completely identical. On the other hand, although the $\mathfrak{B}$ divergence (from Eq. (13)) considers the correlation between subsets, it still reveals a counter-intuitive inflection point. The FBD_{SKL} divergence is conducted with a fractal process, and is able to take into account the similarity within events, which allows for more accurate measurement of conflicts.

Example 2. Suppose there are two BPAs, denoted as m_1 and m_2, in the frame of discernment $\Theta = A, B$. The variable α ranges from 0 to 1 and takes independent values for each BPA, as follows:

$$m_1: \quad m_1(\{A\}) = \alpha, \quad m_1(\{A, B\}) = 1 - \alpha;$$
$$m_2: \quad m_2(\{A\}) = 0.9999, m_2(\{A, B\}) = 0.0001.$$

In Example 2, as α increases, the conflict between m_1 and m_2 decreases until m_1 becomes completely equivalent to m_2 when $\alpha = 0.9999$. Intuitively observed in Fig. 2, the trends of BJS divergence curve and FBD_{SKL} curve are consistent with expectation. However, the curve of $\mathfrak{B}$ divergence in Fig. 2 shows an unexpected and counter-intuitive inflection.

In summary, in Examples 1 and 2, both BJS divergence and $\mathfrak{B}$ divergence show some limitations in certain situations. Although the FBD_{SKL} divergence does not exhibit any intuitive contrast, it does not mean that the model has no room for optimization. As the FBD_{SKL} model assumes only one fractal process, in this study, we need to explore how many fractal processes are optimal. In other words, we need to determine how many fractal processes are needed to accurately characterize the relationship between information quantities among BPAs.

3. A high order fractal-based Kullback–Leibler divergence

In this section, in order to find the ideal fractal process time, the variance of information difference between 2 BPAs is observed. Thus, combined with Deng entropy and FBD_{SKL}, a High Order Fractal-based Kullback–Leibler(HO-FKL) Divergence is devised. Then, the performance of the HO-FKL is analyzed using several concrete examples.

3.1. A high order of fractal-based belief probability assignment

Definition 3.1 (*A High Order Fractal-Based Belief Probability Assignment*)**.** Let m be one BPA in the frame of discernment Θ, as the number of fractal epoch k increases, a high order of fractal-based belief probability assignment (HO-FBBPA) is defined as follows:

$$m_F^k(H_i) = \sum_{H_i \subseteq G_i} \frac{m_F^{k-1}(G_i)}{2^{|G_i|} - 1}, \tag{17}$$

where $H_i, G_i \subseteq \Theta$, k represents the fractal epoch and satisfies $k \geqslant 1$. Specifically, when $k = 1$, which refers to the first fractal transformation and is in line with Eq. (16). Thus, it is obtained that:

$$m_F^1(H_i) = \sum_{H_i \subseteq G_i} \frac{m_F^0(G_i)}{2^{|G_i|} - 1} = \sum_{H_i \subseteq G_i} \frac{m(G_i)}{2^{|G_i|} - 1}, \tag{18}$$

where the HO-FBBPA m_F^0 degenerates into BPA m.

As illustrated in Fig. 3, during each fractal process, multiple subset elements undergo a uniform split, contributing to singleton belief values. High-order splits involve multiple iterations of such fractal processes. When no fractal process is applied, multi-element subsets can increase the system's uncertainty. As the number of fractal iterations approaches infinity, the BPA will eventually degenerate into a probability distribution.

Definition 3.2 (*A High Order Fractal-Based Kullback–Leibler Divergence*)**.** Let $m_{F_1}^k$ and $m_{F_2}^k$ be two HO-FBBPAs in the frame of discernment Θ.

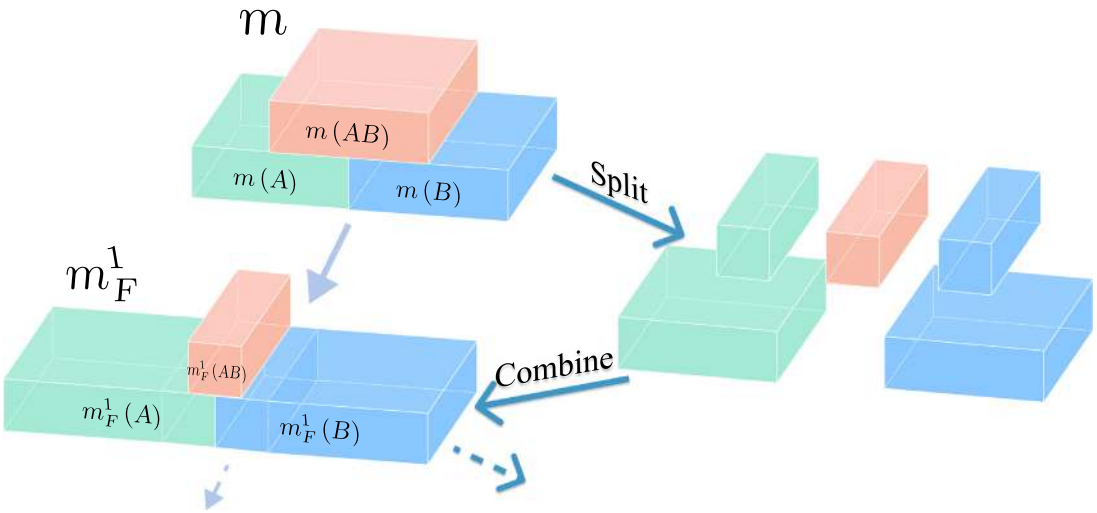


Fig. 3. High order fractal transformation process.

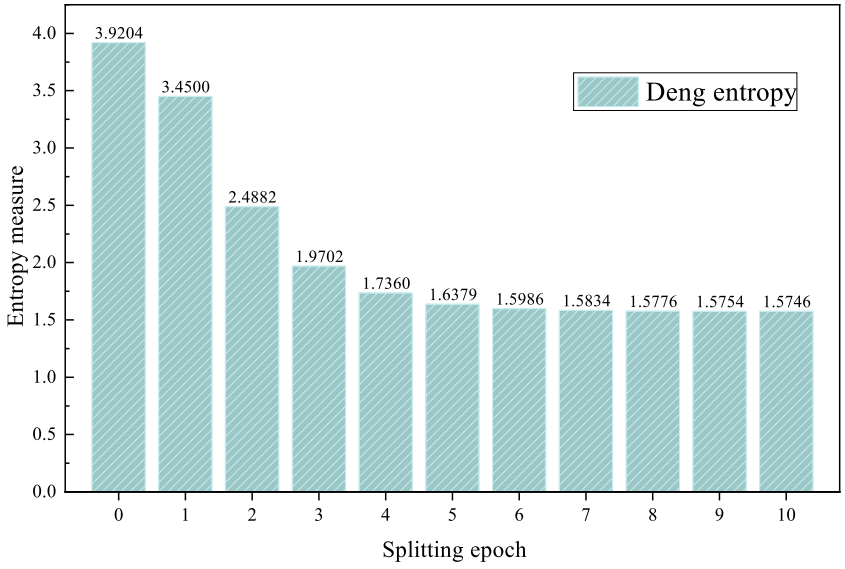


Fig. 4. Entropy measure with varying epoch k in Example 3.

A high order of fractal-based Kullback–Leibler (HO-FKL) divergence of $m_{F_1}^k$ and $m_{F_2}^k$ is defined as follows:

$$\text{HO-FKL}\left(m_{F_1}^k, m_{F_2}^k\right) = \frac{1}{2}\left[\begin{array}{l} \sum_i m_{F_1}^k(A_i)\log\left(\frac{m_{F_1}^k(A_i)}{\sqrt{m_{F_1}^k(A_i)\times m_{F_2}^k(A_i)}}\right) \\ +\sum_i m_{F_2}^k(A_i)\log\left(\frac{m_{F_2}^k(A_i)}{\sqrt{m_{F_1}^k(A_i)\times m_{F_2}^k(A_i)}}\right) \end{array}\right], \quad (19)$$

where k denotes the order of the fractal and satisfies the condition of $k \geqslant 1$.

3.2. Analysis of Deng entropy change of a BPA during fractal process

In this section, during continuous fractal processes, changes and characteristics in Deng entropy of a BPA are investigated. Then, based on the analysis of the changes in entropy difference between two BPAs during fractal processes, a criterion for selecting the ideal number of the fractal epoch is proposed (see Fig. 4).

Example 3. Suppose that there is one BPA m_1 in the frame of discernment $\Theta = \{A, B, C, D\}$:

$$m_1(\{A, B, C, D\}) = 1,$$

explore the changes of Deng entropy of m_1^k during the fractal process, where k represents the number of fractal epoches.

According to the formula of Deng entropy in Eq. (10), it assigns more uncertainty and information to multisets. Therefore, in Example 3, when $k = 0$ (i.e., before the first fractal process), the system is in the most uncertain state with $m_1(A, B, C, D) = 1$, thus the information entropy should be maximal. As shown in Eq. (16), with the increase of fractal epoch k, m_1^k gradually approaches a probability distribution, so Deng entropy gradually degenerates into Shannon entropy. From Table 1 and Fig. 1, the change of H_{DE} value can be intuitively observed.

Based on the observations above, it is analyzed that for m_1 in this case: if it is not conducted with fractal process, its internal uncertainty will be maximized. However, The maximized uncertainty is not conducive to effectively measuring the discrepancy between BPAs. On the

Table 1
Changes in Deng entropy during continuous fractal process.

Epoch k	H_{DE} value	Epoch k	H_{DE} value
0	3.920358814	6	1.598649699
1	3.44996417	7	1.583373365
2	2.488201925	8	1.577554733
3	1.970179059	9	1.575373318
4	1.735978204	10	1.574565606
5	1.637923584		

Table 2
Entropy difference and HO-FKL divergence with varying epoch k.

k	DH_{DE} value	HO-FKL value	k	DH_{DE} value	HO-FKL value
0	0.190013453	0.08157586	6	0.112744975	0.044185718
1	0.431655518	0.063295221	7	0.11031425	0.044034492
2	0.261136243	0.051090441	8	0.109394737	0.044036381
3	0.173590225	0.046509382	9	0.109051903	0.044064025
4	0.134967606	0.044885521	10	0.108925531	0.044082277
5	0.11904881	0.04435566			

other hand, if the fractal epoch is too high, a BPA is almost completely transformed into a probability distribution. Therefore, it is crucial to find the ideal fractal epoch for the BPA, and use it as a preparation for measuring the conflict between BPAs in the subsequent analysis.

Before we dive into Example 4, the concept of difference of Deng entropy between two HO-FBBPAs is firstly introduced in order to better analyze its relationship with HO-FKL in Example 4.

Definition 3.3 (*Difference of Deng Entropy Between Two HO-FBBPAs*)**.** Let $m^k_{F_1}$ and $m^k_{F_2}$ be two HO-FBBPAs in the frame of discernment Θ, where k represents the fractal epoch and satisfies $k \geqslant 1$. The Difference of Deng entropy (DH_{DE}) between two HO-FBBPAs is defined as follows:

$$DH_{DE}\left(m^k_{F_1}, m^k_{F_2}\right) = \left|H_{DE}\left(m^k_{F_1}\right) - H_{DE}\left(m^k_{F_2}\right)\right|. \tag{20}$$

It should be noted that since HO-FBBPA is a special type of BPA, thus it can also be calculated using Eq. (10).

Example 4. Suppose that there are two BPAs m_1 and m_2 in the frame of discernment $\Theta = \{A, B, C\}$:

$$\begin{aligned} m_1 &: \quad m_1(\{A\}) = 0.5, m_1(\{B\}) = 0.1,\ m_1(\{C\}) = 0.2, m_1(\{A, B, C\}) = 0.2; \\ m_2 &: \quad m_2(\{A\}) = 0.6, m_2(\{B\}) = 0.2,\ m_2(\{C\}) = 0.1, m_2(\{A, B, C\}) = 0.1, \end{aligned}$$

explore the relationship between the H_{DE}, DH_{DE} and HO-FKL during the fractal processes, where k represents the epoch of fractal epoches.

From Fig. 5(a), it is clearly noticed the traits of Deng entropy value during continuous fractal processes. The bars in Fig. 5(a) show the respective change of m_1 and m_2 in entropy value in this process, while the dotted line shows the difference in entropy value between m_1 and m_2. This entropy difference can effectively measure the difference in information between the two systems. Therefore, when the information difference value converges, it is considered as the best epoch to measure the discrepancy between the two systems. Specifically, from Table 2, it can be seen that starting from epoch $k = 5$, the change of DH_{DE} are all less than $0.11904881 - 0.112744975 = 0.006303835$. Therefore, the entropy difference is considered to have converged at this point. As shown in Fig. 5(b), two lines are used to show the trends of entropy difference value and $FBD_{SKL}\left(m_1, m_2\right)$ during the continuous fractal processes. It is also noted that in Fig. 5(b), at the point when the information discrepancy between the two systems converges, the divergence between m_1 and m_2 is obtained:

$$\text{HO} - \text{FKL}\left(m^5_{F_1}, m^5_{F_2}\right) = 0.04435566.$$

3.3. A $\mathcal{K}$-order fractal-based Kullback–Leibler divergence

Definition 3.4 (*A $\mathcal{K}$-order Fractal-Based Kullback–Leibler Divergence*)**.** Let m_1 and m_2 be two BPAs in the frame of discernment Θ. A $\mathcal{K}$-order fractal-based Kullback–Leibler ($\mathcal{K}$O-FKL) divergence of m_1 and m_2 is defined as follows:

$$\mathcal{K}\text{O-FKL}\left(m_1, m_2\right) = \frac{1}{2}\left[\begin{array}{l} \sum_i m^{\mathcal{K}}_{F_1}\left(A_i\right) \log\left(\dfrac{m^{\mathcal{K}}_{F_1}(A_i)}{\sqrt{m^{\mathcal{K}}_{F_1}(A_i) \times m^{\mathcal{K}}_{F_2}(A_i)}}\right) \\ + \sum_i m^{\mathcal{K}}_{F_2}\left(A_i\right) \log\left(\dfrac{m^{\mathcal{K}}_{F_2}(A_i)}{\sqrt{m^{\mathcal{K}}_{F_1}(A_i) + m^{\mathcal{K}}_{F_2}(A_i)}}\right) \end{array}\right], \tag{21}$$

where $\mathcal{K}$ represents the best ideal epoch and satisfies the following convergence criterion:

$$DH_{DE}\left(m^{\mathcal{K}}_{F_1}, m^{\mathcal{K}}_{F_2}\right) - DH_{DE}\left(m^{\mathcal{K}+1}_{F_1}, m^{\mathcal{K}+1}_{F_2}\right) < \varepsilon, \tag{22}$$

where ε is the allowable error.

It is noted that $\mathcal{K}$O-FKL is a manifestation of the high order of fractals after the generalization of FBD_{SKL}, it also satisfies the three desirable properties of FBD_{SKL}:

[1] Nonnegativity: $0 \leqslant \mathcal{K}\text{O-FKL}(m_1, m_2)$.
[2] Nondegeneracy: $\mathcal{K}\text{O-FKL}(m_1, m_2) = 0$ if and only if $m_1 = m_2$.
[3] Symmetry: $\mathcal{K}\text{O-FKL}(m_1, m_2) = \mathcal{K}\text{O-FKL}(m_2, m_1)$.

In Example 4, when the allowable error is set to be 0.001, it can be observed that DH_{DE} converges at epoch $k = 5$. In other words, when $\mathcal{K} = 5$ in Eq. (21), the $\mathcal{K}$O-FKL value between m_1 and m_2 can be computed for Example 4 as follows:

$$\mathcal{K}\text{O-FKL}\left(m_1, m_2\right) = \text{HO-FKL}\left(m^5_1, m^5_2\right) = 0.04435566.$$

3.4. Numerical examples and discussions of $\mathcal{K}$O-FKL divergence

In this chapter, the proposed $\mathcal{K}$O-FKL divergence measure is analyzed first in Examples 1 and 2, and then some new numerical examples will be provided to verify the characteristics of $\mathcal{K}$O-FKL. Moreover, in the following examples, the base of the logarithmic function is fixed to 2, and the allowable error ε in Eq. (22) is set to 0.001.

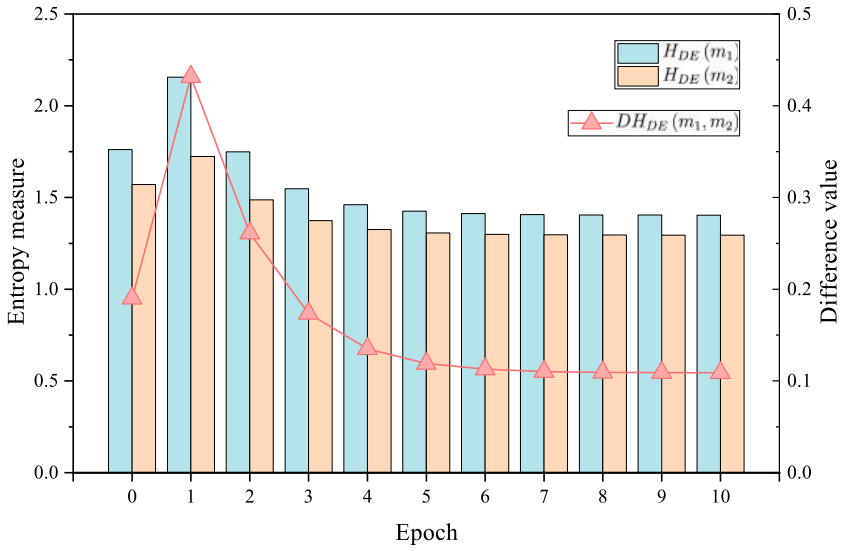
Firstly, let us return to Example 1 in Section 3.2, it is easy to obtain that when $\alpha = 0$, then $\alpha = 0$, it obtained that $\begin{cases} m_1(\{A\}) = 1 \\ m_2(\{B\}) = 1 \end{cases}$. At this point, there is no intersection between m_1 and m_2, so the divergence between them should be the maximum. As shown in Fig. 6(a), when $\alpha = 0$, the proposed $\mathcal{K}$O-FKL measure obtains a divergence value of 1, which is the maximum value during the increment of α. According to the former analysis in Example 1: as α gradually increases, the conflict between m_1 and m_2 should decrease continuously to a value greater than 0 (because there is still some conflict between m_1 and m_2). According to Fig. 6(a), the value of proposed $\mathcal{K}$O-FKL measure decreases continuously with the increase of α, and finally ends at $(1, 0.311)$, which is in line with our expectations.

Then, the proposed $\mathcal{K}$O-FKL is calculated for Example 2 and plotted in Fig. 6(b). In Example 2, it is easy to notice that as α increases, m_1 and m_2 will become increasingly consistent until they are completely identical ($\alpha = 0.9999$). In Fig. 6(b), it is found that the proposed $\mathcal{K}$O-FKL value decreases continuously as α increases, until the divergence measure, which reflects the conflict level, finally becomes 0 when $\alpha = 0.9999$. Therefore, in Example 1 and Example 2, both the trend of conflict and the boundary values are consistent with expectations. In the remaining content, the performance of the proposed $\mathcal{K}$O-FKL divergence model is demonstrated and analyzed comprehensively with more examples.

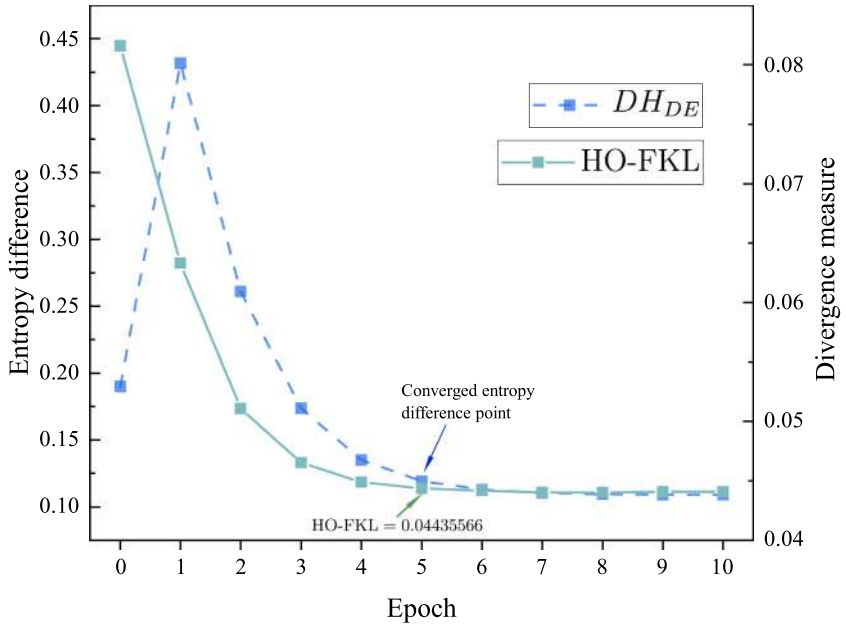
Example 5. Suppose there are two BPAs m_1 and m_2 in the frame of discernment $\Theta = \{A, B\}$. They are defined with varied α and β:

$$\begin{aligned} m_1 &: \quad m_1(\{A\}) = \alpha, m_1(\{B\}) = \beta, m_1(\{A, B\}) = 1 - \alpha - \beta; \\ m_2 &: \quad m_2(\{A\}) = \beta, m_2(\{B\}) = \alpha,\ m_2(\{A, B\}) = 1 - \alpha - \beta, \end{aligned}$$

where α and β ranges from $[0, 1]$, respectively.



(a) Entropy difference with varying epoch k



(b) Entropy difference and HO-FKL value with varying epoch k

Fig. 5. Entropy difference and divergence value in Example 4.

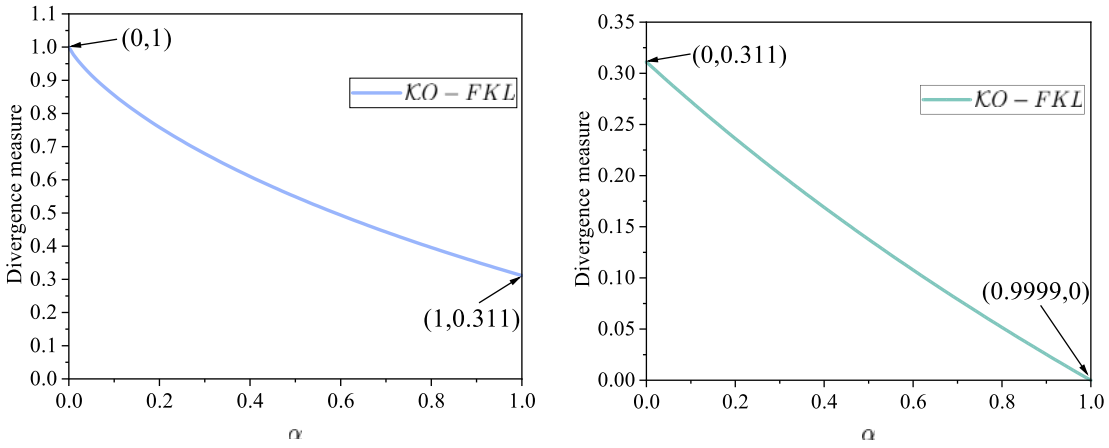
To fulfill the fundamental property of a mass function (in Eq. (4)), it is straightforward to infer that α and β should abide by the constraint of $\alpha + \beta \leqslant 1$. Based on the observation of Fig. 7(a), it can be seen that the value of $\mathcal{K}$O-FKL is always greater than 0, which satisfies the nonnegativity property mentioned in Definition 3.4. Moreover, since α and β are symmetric in this case, the three-dimensional plot of $\mathcal{K}$O-FKL in Fig. 7(a) is symmetric with respect to the plane $X + Y = 0$. Shown in Fig. 7(a), the distribution of the $\mathcal{K}$O-FKL values with the variation of α and β is displayed in a two-dimensional plane. It is visually observed that the $\mathcal{K}$O-FKL divergence value is relatively large when the difference between α and β is large. The definition of m_1 and m_2 in Example 5 also illustrates the same pattern: as the difference between α and β increases, the discrepancy caused by the singleton events A and B will increase. Therefore, the results in Fig. 7(a) are consistent with expectations.

Example 6. Suppose there are two BPAs m_1 and m_2 in the frame of discernment Θ. They are defined with varied α ranging from $[0, 1]$. Additionally, assume E represent a varied event whose length(denoted by l_E) varies from $[1, 8]$ and is defined in Table 3:

$$m_1: \quad m_1(\{dempster1967upperB\}) = \alpha, m_1(E) = 1 - \alpha;$$

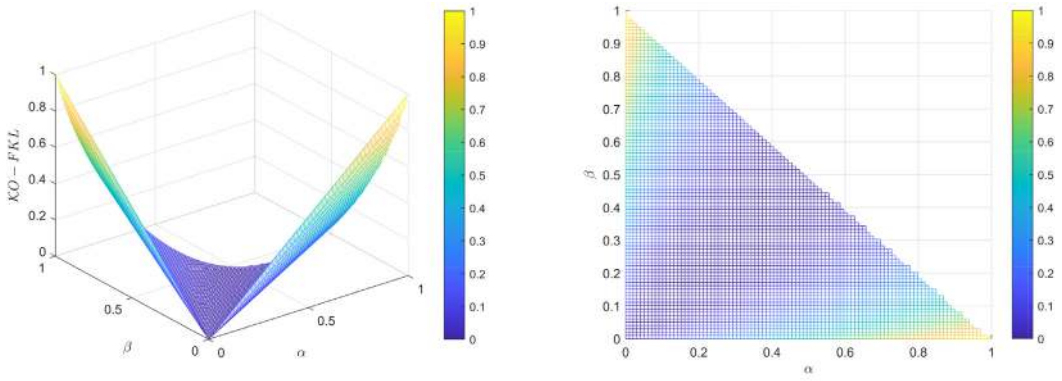
$$m_2: \quad m_2(\{B\}) = 0.95,\ m_2(E) = 0.05.$$

In this example, the variation of proposed $\mathcal{K}$O-FKL with respect to the parameters is shown in Fig. 8. To begin with, Fig. 8(a) presents an overview of the $\mathcal{K}$O-FKL model in three dimensions; while Fig. 8(b) shows the range of varied parameters α and l_E. Based on Figs. 8(a) and 8(b), the distribution range of the values of $\mathcal{K}$O-FKL is clearly obtained. And it is observed that at any time, it satisfies $\mathcal{K}$O-FKL$(m_1, m_2) \geqslant 0$. According to Fig. 8(c), the $\mathcal{K}$O-FKL value calculated at $t = 1$ is higher than those at other values of t. This is because when $t = 1$, m_1 and m_2 are



(a) $\mathcal{K}$O-FKL measure in Example 1 (b) $\mathcal{K}$O-FKL measure in Example 2

Fig. 6. $\mathcal{K}$O-FKL divergence measure with varied α.



(a) Divergence measure varying with α and β (b) Varation of α and β

Fig. 7. The proposed $\mathcal{K}$O-FKL measure in Example 5.

defined as follows: $\begin{cases} m_1: & m_1(B)=\alpha, m_1(A)=1-\alpha; \\ m_2: & m_2(B)=0.95,\ m_2(A)=0.05. \end{cases}$ At this time, the event E does not contain any subset containing event B. In other words, E and B have no intersection at all in this case. So the conflict value is greater than the conflict value in other cases where there is an intersection. As l_E gradually increases from $[2, 8]$, since E contains more subsets, the uncertainty becomes greater. Therefore, $\mathcal{K}$O-FKL will also gradually increase. In Fig. 8(d), it is observed that as α increases gradually from 0 to 0.95, the value of $\mathcal{K}$O-FKL gradually decreases. And in this example, m_1 also gradually approaches m_2 as α increases. When $\alpha = 0.95$, m_1 is completely identical to m_2, so the divergence value between them should be 0. Subsequently, as α continues to increase from 0.95 to 1, the conflict between m_1 and m_2 increases. So the divergence value in this interval of $[0, 0.95]$ also increases. Overall, this example demonstrates that the $\mathcal{K}$O-FKL model is applicable to more complex situations than the previous examples and yields results as expected.

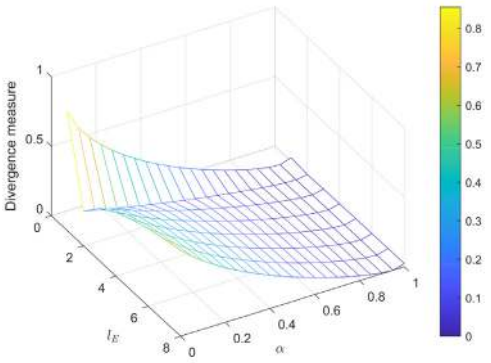
Example 7. Suppose there are two BPAs m_1 and m_2 in the frame of discernment Θ. Additionally, assume E represent a varied event whose length(denoted by l_E) varies from $[1, 8]$ and is defined in Table 3:

$m_1: \quad m_1(\{B\}) = 0.05, m_1(E) = 0.95;$
$m_2: \quad m_2(\{B\}) = 0.95,\ m_2(E) = 0.05.$

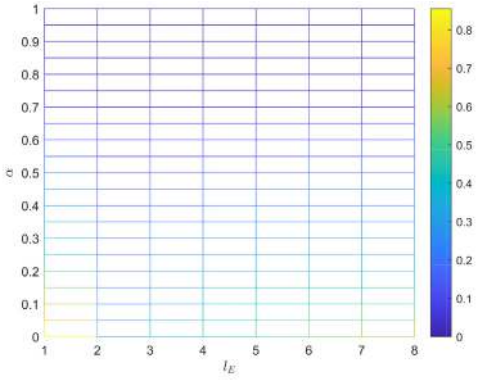
To begin with, it is noticed that this example is a specific case of Fig. 8 under fixed α. As analyzed in Fig. 8(d), when $t = 1$, the following equations are obtained: $\begin{cases} m_1: & m_1(B)=0.05, m_1(A)=0.95; \\ m_2: & m_2(B)=0.95,\ m_2(A)=0.05. \end{cases}$ At this time, the conflict between m_1 and m_2 is the highest. When $t = 2$, we have $\begin{cases} m_1: & m_1(B)=0.05, m_1(A)=0.95; \\ m_2: & m_2(B)=0.95,\ m_2(A)=0.05. \end{cases}$ Because m_1 and m_2 contain events with intersection, the conflict value in this case ($t = 2$) is smaller than that at $t = 1$. Subsequently, as t gradually increases from 2 to 8, the subsets contained in event E set become increasingly more from Table 3, but only event B can intersect with E. Thus, the uncertainty between m_1 and m_2 will increase. According to Fig. 9, both BJS and $\mathfrak{B}$ divergence produce results that conflict with intuition: their divergence values remain unchanged. For FBD_{SKL} and proposed $\mathcal{K}$O-FKL measures, their trends are consistent with our analysis of

Table 3
The varied event E in Examples 6 and 7.

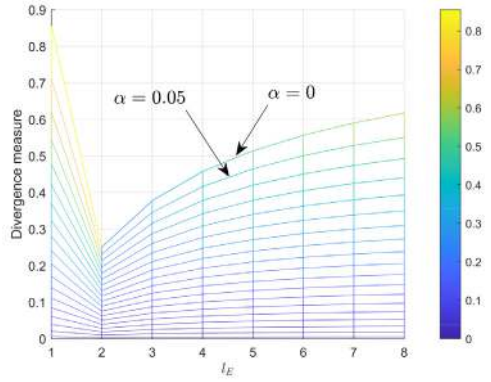
l_E	E
1	A
2	A, B
3	A, B, C
4	A, B, C, D
5	A, B, C, D, E
6	A, B, C, D, E, F
7	A, B, C, D, E, F, G
8	A, B, C, D, E, F, G, H



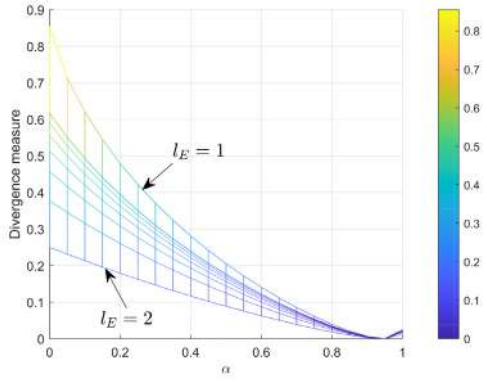
(a) Divergence measure varying with varied α and l_E



(b) Varation of α and l_E



(c) Divergence measure with varied l_E



(d) Divergence measure with varied α

Fig. 8. The proposed $\mathcal{K}$O-FKL measure in Example 6.

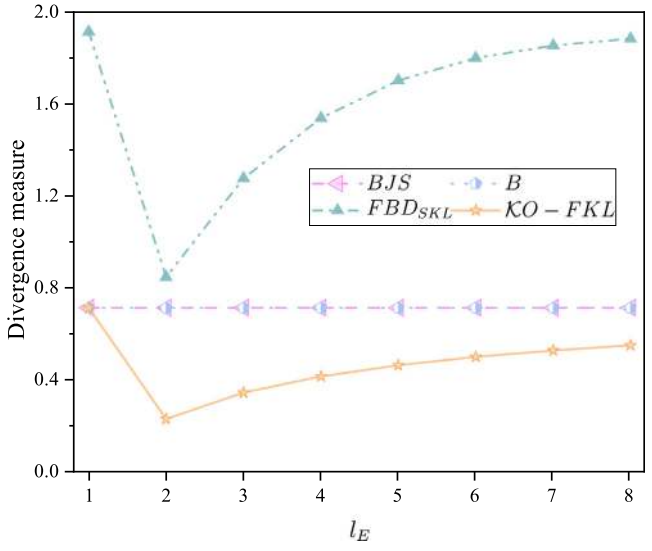


Fig. 9. Divergence measure with varied E in Example 7.

Example 7 above. Next, in subsequent classification applications, it is needed to further compare and analyze the characteristics of the results obtained by the two methods (FBD_{SKL} and proposed $\mathcal{K}$O-FKL) in order to demonstrate the importance of the process of seeking higher-order $\mathcal{K}$.

4. A novel $\mathcal{K}$o-FKL-based multi-source data fusion algorithm

This section applies the $\mathcal{K}$O-FKL-based divergence measure to multi-source data fusion. Thus, a $\mathcal{K}$O-FKL-based multi-source data fusion ($\mathcal{K}$O-FKL-MSDF) is proposed. Firstly, the main steps of the $\mathcal{K}$O-FKL-MSDF algorithm are illustrated. Then, the computational complexity is analyzed in comparison with other algorithms.

4.1. Main steps of $\mathcal{K}$o-FKL-MSDF

In this section, the detailed steps of the $\mathcal{K}$O-FKL-MSDF algorithm is presented. Then, the corresponding pseudocode is provided in Algorithm 1. To facilitate intuitive understanding, the entire algorithm process is also presented in the form of a flowchart in Fig. 10.

Step 1 Consider N independent BPAs, each containing n mutually exclusive events. To measure the similarity between two BPAs m_i and m_j $(i, j = 1, 2, \dots, N)$, the proposed $\mathcal{K}$O-FKL divergence measure (in Eq. (21)) is implemented. And the divergence value obtained is utilized as an element in constructing the conflict measure matrix (CMM). The CMM is defined as:

$$CMM = \begin{bmatrix} 0 & \mathcal{K}\text{O-FKL}_{12} & \cdots & \mathcal{K}\text{O-FKL}_{1N} \\ \mathcal{K}\text{O-FKL}_{21} & 0 & \cdots & \mathcal{K}\text{O-FKL}_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{K}\text{O-FKL}_{N1} & \mathcal{K}\text{O-FKL}_{N2} & \cdots & 0 \end{bmatrix}. \quad (23)$$

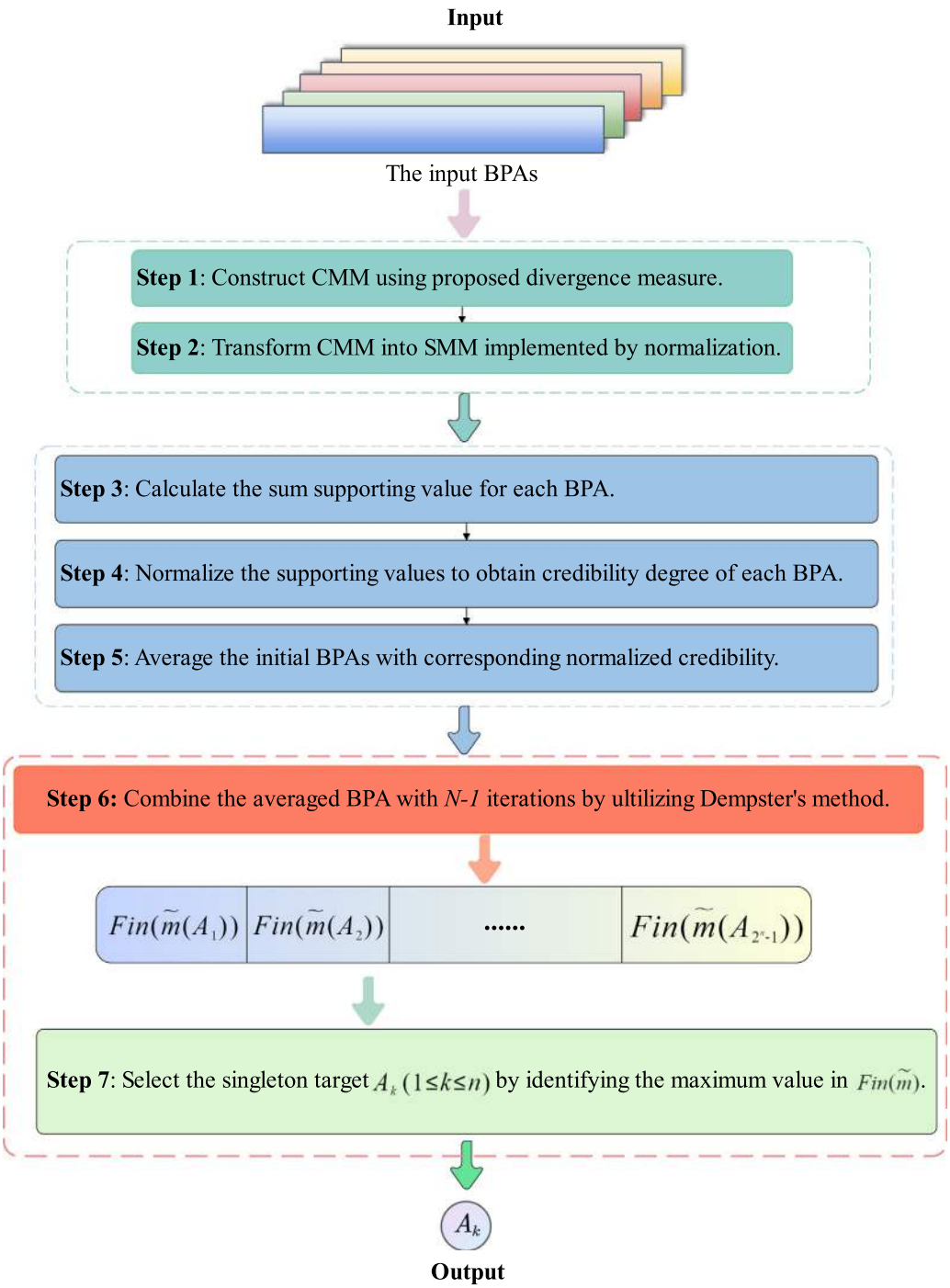


Fig. 10. The flowchart of $\mathcal{K}$O-FKL-MSDF algorithm.

Step 2 Subsequently, a normalization method is applied to construct a similarity measure matrix (SMM).

$$SMM = \begin{bmatrix} 1 & S_{12} & \cdots & S_{1N} \\ S_{21} & 1 & \cdots & S_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ S_{N1} & S_{N2} & \cdots & 1 \end{bmatrix}, \tag{24}$$

where $S_{ij} = \frac{\max(\mathcal{K}\text{O-FKL}_{ij}) - \mathcal{K}\text{O-FKL}_{ij}}{\max(\mathcal{K}\text{O-FKL}_{ij}) - \min(\mathcal{K}\text{O-FKL}_{ij})}$ and the diagonal element S_{ii} is assigned a constant value of 1 to represent the self-similarity of a BPA itself.

Step 3 The supporting value $Sup(m_i)$ of a BPA is calculated by summing up the support values that other BPAs hold for it in SMM. Mathematically, it is defined as:

$$Sup\left(m_i\right) = \sum_{j=1, j\neq i}^{N} e^{S_{ij}}. \tag{25}$$

Step 4 Once the total supporting value $Sup\left(m_i\right)$ of each BPA has been calculated, the credit value Crd_i of each BPA is obtained by normalizing the supporting values with respect to the sum of all supporting values:

$$Crd_i = \frac{Sup\left(m_i\right)}{\sum_{i=1}^{N} Sup\left(m_i\right)}. \tag{26}$$

Step 5 Every BPA is updated using the normalized supporting value Crd_i, and an averaged BPA $\tilde{m}$ is obtained by weighted averaging of these normalized BPAs m_i:

$$\tilde{m} = \sum_{i=1}^{N} \left(Crd_i \times m_i\right). \tag{27}$$

Step 6 The averaged BPA $\tilde{m}$ is further combined $N-1$ times using Dempster's combination rule, resulting in the final combined belief Fin $(\tilde{m})$:

$$Fin\left(\tilde{m}\right) = \left(\left(\left(\left(\tilde{m} \oplus \tilde{m}\right)_1 \oplus \tilde{m}\right)_2 \oplus \ldots\right)_{N-2} \oplus \tilde{m}\right)_{N-1}. \tag{28}$$

Algorithm 1: Pseudocode procedure for $\mathcal{K}$O-FKL-MSDF.

Input: A set of N BPAs $\mathcal{M} = \left\{m_1, m_2, \ldots, m_N\right\}$ (each BPA contains n mutually exclusive events)
Output: The supported event based on fusion results

```
for i ← 1 to N do
    for j ← 1 to N do
        Calculate the KO-FKL divergence between m_i and m_j using Eq. (21);
    end
end
Construct the conflict measure matrix CMM using Eq. (23);
Compute the similarity measure matrix SMM using Eq. (24);
for i ← 1 to N do
    Calculate the support Sup_i of each BPA using Eq. (25);
end
for i ← 1 to N do
    Normalize Sup_i into credibility weights Crd_i using Eq. (26);
end
for i ← 1 to N do
    Compute the average of the input BPAs using Crd_i weights to obtain m̃ based on Eq. (27);
end
for i ← 1 to N − 1 do
    Combine m̃ with itself N − 1 times using Dempster's method to obtain Fin(m̃) based on Eq. (28);
end
Select the target event A_k (1 ⩽ k ⩽ n) with the highest singleton-supporting value from Fin(m̃) using Eq. (29);
```

Step 7 The target event A_k $(1 \leqslant k \leqslant n)$ is chosen based on the highest supporting value in $Fin\left(\tilde{m}\right)$, which is given by:

$$A_k \leftarrow \max_{1 \leqslant i \leqslant n} \left\{Fin\left(\tilde{m}\left(A_i\right)\right)\right\}. \tag{29}$$

4.2. Computational complexity analysis of Ko-FKL-MSDF

In this section, time and space complexity of each step in the proposed $\mathcal{K}$O-FKL-MSDF method is first analyzed. Furthermore, the proposed algorithm is compared with other classical multi-source data fusion algorithms in the computational complexity.

Firstly, a detailed analysis of the computational complexity of the proposed $\mathcal{K}$O-FKL-MSDF algorithm is provided in Table 4. In **Step 1**, the CMM is constructed by calculating the $\mathcal{K}$O-FKL divergence between each pair of BPAs. There are N^2 elements in the CMM. And in each process of $\mathcal{K}$O-FKL algorithm, it is executed at least $\mathcal{K}$ times in Eq. (21). Thus, this step results in the time complexity of $\mathcal{O}(N^2) \times \mathcal{O}(\mathcal{K} \times 2^n)$ and the space complexity of $\mathcal{O}(N^2) \times \mathcal{O}(2^n)$. In **Step 2**, the CMM is transformed into an SMM, which takes up the time complexity of $\mathcal{O}(N^2)$ and the space complexity of $\mathcal{O}(N^2) \times \mathcal{O}(2^n)$. **Step 3** computes the total supporting value of each BPA, and has the time complexity of $\mathcal{O}(N^2)$ and the space complexity of $\mathcal{O}(N)$. In **Step 4**, the credit value of each BPA is computed by normalizing the supporting obtained in the last step, and consumes $\mathcal{O}(N)$ in time complexity and $\mathcal{O}(N)$ in space complexity. **Step 5** averages all the BPAs to obtain $\tilde{m}$, which has the time complexity of $\mathcal{O}(N)$ and the space complexity of $\mathcal{O}(1) \times \mathcal{O}(2^n)$. Finally, in **Step 6**, Dempster's rule (in Eq. (7)) is applied $n-1$ times. So this consumes $\mathcal{O}(N) \times \mathcal{O}(2^n)$ in time complexity and $\mathcal{O}(1) \times \mathcal{O}(2^n)$ in space complexity. Therefore, the overall time complexity and space complexity of $\mathcal{K}$O-FKL-MSDF algorithm are $\mathcal{O}(N^2) \times \mathcal{O}(\mathcal{K} \times 2^n)$ and $\mathcal{O}(N^2) \times \mathcal{O}(2^n)$, respectively.

In the following, a comprehensive computational complexity analysis for the other classical multi-source fusion algorithms, which are introduced in Section 2, is discussed and summarized in Table 5. This analysis will allow us to better understand the strengths and weaknesses of each algorithm, and to evaluate the performance of the proposed method in practical applications. First, one advantage of the Dempster's combination rule based multi-source fusion algorithm (DCR-MSDF) is its low computational complexity, with a time complexity of $\mathcal{O}(N) \times \mathcal{O}(2^n)$ and a space complexity of $\mathcal{O}(2^n)$. This is because it does not require the construction of a matrix like CMM or SMM to reflect the relationship between two BPAs. Establishing a CMM and adopting a similar algorithmic framework, the BJS-MSDF, $\mathfrak{B}$-MSDF, and FBD_{SKL}-MSDF have time and space complexity of $\mathcal{O}(N^2) \times \mathcal{O}(2^n)$. As for the $\mathcal{K}$O-FKL-MSDF algorithm, even though it also utilizes the CMM matrix to measure the degree of conflict between BPAs, its time complexity is affected by $\mathcal{K}$ due to the high-order fractal iteration in $\mathcal{K}$O-FKL. Therefore, the time complexity of $\mathcal{K}$O-FKL-MSDF algorithm can be expressed as $\mathcal{O}(N^2) \times \mathcal{O}(\mathcal{K} \times 2^n)$. As for its space complexity, because there is no need to store all HO-FBBPAs during the iteration process, the space complexity of the proposed $\mathcal{K}$O-FKL-MSDF algorithm can be optimized to $\mathcal{O}(N^2) \times \mathcal{O}(2^n)$.

Even though the proposed $\mathcal{K}$O-FKL-MSDF has a higher time complexity, a more comprehensive analysis of these algorithms is conducted in Sections 5 and 6 to further demonstrate the advantages of this algorithm.

5. Experiments

In this section, a concrete multi-source data fusion experiment is conducted using a dataset obtained from Deng et al. (2004). To compare the proposed $\mathcal{K}$O-FKL-MSDF algorithm with existing methods, the classical Dempster's combination rule based multi-source fusion algorithm (DCR-MSDF) (Dempster, 1967) is firstly considered, and the divergence-based multi-source fusion algorithms are taken into consideration. They include BJS-MSDF (Xiao, 2019), $\mathfrak{B}$-MSDF (Xiao, 2020), and FBD_{SKL}-MSDF (Zeng & Xiao, 2023), which are introduced in Section 2. In the process of utilizing the proposed $\mathcal{K}$O-FKL-MSDF algorithm, based on the detailed steps given in Section 4, the intermediate results are provided. Furthermore, a sensitivity analysis is performed on these algorithms in order to verify their robustness.

5.1. Problem statement

Suppose the frame of discernment Θ contains three mutually exclusive and exhaustive hypotheses, $\{A\}$, $\{B\}$, and $\{C\}$, which are independent of each other. Assume there are five BPAs m_1, m_2, m_3, m_4, and m_5 defined on this discernment Θ and are shown in Table 6.

5.2. Implementation of Ko-FKL-MSDF algorithm

Step 1 The CMM is constructed according to the proposed $\mathcal{K}$O-FKL measure, and the resulting matrix is presented as follows::

$$CMM = \begin{bmatrix} 0 & 0.0439 & 0.0045 & 0.0111 & 0.3043 \\ 0.0439 & 0 & 0.0410 & 0.0653 & 0.4282 \\ 0.0045 & 0.0410 & 0 & 0.0038 & 0.3176 \\ 0.0111 & 0.0653 & 0.0038 & 0 & 0.3653 \\ 0.3043 & 0.4282 & 0.3176 & 0.3653 & 0 \end{bmatrix}.$$

Table 4
Detailed computational analysis of $\mathcal{K}O$-FKL-MSDF algorithm.

Steps	Time complexity	Space complexity	Description
Step 1	$\mathcal{O}(N^2)\times\mathcal{O}(\mathcal{K}\times 2^n)$	$\mathcal{O}(N^2)\times\mathcal{O}(2^n)$	Construct the CMM using the $\mathcal{K}$O-FKL measure
Step 2	$\mathcal{O}(N^2)$	$\mathcal{O}(N^2)\times\mathcal{O}(2^n)$	Transform CMM into SMM
Step 3	$\mathcal{O}(N^2)$	$\mathcal{O}(N)$	Compute the total supporting value of each BPA
Step 4	$\mathcal{O}(N)$	$\mathcal{O}(N)$	Assign credit values to each BPA by normalizing the supporting values
Step 5	$\mathcal{O}(N)$	$\mathcal{O}(1)\times\mathcal{O}(2^n)$	Combine the BPAs into $\widetilde{m}$ by computing the weighted average using the allocated credit values.
Step 6	$\mathcal{O}(N)\times\mathcal{O}(2^n)$	$\mathcal{O}(1)\times\mathcal{O}(2^n)$	Combine the BPAs $\widetilde{m}$ using Dempster's rule to obtain the final fused BPA, $Fin(\widetilde{m})$.
Overall	$\mathcal{O}(N^2)\times\mathcal{O}(\mathcal{K}\times 2^n)$	$\mathcal{O}(N^2)\times\mathcal{O}(2^n)$	Computational complexity

Notations:
N : represents the number of BPAs to be fused.
n : represents the number of exclusive events in the frame of discernment Θ.

Table 5
Comparison of computational complexity across different algorithms.

Algorithm	Time complexity	Space complexity
DCR-MSDF (Dempster, 1967)	$\mathcal{O}(N)\times\mathcal{O}(2^n)$	$\mathcal{O}(2^n)$
BJS-MSDF (Xiao, 2019)	$\mathcal{O}(N^2)\times\mathcal{O}(2^n)$	$\mathcal{O}(N^2)\times\mathcal{O}(2^n)$
$\mathfrak{B}$-MSDF (Xiao, 2020)	$\mathcal{O}(N^2)\times\mathcal{O}(2^n)$	$\mathcal{O}(N^2)\times\mathcal{O}(2^n)$
FBD_{SKL}-MSDF (Zeng & Xiao, 2023)	$\mathcal{O}(N^2)\times\mathcal{O}(2^n)$	$\mathcal{O}(N^2)\times\mathcal{O}(2^n)$
$\mathcal{K}$O-FKL-MSDF	$\mathcal{O}(N^2)\times\mathcal{O}(\mathcal{K}\times 2^n)$	$\mathcal{O}(N^2)\times\mathcal{O}(2^n)$

Table 6
Initial BPAs with their belief values.

BPA	$\{A\}$	$\{B\}$	$\{C\}$	$\{A,C\}$
m_1	0.55	0.15	0.15	0.15
m_2	0.56	0.04	0.01	0.39
m_3	0.60	0.10	0.20	0.10
m_4	0.60	0.09	0.11	0.20
m_5	0.01	0.25	0.65	0.09

Step 2 The constructed CMM above is then converted into an SMM as follows:

$$SMM=\begin{bmatrix} 1 & 0.9056 & 0.9984 & 0.9828 & 0.2920\\ 0.9056 & 1 & 0.9123 & 0.8552 & 0.0000\\ 0.9984 & 0.9123 & 1 & 1.0000 & 0.2605\\ 0.9828 & 0.8552 & 1.0000 & 1 & 0.1483\\ 0.2920 & 0.0000 & 0.2605 & 0.1483 & 1 \end{bmatrix}.$$

Step 3 The sum of supporting values for each BPA is calculated as follows:

$$Sup(m_1)=9.1983,$$
$$Sup(m_2)=8.3151,$$
$$Sup(m_3)=9.2197,$$
$$Sup(m_4)=8.9018,$$
$$Sup(m_5)=4.7965.$$

Step 4 The sum supporting values are then normalized to obtain credit weights for each BPA as follows:

$$Crd(m_1)=0.2275,$$
$$Crd(m_2)=0.2057,$$
$$Crd(m_3)=0.2280,$$
$$Crd(m_4)=0.2202,$$
$$Crd(m_5)=0.1186.$$

Step 5 The initial BPAs are averaged to generate $\widetilde{m}$ as follows:

$$\widetilde{m}(\{A\})=0.5105,$$
$$\widetilde{m}(\{B\})=0.1146,$$
$$\widetilde{m}(\{C\})=0.1831,$$
$$\widetilde{m}(\{A,C\})=0.1918.$$

Table 7
Recognition results using different fusion algorithms (in %).

Method	$\{A\}$	$\{B\}$	$\{C\}$	$\{A,C\}$	Target
DCR-MSDF (Dempster, 1967)	81.78	0.03	17.96	0.23	A
BJS-MSDF (Xiao, 2019)	95.19	0.01	4.66	0.15	A
$\mathfrak{B}$-MSDF (Xiao, 2020)	94.83	0.02	5.02	0.14	A
FBD_{SKL}-MSDF (Zeng & Xiao, 2023)	95.12	0.01	4.72	0.15	A
$\mathcal{K}$O-FKL-MSDF	95.83	0.01	4.02	0.15	A

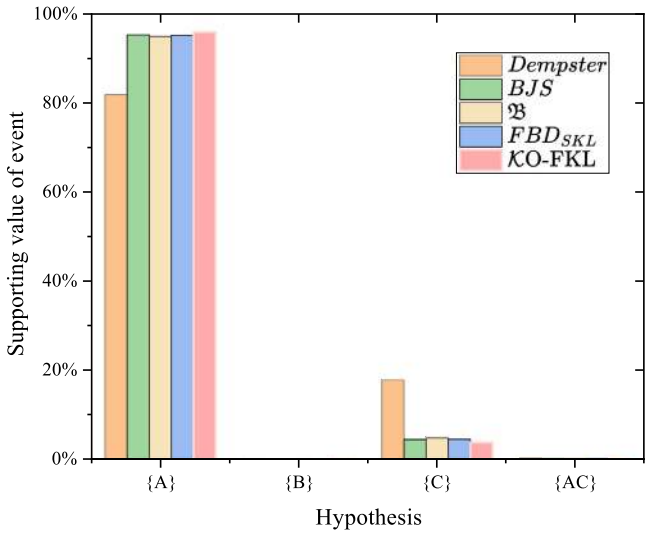


Fig. 11. Sensitivity analysis generated by different methods.

Step 6 The averaged BPA $\widetilde{m}$ is then combined $n-1$ times using Dempster's rule to obtain the final answer $Fin(\widetilde{m})$, which is presented in Fig. 11 and Table 7.

Step 7 The target event A is selected.

In Table 7, it can be observed that event A is identified as the target with recognition values generated by $\mathcal{K}$O-FKL-MSDF algorithm and the other four baseline multi-source algorithms.

Specifically, DCR-MSDF method yields a relatively low belief value of 81.78% for the target A among the five methods. This result may be attributed to the highly conflicting BPA m_5, which results in a lower belief value for A. However, DCR-MSDF method is not capable of properly handling highly conflicting BPAs. As for the belief-based divergence algorithms, BJS-MSDF and $\mathfrak{B}$-MSDF recognized the target A with recognition values of 95.19% and 94.83%, respectively. However, as shown in Examples 1 and 2, their performance may be limited in certain situations. On the other hand, the proposed FBD_{SKL}-MSDF algorithm recognized the target A with the highest belief value of 95.12%. It should be noted, however, that only one fractal iteration may not fully demonstrate the advantages of the fractal method. Therefore, in the $\mathcal{K}$O-FKL-MSDF algorithm, a higher order can better

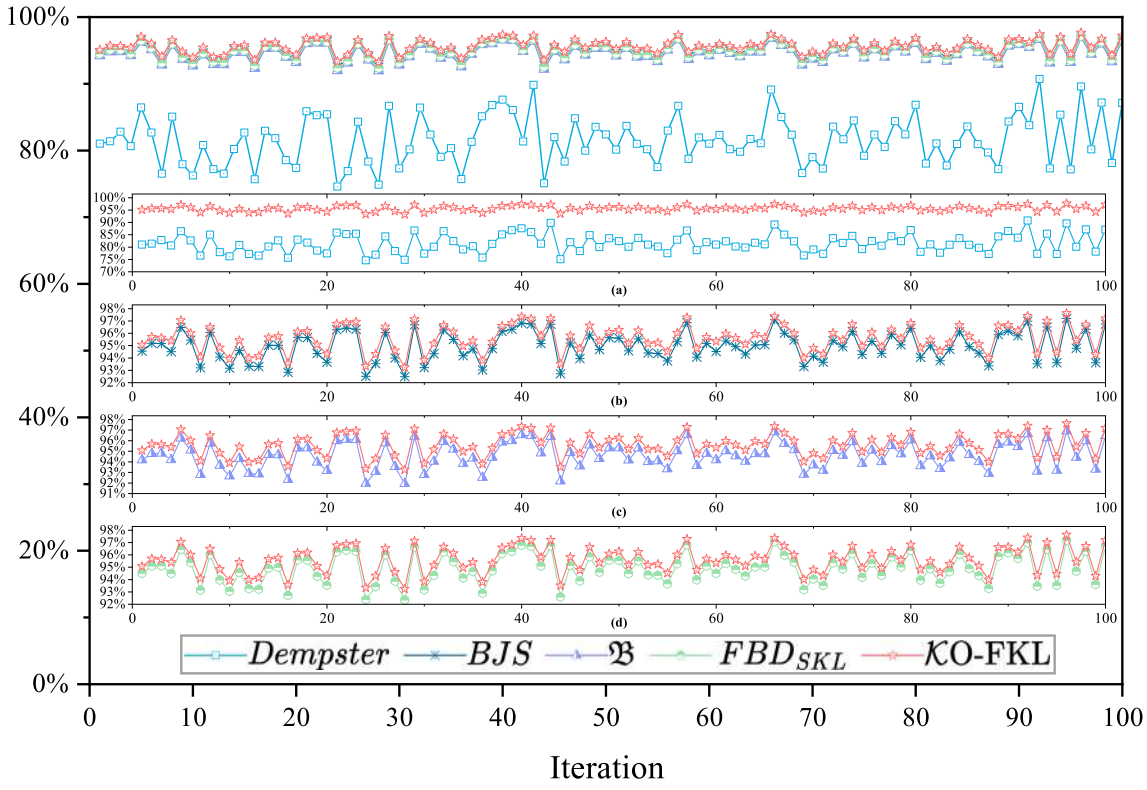


Fig. 12. Sensitivity analysis results of different fusion methods. (**a**) Performance comparison: Dempster vs. $\mathcal{K}$O-FKL-MSDF. (**b**) Performance comparison: BJS vs. $\mathcal{K}$O-FKL-MSDF. (**c**) Performance comparison: $\mathfrak{B}$ vs. $\mathcal{K}$O-FKL-MSDF. (**d**) Performance comparison: FBD_{SKL} vs. $\mathcal{K}$O-FKL-MSDF.

Table 8
Comparison of recognition rates and variances among different fusion methods.

Method	Mean (%)	Variance ($\times 10^{-3}$)	Target
DCR-MSDF (Dempster, 1967)	81.63	1.3589	A
BJS-MSDF (Xiao, 2019)	94.98	0.1371	A
$\mathfrak{B}$-MSDF (Xiao, 2020)	94.61	0.1516	A
FBD_{SKL}-MSDF (Zeng & Xiao, 2023)	94.91	0.1414	A
$\mathcal{K}$O-FKL-MSDF	95.61	0.1144	A

reveal the relationships between BPAs. This results in a recognition value of 95.83%, which is the highest among all the methods. The recognition values are also visually presented in Fig. 11. The figure indicates that $\mathcal{K}$O-FKL-MSDF algorithm can effectively manage highly conflicting data and precisely identify the correct target in multi-source data fusion.

5.3. Sensitivity analysis

In this section, a sensitivity analysis is conducted to examine the superiority and robustness of the proposed $\mathcal{K}$O-FKL-MSDF method. To accomplish this sensitivity analysis, a random fluctuation error with a range of $[-0.1, 0.1]$ is defined for each focal element of m_1. After adding noise to all focal elements of m_1, it is then normalized. Then, the recognition accuracy of target $\{A\}$ is compared among DCR-MSDF, divergence-based fusion algorithms (including BJS-MSDF, $\mathfrak{B}$-MSDF, and FBD_{SKL}-MSDF), and our proposed $\mathcal{K}$O-FKL-MSDF method. To eliminate experimental randomness and further verify the robustness of the model, we repeat the experiment independently 100 times and summarize the final results in Fig. 12.

From Fig. 12(a), a comparison between the proposed $\mathcal{K}$O-FKL-MSDF algorithm and DCR-MSDF algorithm is made. It can be observed that DCR-MSDF method mostly focuses on a recognition rate ranging from 70% to 85%, and is greatly affected by noise. According to Table 8, it can also be seen that the average recognition rate and variance of DCR-MSDF method in this experiment are 81.63% and 1.3589×10^{-3}, respectively, which differs significantly from the results of other methods. This is because the Dempster's classic algorithm is easily affected by conflicting BPAs in multi-source data fusion. However, abnormal samples are common in practical applications, and using DCR-MSDF method may significantly affect the results in such cases. As for the divergence-based algorithms, shown in Fig. 12(b), (c), and (d), it can be observed that the results of the proposed $\mathcal{K}$O-FKL-MSDF algorithm show significant improvement compared to BJS-MSDF, $\mathfrak{B}$-MSDF, and FBD_{SKL}-MSDF algorithms. In Table 8, the average recognition rates obtained by BJS-MSDF, $\mathfrak{B}$-MSDF, FBD_{SKL}-MSDF and proposed $\mathcal{K}$O-FKL-MSDF are 94.98%, 94.61%, 94.91% and 95.61%, respectively. Meanwhile, the variances of these methods' recognition rates (10^{-3}) during the experiment are 0.1371, 0.1516, 0.1414 and 0.1144, respectively. In terms of average recognition rate, $\mathcal{K}$O-FKL-MSDF outperforms other divergence-based methods by about 0.5% to 1%, and the significant decrease in variance also indicates that the method has stronger robustness compared to other methods.

In general, the $\mathcal{K}$O-FKL-MSDF algorithm outperformed other algorithms in terms of average recognition rate and stability in this sensitivity analysis experiment.

6. Application

In recent years, pattern classification has attracted significant attention (Ko & Koo, 2023; Liu, Fu et al., 2022; Liu et al., 2021; Xiao, Cao et al., 2022; Xiao, Wen et al., 2022), as it serves as a benchmark for evaluating the performance and effectiveness of various methods in tackling complex problems. To showcase the practical applicability and robustness of the proposed $\mathcal{K}$O-FKL-MSDF algorithm, this section focuses on its application to classification tasks using real-world datasets,

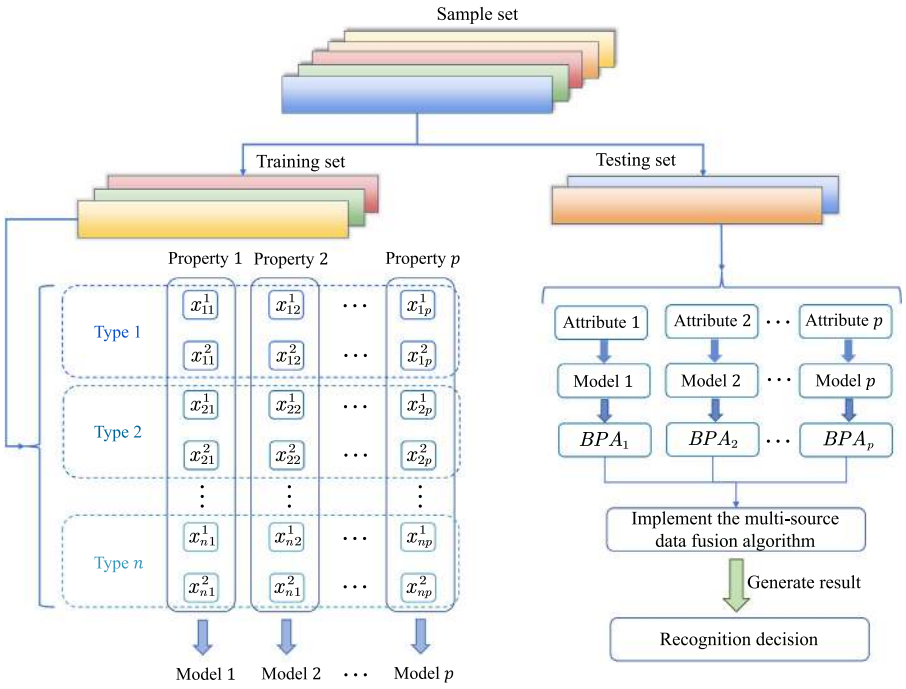


Fig. 13. The flowchart of experiment process.

Table 9
Dataset information.

Dataset	Classes number	Instances number	Attributes number
Iris	3	150	4
Seed	3	210	7
Wine	3	178	13
Breast Cancer	2	569	32
Skin	2	245 057	3

comparing its performance against both classical DCR-MSDF (Dempster, 1967) and divergence-based algorithms, such as BJS-MSDF (Xiao, 2019), $\mathfrak{B}$-MSDF (Xiao, 2020), and FBD_{SKL}-MSDF (Zeng & Xiao, 2023). By doing so, the strengths of the proposed $\mathcal{K}$O-FKL-MSDF algorithm are highlighted in addressing complex classification problems, further substantiating its potential for real-world applications.

6.1. Experiment design and implementation

Firstly, to demonstrate the robustness and versatility of the proposed $\mathcal{K}$O-FKL-MSDF algorithm, five diverse datasets, each of varying complexity and from different fields, are assembled. These datasets are drawn from the UCI repository,[1] and comprised the Iris, Seeds, Wine, and Breast Cancer Wisconsin (Diagnostic) datasets. Importantly, the Skin dataset is incorporated into the study. With an impressive total of 245 057 instances, this dataset serves as a benchmark for testing the scalability of the algorithm, thus enhancing the reliability and generalizability of the results. The datasets also exhibit a wide range of feature dimensions, reflecting the diversity in complexity. The characteristics of these datasets are summarized in Table 9.

In the experiment, for each type in a dataset, 40% of the samples from each type are randomly selected as the test set, while the remaining 60% is used as the training set. As shown in Fig. 13, the values of each sample can be converted into a BPA. For the training dataset, the number of attributes p can generate p Gaussian models. In the process of recognizing each sample in the test set, each trained Gaussian model converts each attribute of a test sample into a BPA. Through a specific multi-source data fusion algorithm, these BPAs can be fused to generate the final decision result. Therefore, the recognition rate can be calculated by confirming whether each sample is classified successfully. This process is repeated 100 times, with the training set and test set randomly reselected in each independent experiment.

To compare the practical application ability of the algorithms in classification tasks, DCR-MSDF, classic divergence-based methods (BJS-MSDF, $\mathfrak{B}$-MSDF, and FBD_{SKL}-MSDF), and the proposed $\mathcal{K}$O-FKL-MSDF algorithm are selected for result comparison. The weighted average recognition rate, denoted as $\overline{Accuracy}$, is calculated as follows:

$$\overline{Accuracy} = \sum_{j=1}^{T} Accuracy_j \times \frac{n_j}{N}, \tag{30}$$

where T represents the total number of types, $Accuracy_j$ denotes the recognition rate for type j, n_j denotes the number of samples for type j, and N denotes the total number of samples. This formula provides a generalizable way to calculate the overall recognition rate across different types within a dataset.

In Fig. 14, the variations in recognition rates of DCR-MSDF, classic divergence-based methods (BJS-MSDF, $\mathfrak{B}$-MSDF, and FBD_{SKL}-MSDF), and the proposed $\mathcal{K}$O-FKL-MSDF algorithm across 100 independent experiments on the five datasets can be visually observed. From the characteristics of the datasets, it is observed that the Iris dataset generally yields higher recognition rates than the other four datasets. This may be attributed to its fewer features, as shown in Table 9, which reduces complexity and potential interference from anomalous samples. For the proposed $\mathcal{K}$O-FKL-MSDF algorithm, it demonstrates relatively high recognition rates on all five datasets, with the rates exceeding 90% in most independent experiments.

The mean recognition rates and corresponding variances of five fusion algorithms across five datasets are detailed in Tables 10 and 11, providing a comprehensive perspective on the performance and stability of each algorithm over 100 independent experiments. Among the algorithms, Dempster's combination rule multi-source data fusion (DCR-MSDF) is included for comparison. It is observed that the proposed $\mathcal{K}$O-FKL-MSDF algorithm consistently yields superior recognition rates, thereby outperforming the other fusion algorithms across

[1] https://archive.ics.uci.edu/ml/datasets/.

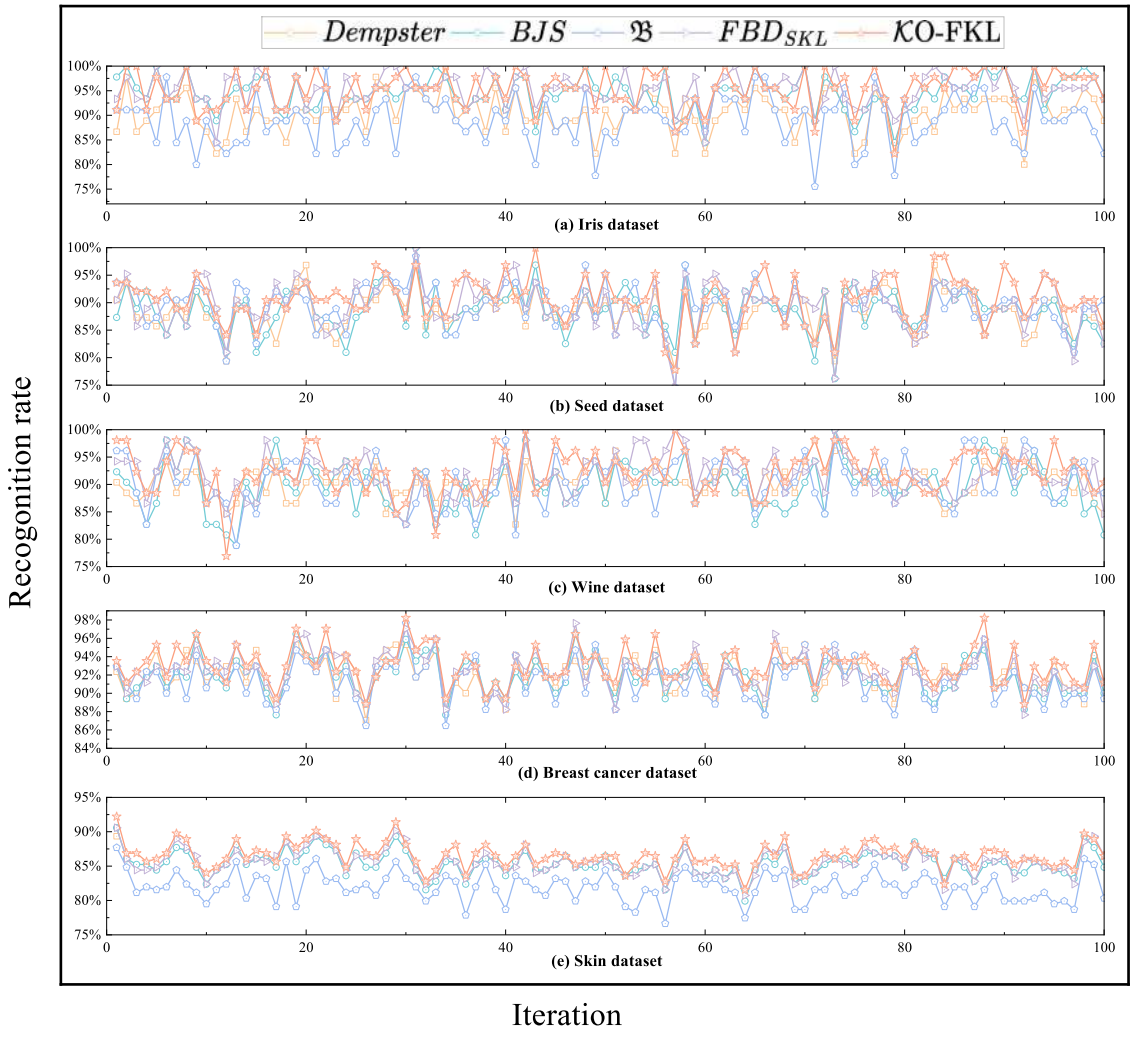


Fig. 14. Recognition rates generated by fusion algorithms on different datasets. (**a**) Application performance on Iris dataset. (**b**) Application performance on seed dataset. (**c**) Application performance on wine dataset. (**d**) Application performance on Breast Cancer Wisconsin dataset. (**e**) Application performance on Skin dataset.

Table 10
Comparison of mean recognition rates (in %) for different fusion algorithms on five datasets.

Method	Iris	Seed	Wine	Breast cancer	Skin
DCR-MSDF	90.56	88.37	90.52	92.11	84.06
BJS-MSDF	94.53	88.67	89.75	91.96	85.40
$\mathfrak{B}$-MSDF	89.00	89.32	90.38	91.51	81.97
FBD_{SKL}-MSDF	95.20	89.32	91.67	92.38	85.62
$\mathcal{K}$O-FKL-MSDF	95.33	90.62	92.37	92.98	86.56

Table 11
Comparison of variances (in units of 10^{-4}) for different fusion algorithms on five datasets.

Method	Iris	Seed	Wine	Breast cancer	Skin
DCR-MSDF	16.10	13.69	9.21	3.63	3.95
BJS-MSDF	12.51	16.34	19.88	3.97	3.26
$\mathfrak{B}$-MSDF	25.38	16.13	19.20	5.02	4.53
FBD_{SKL}-MSDF	11.94	20.45	14.72	4.12	3.70
$\mathcal{K}$O-FKL-MSDF	15.91	17.41	16.25	4.08	3.47

all datasets. Specifically, the $\mathcal{K}$O-FKL-MSDF algorithm attained mean recognition rates of 95.33%, 90.62%, 92.37%, 92.98%, and 86.56% on the Iris, Seed, Wine, Breast cancer, and Skin datasets, respectively. This superior performance over the other four fusion methods is a clear indication of the enhanced recognition capability of the proposed $\mathcal{K}$O-FKL-MSDF algorithm.

With an overview of the comparative performance of the $\mathcal{K}$O-FKL-MSDF algorithm with the initial four fusion methods (DCR-MSDF, BJS-MSDF, $\mathfrak{B}$-MSDF, and FBD_{SKL}-MSDF) in Table 10, it is apparent that the proposed $\mathcal{K}$O-FKL-MSDF algorithm offers a considerable enhancement in recognition rates across all datasets. This clear-cut advantage sets the stage for a more focused comparison with the FBD_{SKL}-MSDF algorithm. Upon a more focused comparison, it is noted that the $\mathcal{K}$O-FKL-MSDF algorithm offers a marginal improvement over the FBD_{SKL}-MSDF on the Iris dataset, with an increase in the mean recognition rate of just 0.13%. However, on the Seed, Wine, Breast cancer, and Skin datasets, the improvements in recognition rates are considerably more pronounced, reaching 1.3%, 0.7%, 0.6%, and 0.94%, respectively.

In addition to its impressive recognition rates, the $\mathcal{K}$O-FKL-MSDF algorithm also demonstrates a relatively stable performance as evidenced by the small variances observed over 100 epochs in Table 11. This suggests a level of robustness in the proposed method, indicating that it can consistently deliver high-quality results across different

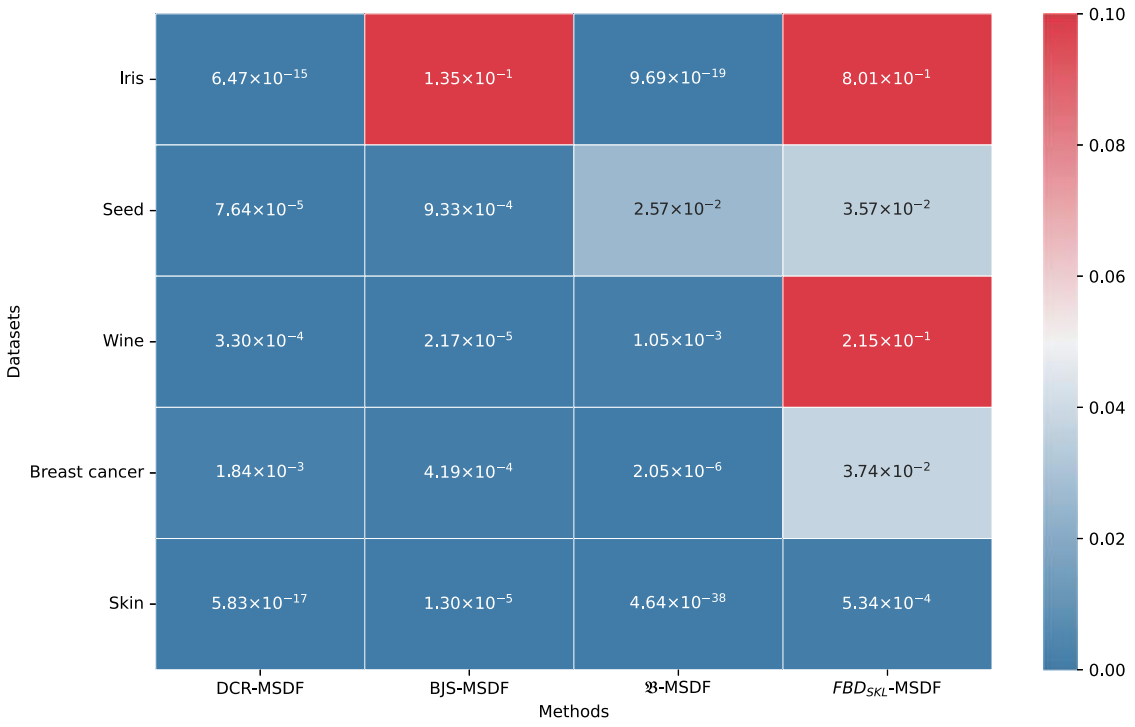


Fig. 15. Heatmap of p-values from Welch's t-test based on application results.

applications. These findings intuitively underscore the superiority and robustness of the $\mathcal{K}$O-FKL-MSDF algorithm in practical classification applications.

6.2. Statistical analysis

To statistically substantiate the superior performance of our $\mathcal{K}$O-FKL-MSDF algorithm, as suggested by the higher recognition rates and low variances observed in Section 6.1, the Welch's t-test is conducted. The t-test is applied to the recognition rates obtained from the 100 independent experiments performed on each dataset using each fusion algorithm, as illustrated in Fig. 14. Note that standard t-tests assume equal variances in the populations from which samples are drawn. However, from Table 11, it is evident that there are variations in the variances of the results obtained from different fusion algorithms. Therefore, the Welch's t-test is opted, which accommodates unequal variances and offers a more reliable option when the assumption of equal variances is violated (Welch, 1947).

For each dataset, a baseline method is selected and the Welch's t-test is conducted between the recognition rates of this baseline method and those of our $\mathcal{K}$O-FKL-MSDF algorithm. The resultant p-value is examined, with a value less than 0.05 indicating a statistically significant difference between the recognition rates of the two methods, at a confidence level of 95% or higher, as established by Welch (1947). Furthermore, the smaller the p-value, the more significant this difference. Accordingly, the p-values are presented and visualized in a heatmap, depicted in Fig. 15.

From Fig. 15, those p-values less than 0.05 are colored on a gradient from white to dark blue, whereas p-values greater than 0.05 are represented on a gradient from white to dark red. It is clear from the heatmap that out of the 20 t-tests conducted to compare our method with four baseline methods across five datasets, only three result in p-values greater than 0.05. Three tests generated p-values less than 0.05 but greater than 0.01, indicating a confidence level of at least 95% that there is a significant difference in the recognition rates of the two methods across the 100 independent experiments. The remaining 14 tests yield p-values far less than 0.01, suggesting a confidence level exceeding 99% that a significant difference exists in the recognition rates of the two methods over 100 independent experiments. In addition, when focusing specifically on the Skin dataset, which comprises 245 057 instances and thus provides a robust verification for the reliability and generalizability of our results, even the highest p-value obtained when comparing our proposed $\mathcal{K}$O-FKL-MSDF algorithm with the four baselines is a mere 5.34×10^{-4}. This further underscores the superior recognition rate and convincing performance of our method on this broadly applicable dataset.

6.3. Error analysis

In Section 6.2, the superiority in recognition rates and robustness of the proposed $\mathcal{K}$O-FKL-MSDF algorithm has been statistically validated using Welch's t-test. This section delves further into the behavior of the proposed $\mathcal{K}$O-FKL-MSDF algorithm by investigating its recognition patterns and associated errors. This analysis aims to uncover the underlying challenges that might influence the algorithm's performance, hence identifying potential areas for further refinement.

As seen in Table 10 from Section 6.1, while the proposed $\mathcal{K}$O-FKL-MSDF algorithm exhibits superior performance across various datasets, the average recognition rate over 100 independent experiments on the Skin dataset is 86.56%. Even though this rate represents a significant improvement over other baseline algorithms, it is comparatively lower than the recognition rates achieved on other datasets, as can be visually observed from Fig. 14 in Section 6.1. This discrepancy suggests that there may be potential for further refining the proposed algorithm to improve its performance on this particular dataset. The Skin dataset is composed of two types, with 194,198 samples for type 1 and 50,859 samples for type 2. For each type, 40% of the samples are randomly selected for testing during each run of the experiment. The recognition rates for each type are then calculated, and the weighted average recognition rate, denoted by $\overline{Accuracy}$, is computed using the formula given in Eq. (30). After 100 such independent experiments, the final weighted recognition rate is found to be 86.56%. Understanding the recognition rates of each type could help uncover the factors contributing to the comparatively lower overall recognition rate observed

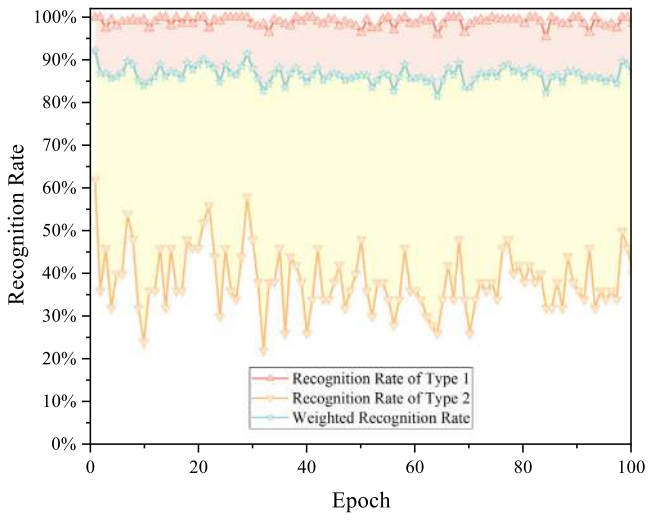


Fig. 16. Comparative analysis of recognition rates for the Skin dataset.

in this dataset, thereby providing insights for the further refinement of the algorithm.

As depicted in Fig. 16, the recognition rates of Type 1, Type 2, and the weighted recognition rate calculated using Eq. (30) across the 100 independent experiments are presented. It is evident from the figure that the recognition rate for Type 1 remains consistently high, almost reaching a 100% recognition rate. However, the recognition rate for Type 2 is considerably lower, hovering around 45% with noticeable fluctuations. These discrepancies in recognition rates are visually represented in the figure using a colored area chart. Upon statistical analysis, the recognition rates for Type 1 and Type 2 across the 100 experiments are found to be 98.93% and 38.59%, respectively. The final weighted recognition rate, computed using Eq. (30), is 86.56%.

In an attempt to understand the underlying causes for the discrepancies in recognition rates between Type 1 and Type 2, an analysis of the individual attributes within the Skin dataset is conducted. As depicted in Fig. 17, the box plots for Attributes 1, 2, and 3 are presented, comparing Classification Types 1 and 2. From the box plots, it is observed that Attributes 1 and 2 do not show substantial differences between the two types, with Type 1's value range almost entirely encompassing that of Type 2. This overlap potentially leads to confusion and ambiguity in classification, as Type 2 may be misclassified as Type 1. This observation aligns with the phenomenon seen in Fig. 16, where Type 2 has a considerably lower recognition rate. In contrast, Attribute 3 demonstrates a distinct separation between the types, indicating its significance in differentiating between Types 1 and 2.

A further analysis of different attribute combinations on recognition error rates is further explored and presented in Fig. 18. In this figure, the average error rates over 100 independent experiments are shown for four distinct attribute combinations: Attributes 1, 2, 3; Attributes 1, 2; Attributes 1, 3; and Attributes 2, 3. For each combination, the error rates for Type 1, Type 2, and the weighted error rate are provided. The analysis reveals that the combination of Attributes 1 and 2 leads to the highest error rate for Type 2, reaching 74.36%. On the other hand, the combinations of Attributes 1 and 3 or Attributes 2 and 3 result in a significant reduction in the error rate for Type 2, falling to approximately 29.36% and 29.46%, respectively. Furthermore, the weighted error rates for these two combinations are 12.23% and 11.64%, respectively, lower than the 13.44% error rate when all three attributes are considered.

An in-depth examination of the Skin dataset's features reveals the underlying errors made by the algorithm. A significant factor contributing to these errors is the algorithm's treatment of attributes. Specifically, the attributes that contribute less to type classification are not assigned lower weights, highlighting an area of potential improvement in the weighting strategy. An examination of Figs. 17 and 18 reveals that Attributes 1 and 2 are less effective in distinguishing between Types 1 and 2, whereas Attribute 3 shows a clear separation between the types. This finding exposes a limitation in the current approach. While our algorithm is designed to effectively handle BPA conflicts, it assumes that noisy BPAs are in the minority. In the case of the Skin dataset, Attributes 1 and 2 are not assigned lower weights, even though they are less effective in distinguishing the two types. Consequently, when all three attributes are combined, the decision-making process is disproportionately influenced by the less classification-effective attributes, diminishing the impact of Attribute 3's BPA, thus leading to lower recognition rates. This analysis underscores the need to recognize the unique characteristics of individual attributes in the fusion process. Specifically, it emphasizes the importance of assessing the contribution of each attribute to classification, taking into account how effectively they differentiate between types. By adjusting the weighting strategy to reflect these insights, $\mathcal{K}$O-FKL-MSDF algorithm's performance could be further fine-tuned. Such refinement could lead to enhanced robustness and effectiveness in handling complex multi-source data fusion tasks, ensuring that the algorithm leverages the most informative attributes for improved recognition rates.

In summary, the proposed $\mathcal{K}$O-FKL-MSDF algorithm exhibited outstanding performance in practical classification tasks on various real-world datasets. Its superiority is statistically validated using Welch's t-test. Despite some challenges, the algorithm showed promising scalability and adaptability, highlighting its potential for complex real-world pattern classification applications.

7. Conclusion

In conclusion, this study proposes a novel divergence measure, the $\mathcal{K}$ order fractal-based Kullback–Leibler ($\mathcal{K}$O-FKL) divergence, which incorporates a high order fractal preprocessing method into the traditional KL divergence. The primary contribution of this study is the determination of the optimal fractal epoch by studying the convergence of the Deng entropy difference between two BPA systems. In Section 3, the proposed $\mathcal{K}$O-FKL divergence model is thoroughly investigated through several numerical examples to demonstrate its desirable properties. Based on $\mathcal{K}$O-FKL divergence, the $\mathcal{K}$O-FKL-MSDF algorithm is developed for multi-source data fusion. In Section 5, the superiority and robustness of the model are validated through experiments and a sensitivity analysis. Finally, in Section 6, the proposed $\mathcal{K}$O-FKL-MSDF algorithm is tested on five datasets with different complexities to demonstrate its recognition accuracy. The results are compared with those of classical algorithms and divergence-based fusion algorithms and are verified with improved recognition rates. Moreover, the proposed model is compared and analyzed with the FBD_{SKL} divergence model in both Sections 5 and 6, demonstrating the necessity of finding higher-order $\mathcal{K}$ fractals. In summary, this study provides a new approach for measuring the conflict between BPAs and contributes to multi-source data fusion applications.

Despite the demonstrated superiority in terms of accuracy and robustness of the $\mathcal{K}$O-FKL-MSDF algorithm, there are still areas for improvement: As discussed in Section 4.2, the time and space complexities of the $\mathcal{K}$O-FKL-MSDF algorithm can reach $\mathcal{O}(N^2)\times\mathcal{O}(\mathcal{K}\times 2^n)$ and $\mathcal{O}(N^2)\times\mathcal{O}(2^n)$, respectively. The exponential nature of these complexities is a critical concern, as it may lead to considerable computational costs when dealing with a large number of BPAs.

In view of the limitations of the proposed $\mathcal{K}$O-FKL-MSDF algorithm, future work should primarily focus on enhancing its efficiency. First, with respect to time complexity, constructing the CMM is the most time-consuming aspect of the algorithm, as each element in the CMM requires divergence computation. To improve calculation efficiency, future research could investigate the potential of combining the proposed $\mathcal{K}$O-FKL measure with quantum computing for parallel processing.

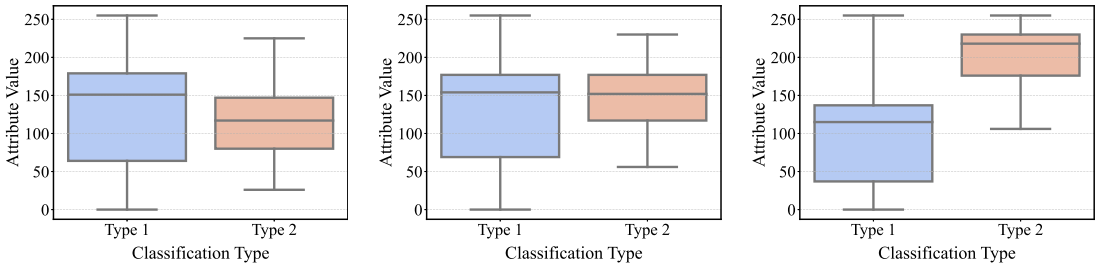


Fig. 17. Comparison of Attributes between Types 1 and 2 in three attributes.

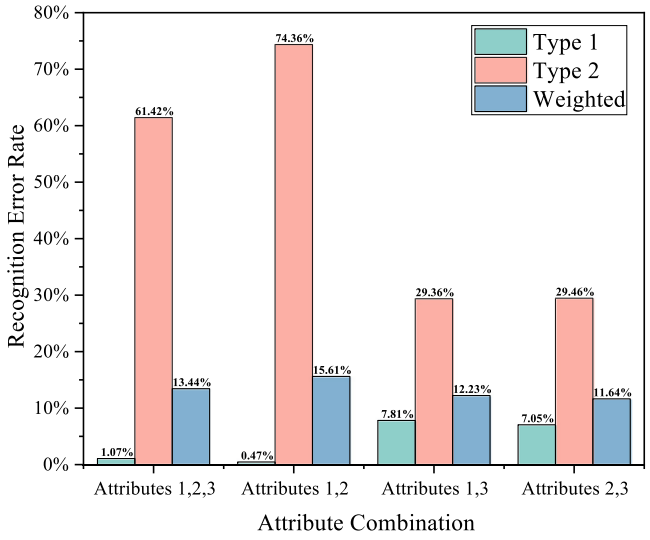


Fig. 18. Comparative analysis of recognition error rates for different attribute combinations in the Skin dataset.

Furthermore, with regards to space complexity, the redundancy in the matrix structures of CMM and SMM due to the duplication of original elements should be addressed. An optimized data structure could be proposed to address this issue by storing divergence values more efficiently. In addition, as indicated in Table 5, the time and space complexities of existing belief divergence algorithms, including BJS, $\mathfrak{B}$ and FBD_{SKL} divergence, all reach $\mathcal{O}(N^2)\times\mathcal{O}(2^n)$. Hence, there is room for optimization in the time complexity of these divergence measures. However, more importantly, these algorithms are primarily applicable to the real number domain and may not be suitable for more complex applications, particularly those involving complex numbers. In future work, it is a promising orientation to extend the proposed $\mathcal{K}$O-FKL measure to the complex domain, making it a generalized method that can be applied to both real and complex number domains.

CRediT authorship contribution statement

Jie Zeng: Validation, Writing – original draft. **Fuyuan Xiao:** Methodology, Writing – revised draft.

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.

Data availability

Data was used for the research described in the article.

Acknowledgments

The authors greatly appreciate the reviewers' valuable suggestions and the editor's great encouragement. This research is supported by the National Natural Science Foundation of China (No. 62003280), Chongqing Talents: Exceptional Young Talents Project (No. cstc2022ycjh-bgzxm0070), Natural Science Foundation of Chongqing, China (No. CSTB2022NSCQ-MSX0531), Chongqing Overseas Scholars Innovation Program (No. cx2022024), and Fundamental Research Funds for the Central Universities, China (No. 2023CDJXY-035).

References

Anjaria, K. (2022). Dempster-Shafer theory and linguistic intuitionistic fuzzy number-based framework for blending knowledge from knowledge repositories: An approach for knowledge management. *Expert Systems with Applications*, *199*, Article 117142.

Bisht, K., & Kumar, A. (2023). A portfolio construction model based on sector analysis using Dempster-Shafer evidence theory and granger causal network: An application to national stock exchange of India. *Expert Systems with Applications*, *215*, Article 119434.

Cao, Z., Lin, C.-T., Lai, K.-L., Ko, L.-W., King, J.-T., Liao, K.-K., Fuh, J.-L., & Wang, S.-J. (2019). Extraction of SSVEPs-based inherent fuzzy entropy using a wearable headband EEG in migraine patients. *IEEE Transactions on Fuzzy Systems*, *28*(1), 14–27.

Chang, L., Zhang, L., Fu, C., & Chen, Y.-W. (2021). Transparent digital twin for output control using belief rule base. *IEEE Transactions on Cybernetics*, http://dx.doi.org/10.1109/TCYB.2021.3063285.

Che, Y., Deng, Y., & Yuan, Y.-H. (2022). Maximum-entropy-based decision-making trial and evaluation laboratory and its application in emergency management. *Journal of Organizational and End User Computing (JOEUC)*, *34*(7), 1–16.

Chen, L., & Deng, Y. (2023). Entropy of random permutation set. *Communications in Statistics. Theory and Methods*, http://dx.doi.org/10.1080/03610926.2023.2173975.

Chen, L., Deng, Y., & Cheong, K. H. (2021). Probability transformation of mass function: A weighted network method based on the ordered visibility graph. *Engineering Applications of Artificial Intelligence*, *105*, Article 104438.

Chen, L., Deng, Y., & Cheong, K. H. (2023a). The distance of random permutation set. *Information Sciences*, *628*, 226–239.

Chen, L., Deng, Y., & Cheong, K. H. (2023b). Permutation Jensen-Shannon divergence for random permutation set. *Engineering Applications of Artificial Intelligence*, *119*.

Chu, C., Li, Y., Liu, J., Hu, S., Li, X., & Wang, Z. (2022). A formal model for multiagent Q-learning dynamics on regular graphs. In *Proceedings of the thirty-first international joint conference on artificial intelligence, IJCAI* (pp. 194–200).

Dempster, A. P. (1967). Upper and lower probabilities induced by a multivalued mapping. *The Annals of Mathematical Statistics*, 325–339.

Deng, Y. (2020a). Information volume of mass function. *International Journal of Computers Communications & Control*, *15*(6), 3983.

Deng, Y. (2020b). Uncertainty measure in evidence theory. *Science China Information Sciences*, *63*(11), Article 210201.

Deng, Y. (2022). Random permutation set. *International Journal of Computers Communications & Control*, *17*(1), 4542.

Deng, X., & Cui, Y. (2021). An improved belief structure satisfaction to uncertain target values by considering the overlapping degree between events. *Information Sciences*, *580*, 398–407.

Deng, Y., Shi, W., Zhu, Z., & Liu, Q. (2004). Combining belief functions based on distance of evidence. *Decision Support Systems*, *38*(3), 489–493.

Fan, W., & Xiao, F. (2022). A complex Jensen-Shannon divergence in complex evidence theory with its application in multi-source information fusion. *Engineering Applications of Artificial Intelligence*, http://dx.doi.org/10.1016/j.engappai.2022.105362.

Fang, R., Liao, H., & Mardani, A. (2022). How to aggregate uncertain and incomplete cognitive evaluation information in lung cancer treatment plan selection? A method based on Dempster-Shafer theory. *Information Sciences*, *603*, 222–243.

Fujita, H., Gaeta, A., Loia, V., & Orciuoli, F. (2020). Hypotheses analysis and assessment in counter-terrorism activities: a method based on OWA and fuzzy probabilistic rough sets. *IEEE Transactions on Fuzzy Systems*, *28*, 831–845.

Garg, H., & Rani, D. (2022). Novel distance measures for intuitionistic fuzzy sets based on various triangle centers of isosceles triangular fuzzy numbers and their applications. *Expert Systems with Applications*, *191*, Article 116228.

Hua, Z., Fei, L., & Xue, H. (2022). Consensus reaching with dynamic expert credibility under Dempster-Shafer theory. *Information Sciences*, *610*, 847–867.

Huang, J., Song, X., Xiao, F., Cao, Z., & Lin, C.-T. (2023). Belief f-divergence for EEG complexity evaluation. *Information Sciences*, *643*, Article 119189.

Ko, S., & Koo, D. (2023). A novel approach for wafer defect pattern classification based on topological data analysis. *Expert Systems with Applications*, *231*, Article 120765.

Kurban, T. (2022). Region based multi-spectral fusion method for remote sensing images using differential search algorithm and IHS transform. *Expert Systems with Applications*, *189*, Article 116135.

Li, Y.-F., Huang, H.-Z., Mi, J., Peng, W., & Han, X. (2022). Reliability analysis of multi-state systems with common cause failures based on Bayesian network and fuzzy probability. *Annals of Operations Research*, *311*, 195–209.

Li, Y., Pelusi, D., Cheong, K. H., & Deng, Y. (2022). The arithmetics of two dimensional belief functions. *Applied Intelligence*, *52*(4), 4192–4210.

Liu, Z.-g., Fu, Y.-m., Pan, Q., & Zhang, Z.-w. (2022). Orientational distribution learning with hierarchical spatial attention for open set recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(7), 8757–8772.

Liu, P., Li, Y., & Wang, P. (2022). Consistency threshold-and score function-based multi-attribute decision-making with Q-rung orthopair fuzzy preference relations. *Information Sciences*, *618*, 356–378.

Liu, Z.-G., Qiu, G.-H., Wang, S.-Y., Li, T.-C., & Pan, Q. (2021). A new belief-based bidirectional transfer classification method. *IEEE Transactions on Cybernetics*, *52*(8), 8101–8113.

Meng, D., Wang, H., Yang, S., Lv, Z., Hu, Z., & Wang, Z. (2022). Fault analysis of wind power rolling bearing based on EMD feature extraction. *CMES-Computer Modeling in Engineering & Sciences*, *130*(1), 543–558.

Meng, D., Yang, S., He, C., Wang, H., Lv, Z., Guo, Y., & Nie, P. (2022). Multidisciplinary design optimization of engineering systems under uncertainty: a review. *International Journal of Structural Integrity*, *13*(4), 565–593.

Miao, W., Geng, J., & Jiang, W. (2023). Multi-granularity decoupling network with pseudo-label selection for remote sensing image scene classification. *IEEE Transactions on Geoscience and Remote Sensing*, http://dx.doi.org/10.1109/TGRS.2023.3244565.

Qiang, C., Deng, Y., & Cheong, K. H. (2022). Information fractal dimension of mass function. *Fractals*, *30*, Article 2250110.

Ren, M., He, P., & Zhou, J. (2022). Decision fusion of two sensors object classification based on the evidential reasoning rule. *Expert Systems with Applications*, *210*, Article 118620.

Shafer, G. (1976). *A mathematical theory of evidence, Vol. 42* (pp. 237–238). Princeton University Press.

Shannon, C. E. (2001). A mathematical theory of communication. *SIGMOBILE Mobile Computing and Communications Review*, *5*(1), 3–55.

Smets, P. (2005). Decision making in the TBM: the necessity of the pignistic transformation. *International Journal of Approximate Reasoning*, *38*(2), 133–147.

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. *Applied Soft Computing*, *84*, Article 105703.

Verma, R., & Álvarez-Miranda, E. (2023). Group decision-making method based on advanced aggregation operators with entropy and divergence measures under 2-tuple linguistic Pythagorean fuzzy environment. *Expert Systems with Applications*, *231*, Article 120584.

Wang, Z., Hou, D., Gao, C., Huang, J., & Xuan, Q. (2022). A rapid source localization method in the early stage of large-scale network propagation. In *Proceedings of the ACM web conference (WWW-22)* (p. 1372).

Wang, T., Liu, R., & Qi, G. (2022). Multi-classification assessment of bank personal credit risk based on multi-source information fusion. *Expert Systems with Applications*, *191*, Article 116236.

Wang, Z., Mu, C., Hu, S., Chu, C., & Li, X. (2022). Modelling the dynamics of regret minimization in large agent populations: a master equation approach. In *Proceedings of the 31st international joint conference on artificial intelligence (IJCAI-22)* (pp. 534–540).

Wang, Z., Song, Z., Shen, C., & Hu, S. (2023). Emergence of punishment in social dilemma with environmental feedback anonymous submission. In *Proceedings of the 37th AAAI conference on artificial intelligence (AAAI-23)*. Accepted.

Wang, J., Zhou, Z., Hu, C., Tang, S., He, W., & Long, T. (2022). A fusion approach based on evidential reasoning rule considering the reliability of digital quantities. *Information Sciences*, *612*, 107–131.

Welch, B. L. (1947). The generalization of "student's" problem when several different population variances are involved. *Biometrika*, *34*(1–2), 28–35.

Xiao, F. (2019). Multi-sensor data fusion based on the belief divergence measure of evidences and the belief entropy. *Information Fusion*, *46*, 23–32.

Xiao, F. (2020). A new divergence measure for belief functions in D–S evidence theory for multisensor data fusion. *Information Sciences*, *514*, 462–483.

Xiao, F. (2022a). GEJS: A generalized evidential divergence measure for multisource information fusion. *IEEE Transactions on Systems, Man, and Cybernetics - Systems*, *53*(4), 2246–2258.

Xiao, F. (2022b). Generalized quantum evidence theory. *Applied Intelligence*, *53*(11), 14329–14344.

Xiao, F. (2023). Quantum X-entropy in generalized quantum evidence theory. *Information Sciences*, *643*, 119177.

Xiao, F., Cao, Z., & Lin, C.-T. (2022). A complex weighted discounting multisource information fusion with its application in pattern classification. *IEEE Transactions on Knowledge and Data Engineering*, *53*(4), 2246–2258.

Xiao, F., & Pedrycz, W. (2022). Negation of the quantum mass function for multisource quantum information fusion with its application to pattern classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(2), 2054–2070.

Xiao, F., Wen, J., & Pedrycz, W. (2022). Generalized divergence-based decision making method with an application to pattern classification. *IEEE Transactions on Knowledge and Data Engineering*, *35*(7), 6941–6956.

Xiong, L., Su, X., & Qian, H. (2021). Conflicting evidence combination from the perspective of networks. *Information Sciences*, *580*, 408–418.

Xu, X., Zhang, D., Bai, Y., Chang, L., & Li, J. (2020). Evidence reasoning rule-based classifier with uncertainty quantification. *Information Sciences*, *516*, 192–204.

Yager, R. R., Alajlan, N., & Bazi, Y. (2019). Uncertain database retrieval with measure-based belief function attribute values. *Information Sciences*, *501*, 761–770.

Yang, C., & Xiao, F. (2022). An exponential negation of complex basic belief assignment in complex evidence theory. *Information Sciences*, http://dx.doi.org/10.1016/j.ins.2022.11.160.

Ye, J., Zhan, J., Ding, W., & Fujita, H. (2021). A novel fuzzy rough set model with fuzzy neighborhood operators. *Information Sciences*, *544*, 266–297.

Zeng, J., & Xiao, F. (2023). A fractal belief KL divergence for decision fusion. *Engineering Applications of Artificial Intelligence*, *121*, Article 106027.

Zhou, Q., & Deng, Y. (2022). Fractal-based belief entropy. *Information Sciences*, *587*, 265–282.

Zhou, M., Zheng, Y.-Q., Chen, Y.-W., Cheng, B.-Y., Herrera-Viedma, E., & Wu, J. (2023). A large-scale group consensus reaching approach considering self-confidence with two-tuple linguistic trust/distrust relationship and its application in life cycle sustainability assessment. *Information Fusion*, *94*, 181–199.

Zhou, Y.-J., Zhou, M., Liu, X.-B., Cheng, B.-Y., & Herrera-Viedma, E. (2022). Consensus reaching mechanism with parallel dynamic feedback strategy for large-scale group decision making under social network analysis. *Computers & Industrial Engineering*, *174*, Article 108818.

Zhu, R., Liu, Q., Huang, C., & Kang, B. (2022). Z-ACM: An approximate calculation method of Z-numbers for large data sets based on kernel density estimation and its application in decision-making. *Information Sciences*, *610*, 440–471.

Zhu, C., Xiao, F., & Cao, Z. (2022). A generalized Rényi divergence for multi-source information fusion with its application in EEG data analysis. *Information Sciences*, http://dx.doi.org/10.1016/j.ins.2022.05.012.