A new divergence measure for basic probability assignment and its applications in extremely uncertain environments

Liguo Fei, Yong Hu, Yong Deng and Sankaran Mahadevan

Abstract—**Information fusion under extremely uncertain environments is an important issue in pattern classification and decision-making problem. Dempster-Shafer evidence theory (D-S theory) is more and more extensively applied to information fusion for its advantage to deal with uncertain information. However, the results opposite to common sense are often obtained when combining the different evidences using Dempster's combination rules. How to measure the difference between different evidences is still an open issue. In this paper, a new divergence is proposed based on Kullback-Leibler divergence in order to measure the difference between different basic probability assignments (BPAs). Numerical examples are used to illustrate the computational process of the proposed divergence. Then the similarity for different BPAs is also defined based on the proposed divergence. The basic knowledge about pattern recognition is introduced and a new classification algorithm is presented using the proposed divergence and similarity under extremely uncertain environments, which is illustrated by a small example handling robot sensing. The method put forward is motivated by desperately in need to develop intelligent systems, such as sensor-based data fusion manipulators, which need to work in complicated, extremely uncertain environments. Sensory data satisfy the conditions 1) fragmentary and 2) collected from multiple levels of resolution.**

Index Terms—**Dempster-Shafer evidence theory, Basic probability assignment, Divergence, Pattern classification, Extremely uncertain environments, Uncertainty management.**

I. Introduction

The information fusion problems are investigated for extremely uncertain environments, such as deep-sea exploration [1] and robot sensing [2], [3]. There exist many key problems need to be handled in pattern classification [4], [5] and decision making [6] under extremely uncertain environments. Dempster-Shafer evidence theory (D-S theory) [7], [8] has been widely focused and adopted in recent years with its great advantage to handle and analyze uncertain information. This theory is widely used in object classification [9], decision making [10], [11], [12], [13], risk assessment [14], [15] and information fusion [16], [17]. In these applications, Dempster's combination rule is very important to combine multiple sources for an aggregated result. However, the rule has been criticized by many researchers [18], [19] because the counter-intuitive conditions often occur when fusing the high conflicting evidences using this rules [19], [20]. And this kind of results have a serious influence for the accuracy of evidence fusion.

It is so necessary and significant for researchers to remedy this weakness of Dempster's [7] combination rules. To improve this shortcoming, a series of alternative combination rules are presented [17], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30], [31], [32] currently. Generally speaking, there exist two categories of methods to deal with this problem. One is to improve the Dempster's combination rules and to reallocate the conflict. For example, in [23], [33], *Lefevre* used the part of the conflicting evidence and distributed the conflict into the focal element sets of all the evidence proportionally. In [26], the conflict of evidence is abandoned to utilize because *Yager* believe it is useless and distribute them into the universal set. However, sometimes it enlarges the uncertainty of evidence and gets the unreasonable fusion results. And the other one is to modify the conflicting evidences before the fusion. *Schubert* [17] and *Han* [30] proposed the modified algorithms to obtain the weights of evidence. In [24], *Deng* proposed a method about the evidence support based on the Jousselme distance function and determine a weighted average of all the evidence. In [28], *Murphy* presented a problem, the failure to balance multiple evidence, then illustrated the proposed solutions and described their limitations.

All of these methods can improve the fusion results in part and make up some weakness of the combination rule from different perspectives. However, some essence is ignored to figure out this problem for a long time. To resolve the problem in essence, in this paper, a new divergence measure is proposed for difference between BPAs based on the modified Kullback-Leibler divergence. Some numerical examples are given to illustrate the computational process.

Similarity measures are very useful in clustering and

L. Fei is with the School of Computer and Information Science, Southwest University, Chongqing, 400715, China; Y. Hu is with the Big Data Decision Institute, Jinan University, Tianhe, Guangzhou, 510632, China; Y. Deng is with the the School of Computer and Information Science, Southwest University, Chongqing, 400715, China, and also with Institute of Fundamental and Frontier Science, University of Electronic Science and Technology of China, Chengdu, 610054, China; S. Mahadevan is with the School of Engineering, Vanderbilt University, Nashville, TN, 37235, USA. e-mail: (ydeng@swu.edu.cn, prof.deng@hotmail.com.)

classification. However, classical these measures are not fully reasonable. So, in this paper, a new similarity is proposed based on the presented divergence and with beautiful properties.

Pattern recognition is a branch of machine learning that focuses on the recognition of patterns and classifications in data [34], [35]. Recently, pattern recognition has been widely applied in different areas: character recognition, speech recognition, fingerprint recognition, remote sensing, medical diagnosis and especially extremely uncertain environments. However, the inevitable problems in pattern recognition become increasingly prominent, and the signification of uncertainty has been aware in recent years for the securable information from sensors of the intelligent robot is collected at multiple levels of resolution. It is generally fuzzy, inaccurate and incomplete from extremely uncertain environment. Although a lot of algorithms have been developed for pattern recognition, such as K-Nearest Neighbor(K-NN) [36], Bayes Classifier [37], Principle Component Analysis(PCA) [38] and Linear Discriminant Analysis(LDA) [39]. These mentioned uncertain features of intelligent systems motivate us to explore the evidential classification approaches for sensor-based information fusion. In this paper, a new algorithm is proposed for deducing overall information derived from local information in the field of pattern recognition based on the proposed divergence and similarity. The purpose of this section is to design an algorithmic framework for the fusion of multi-aspect information in the intelligent systems in complicated, irregularly and extremely uncertain environments. An application example is conducted about classification and the results demonstrated the effectiveness of our algorithm.

The remainder of this paper is constituted as follows. Section II introduces the D-S theory and its basic rules and some necessary related concepts about Kullback-Leibler Divergence. The proposed definition of divergence and similarity are presented in Section III. Section IV proposes an algorithm applying in pattern recognition and conducts an example in a correlative application. Conclusion is given in Section V.

II. Preliminaries

A. Dempster-Shafer evidence theory

Dempster-Shafer evidence theory (D-S theory) is proposed by Dempster [7] and developed later by Shafer[8]. This theory extends the elementary event space in probability theory to its power set named as frame of discernment and constructs the basic probability assignment(BPA) on it. In addition, there is a combination rule presented by Dempster to fuse different BPAs. In particular, D-S theory can definitely degenerate to the probability theory if the belief is only assigned to single elements. The basic definitions about D-S theory are shown as follows:

1) Frame of discernment: D-S theory supposes the definition of a set of elementary hypotheses called the frame of discernment, defined as:

$$\Theta = \{H_1, H_2, ..., H_N\} \tag{1}$$

Where Θ is a set of mutually exclusive and collectively exhaustive events. Let us denote 2^{Θ} the power set of Θ.

2) Mass function: A mass function is also called a BPA and defined as follows.

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

which satisfies the following conditions: $m(\phi) = 0$, $\sum_{A\in 2^{\Theta}} m(A) = 1$.

3) Dempster's combination rule: Dempster [7] proposed orthogonal sum to combine these BPAs for a final result. Suppose m_1 and m_2 are two mass functions. The Dempster's rule of combination denoted by $m = m_1 \bigoplus m_2$ is defined as follows:

$$m(A) = \frac{\sum_{B\bigcap C=A} m_1(B)m_2(C)}{1-K} \tag{3}$$

with

$$K = \sum_{B\bigcap C=\phi} m_1(B)m_2(C) \tag{4}$$

The K is also considered as the conflict coefficient to measure the conflict degree of different BPAs.

B. Pignistic probability distance

In 2006, Liu [19] presented the pignistic probability distance to measure the conflict between evidences.

Suppose m is a BPA, the pignistic probability function [40] is defined as:

$$BetP_m(A) = \sum_{B\subseteq\Theta} \frac{|A \cap B|}{|B|} \frac{m(B)}{1-m(\phi)}, \forall A \subseteq \Theta \tag{5}$$

where $|A|$ is the cardinality of A. The pignistic probability distance is defined as follows.

$$difBetP_{m_1}^{m_2} = max_{A\subseteq\Theta}(|BetP_{m_1}(A) - BetP_{m_2}(A)|) \tag{6}$$

C. Kullback-Leibler Divergence

The Kullback-Leibler divergence [41], [42](KL divergence) proposed by Kullback and Leibler is a non-symmetric measure of the difference between two probability distributions P and Q, is also called information divergence, information gain or relative entropy in probability theory and information theory. Usually, P denotes the "real" distribution of data or a precisely calculated theoretical distribution. The probability distribution Q generally represents a theory, model, description, or approximation of P. And its definition is shown as follows:

$$I(P||Q) = \sum_{x\in X} P(x) \ln \frac{P(x)}{Q(x)} \tag{7}$$

where P and Q are the probability distributions of the discrete random variable X and usually have the same type.

In theory, it is the average of the logarithmic difference between the probabilities P and Q, where the average is taken using the probabilities P. The KL divergence is only defined if P and Q both sum to 1 and if $Q(x) > 0$ for any x such that $P(x) > 0$. If the quantity $0 \times ln(0)$ appears in the formula, it is interpreted as zero.

KL divergence has the following properties that KL divergence is always non-negative, namely $I(P||Q) \geq 0$, a result known as Gibbs' [43] inequality, with $I(P||Q) = 0$ if and only if $P = Q$.

III. The proposed divergence and similairty for BPAs

A. The proposed correlative definition for BPAs

As a kind of uncertainty processing method, D-S theory has been widely applied to multi-source information combination. The combination rule plays an important role in dealing with uncertain information, and the counter-intuitive fusion results will be obtained if there exist any errors in combination. As is known that the unreasonable results typically derived from the conflict among different original evidences. There appear so many methods to measure the conflict between evidences.

It is obvious that the conflict coefficient K and the pignistic probability distance $difBetP$ provide a good methodology to determine whether two BPAs are in conflict. However, they can not discriminate two BPAs which are different but coordinate. To be precise, the *zero* value can be obtained of K and $difBetP$ for different BPAs.

Example III.1. *Let us suppose a frame of discernment $X = \{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5\}$, two mass functions are defined as: $m_1(\theta_1) = m_1(\theta_2) = m_1(\theta_3) = m_1(\theta_4) = m_1(\theta_5) = 0.2$; $m_2\{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5\} = 1$.*

It can be seen that the two BPAs are different, but the calculation results are $K = 0$ based on Eq. (4) and $difBetP = 0$ using Eqs. (6) and (5) .

From Example III.1, a conclusion can be drawn that the conflict coefficient K and the pignistic probability distance $difBetP$ could not always be correctly to reflect the difference between BPAs. In many real cases, the mass function m_1 denotes that the system supports each proposition with the same probability 0.2, while m_2 represents the system could not provide any information for all the proposition. That is to say m_1 provides much more information in the specific application than m_2. Through above analysis, a kind of divergence measure needs to be proposed to distinguish different BPAs. It is obvious that the KL divergence is undefined if $Q(x) = 0$ and $P(x) \neq 0$ for any $x \in X$. To break through this shortcoming and apply it into BPAs, a new divergence is proposed and defined as follows.

$$E(m_1||m_2) = \sum_i m_1(F_i) \ln \frac{m_1(F_i)}{\frac{1}{2}m_1(F_i) + \frac{1}{2}m_2(F_i)} \tag{8}$$

where F_i is a proposition in mass function m_1 and m_2, respectively. The new divergence is similar with the KL divergence in form, but it uses mass functions instead of probability distribution functions. Specially, the BPA will turn into probability if it's only assigned to single elements, and the new divergence will also degenerate to modified KL divergence at the same time. It is noteworthy that the Eq. (8) is the divergence degree of mass function m_1 from mass function m_2, and $E(m_1||m_2)$ is not symmetric on account of its properties. So, we propose a symmetric divergence information measure based on $E(m_1||m_2)$, shown as.

$$div(m_1||m_2) = \frac{1}{2}[E(m_1||m_2) + E(m_2||m_1)] \tag{9}$$

There are several properties of the proposed definition. The related theorems and proofs are shown as follows.

Theorem III.1. ***Non-Negativeness:***
$div(m_1||m_2) \geq 0$ and $div(m_1||m_2) = 0$ iff m_1 is same as m_2.

Proof. Based on known knowledge $\ln(x) \leq x - 1$, for Eq. (8), we have:

$$\begin{aligned}
-E(m_1||m_2) &= -\sum_i m_1(F_i) \ln \frac{m_1(F_i)}{\frac{1}{2}m_1(F_i) + \frac{1}{2}m_2(F_i)} \\
&= \sum_i m_1(F_i) \ln \frac{\frac{1}{2}m_1(F_i) + \frac{1}{2}m_2(F_i)}{m_1(F_i)} \\
&\leq \sum_i m_1(F_i) [\frac{\frac{1}{2}m_1(F_i) + \frac{1}{2}m_2(F_i)}{m_1(F_i)} - 1] \\
&= \sum_i m_1(F_i) [\frac{\frac{1}{2}m_2(F_i) - \frac{1}{2}m_1(F_i)}{m_1(F_i)}] \\
&= \sum_i [\frac{1}{2}m_2(F_i) - \frac{1}{2}m_1(F_i)] \\
&= \frac{1}{2}\sum_i m_2(F_i) - \frac{1}{2}\sum_i m_1(F_i) = 0
\end{aligned}$$

So, $E(m_1||m_2) \geq 0$, and based on Eq. (9), we have $div(m_1||m_2) \geq 0$. It is obvious that $div(m_1||m_2) = 0$ if and only if m_1 is same as m_2.

□

Theorem III.2. ***Orthogonality:***
$div(m_1||m_2) \leq 1$ and $div(m_1||m_2) = 1$ iff m_1 and m_2 are orthogonal.

Proof. **(1):**

1) $div(m_1||m_2) \leq 1$

Based on Eq. (8), we have

$$
\begin{aligned}
E(m_1||m_2) &= \sum_i m_1(F_i)\ln\frac{m_1(F_i)}{\frac{1}{2}m_1(F_i)+\frac{1}{2}m_2(F_i)}\\
&= \sum_i m_1(F_i)\ln\frac{m_1(F_i)}{m_1(F_i)+m_2(F_i)}+\sum_i m_1(F_i)\\
&= \sum_i m_1(F_i)\ln\frac{m_1(F_i)}{m_1(F_i)+m_2(F_i)}+1
\end{aligned}
$$

Since

$$m_1(F_i) \le m_1(F_i)+m_2(F_i)$$

it follows that

$$
\begin{aligned}
&\Rightarrow \frac{m_1(F_i)}{m_1(F_i)+m_2(F_i)} \le 1 \Rightarrow \ln\frac{m_1(F_i)}{m_1(F_i)+m_2(F_i)} \le 0\\
&\Rightarrow \sum_i m_1(F_i)\ln\frac{m_1(F_i)}{m_1(F_i)+m_2(F_i)} \le 0 \Rightarrow E(m_1||m_2) \le 1
\end{aligned}
$$

According to Eq. (9), it is obvious $div(m_1||m_2) \le 1$. **(2):**
1) $div(m_1||m_2) = 1$ iff m_1 and m_2 are orthogonal

When m_1 and m_2 are orthogonal, it should satisfy that:
1) $F_i \bigcap F_j = \phi$, where F_i and F_j are propositions in mass function m.
2) $m_1(F_i) = m_2(F_j) = 1$.

In this case it is easy to get the conclusion that: $E(m_1||m_2) = 1$ and $E(m_2||m_1) = 1$, so $div(m_1||m_2) = 1$. □

Theorem III.3. ***Symmetry:*** *$div(m_1||m_2) = div(m_2||m_1)$.*

Proof. According to Eq. (9), we have

$$
\left.\begin{aligned}
div(m_1||m_2) = \tfrac{1}{2}[E(m_1||m_2)+E(m_2||m_1)]\\
div(m_2||m_1) = \tfrac{1}{2}[E(m_2||m_1)+E(m_1||m_2)]
\end{aligned}\right\}
\Rightarrow div(m_1||m_2) = div(m_2||m_1)
$$

□

Theorem III.4. ***Reflexivity:*** *if $E(m_1||m_2) = 0$ or $E(m_2||m_1) = 0$, then $m_1 = m_2$.*

Proof.

$$
\begin{aligned}
E(m_1||m_2) = 0 &\Rightarrow \sum_i m_1(F_i)\ln\frac{m_1(F_i)}{\frac{1}{2}m_1(F_i)+\frac{1}{2}m_2(F_i)} = 0\\
&\Rightarrow \frac{m_1(F_i)}{\frac{1}{2}m_1(F_i)+\frac{1}{2}m_2(F_i)} = 1\\
&\Rightarrow m_1(F_i) = \frac{1}{2}m_1(F_i)+\frac{1}{2}m_2(F_i)\\
&\Rightarrow m_1(F_i) = m_2(F_i)
\end{aligned}
$$

corollaries, if $E(m_2||m_1) = 0$, then $m_2 = m_1$. □

By the definition of the proposed divergence and its related properties, it is obvious that the divergence is also a kind of distance, which is used to measure the difference between BPAs. Similarity measures is the other side of divergence measure. Namely, we can define our similarity measure based on divergence measures.

Definition III.1.

$$sim(m_1, m_2) = 1 - div(m_1||m_2) \tag{10}$$

By Theorem III.1 and Theorem III.2, we know that $0 \le div(m_1||m_2) \le 1$. Therefore, the similarity also has the property $0 \le sim(m_1||m_2) \le 1$.

B. Numerical examples

Let us go back to the Example III.1. The divergence between m_1 and m_2 is 0.6931 using Eq. (9), instead of 0 value of K and $difBetP$. So, the new divergence can distinguish the two different BPAs.

Next, some numerical examples are given to illustrate the effectiveness of the proposed new divergence.

Example III.2. *Let us suppose a frame of discernment X={θ_1,θ_2,θ_3}, two mass functions in three differen conditions as follows:*

***Case 1**: $m_1(\theta_1)$ = 0.4, $m_1(\theta_2) = m_1(\theta_3)$ = 0.3; $m_2(\theta_1)$ = 0.4, $m_2(\theta_2) = m_2(\theta_3)$ = 0.3*

***Case 2**: $m_1(\theta_1,\theta_2)$ = 0.4, $m_1(\theta_1,\theta_3)$ = 0.6 ; $m_2(\theta_1,\theta_2)$ = 0.4, $m_2(\theta_1,\theta_3)$ = 0.6*

***Case 3**: $m_1(\theta_1,\theta_2,\theta_3)$ = 1 ; $m_2(\theta_1,\theta_2,\theta_3)$ = 1*

the calculation process of divergence and similarity for three cases as follows:

Case 1: $div = \frac{1}{2} \times [(0.4 \times \ln\frac{0.4}{\frac{1}{2}\times 0.4+\frac{1}{2}\times 0.4} + 0.3 \times \ln\frac{0.3}{\frac{1}{2}\times 0.3+\frac{1}{2}\times 0.3} + 0.3 \times \ln\frac{0.3}{\frac{1}{2}\times 0.3+\frac{1}{2}\times 0.3}) + (0.4 \times \ln\frac{0.4}{\frac{1}{2}\times 0.4+\frac{1}{2}\times 0.4} + 0.3 \times \ln\frac{0.3}{\frac{1}{2}\times 0.3+\frac{1}{2}\times 0.3} + 0.3 \times \ln\frac{0.3}{\frac{1}{2}\times 0.3+\frac{1}{2}\times 0.3})] = 0$

$sim = 1 - div$ = 1 - 0 = 1

Case 2: $div = \frac{1}{2} \times [(0.6 \times \ln\frac{0.6}{\frac{1}{2}\times 0.6+\frac{1}{2}\times 0.6} + 0.4 \times \ln\frac{0.4}{\frac{1}{2}\times 0.4+\frac{1}{2}\times 0.4}) + (0.6 \times \ln\frac{0.6}{\frac{1}{2}\times 0.6+\frac{1}{2}\times 0.6} + 0.4 \times \ln\frac{0.4}{\frac{1}{2}\times 0.4+\frac{1}{2}\times 0.4})] = 0$

$sim = 1 - div$ = 1 - 0 = 1

Case 3: $div = \frac{1}{2} \times [(1 \times \ln\frac{1}{\frac{1}{2}\times 1+\frac{1}{2}\times 1}) + (1 \times \ln\frac{1}{\frac{1}{2}\times 1+\frac{1}{2}\times 1})] = 0$

$sim = 1 - div$ = 1 - 0 = 1

From the the three conditions of Example III.2, it can be seen that the divergence is zero for the same BPAs.

Example III.3. *Suppose a frame of discernment X={θ_1,θ_2}, two mass functions:*

$m_1(\theta_1)$ = a, $m_1(\theta_2)$ = 1-a, $a \in [0,1]$; $m_2(\theta_1)$ = 0.5, $m_2(\theta_2)$ = 0.5

the divergence degree of Example III.3 is: $div = \frac{1}{2} \times [(a \times \ln\frac{a}{\frac{1}{2}\times a+\frac{1}{2}\times 0.5} + (1-a) \times \ln\frac{1-a}{\frac{1}{2}\times(1-a)+\frac{1}{2}\times 0.5}) + (0.5 \times \ln\frac{0.5}{\frac{1}{2}\times 0.5+\frac{1}{2}\times a} + 0.5 \times \ln\frac{0.5}{\frac{1}{2}\times 0.5+\frac{1}{2}\times(1-a)})]$.

The divergence and similarity between m_1 and m_2 in Example III.3 is shown in Figure 1 and Figure 2 with the parameter a has changed, respectively. It is obvious that when the value of a changes in the interval [0, 0.5], the difference between m_1 and m_2 getting smaller and smaller, and their divergence decreases correspondingly.

Next, the divergence degree turns into 0 when m_1 and m_2 are exactly the same. It is also equally obvious that when the parameter a changes in the interval [0.5, 1], the difference between m_1 and m_2 is growing, and their divergence increases correspondingly. The similarity is in like manner.

IV. A PATTERN CLASSIFICATION ALGORITHM FRAMEWORK UNDER EXTREMELY UNCERTAIN ENVIRONMENTS

Pattern recognition, is also called pattern classification which is a branch of machine learning that focuses on the recognition of patterns and regularities in data [34]. There exist many algorithms for pattern recognition, such as K-Nearest Neighbor(K-NN) [36], Bayes Classifier [37], Principle Component Analysis(PCA) [38], Linear Discriminant Analysis(LDA) [39], Non-negative Matrix Factorization(NMF) [44]. In this section, an algorithm will be introduced for pattern recognition based on the proposed new divergence and similarity between BPAs, then an example will be conducted which is proposed in [2] to illustrate the effectiveness of the proposed method.

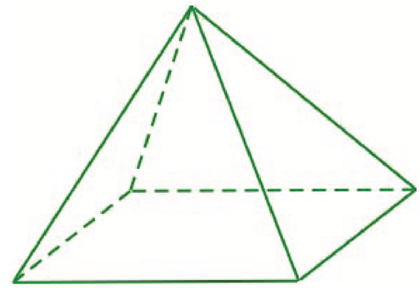
A. Introduction of the background

In this example, the purpose of pattern classification algorithm is to classify 3-D objects from incomplete sensory information. The object which waiting to be identified in this example is the available portion among most objects in a complex environment. There are four kinds of objects which may be extracted from a "coffee cup and a cube of sugar on a saucer" environment in the knowledge base which are shown as Figure 3.

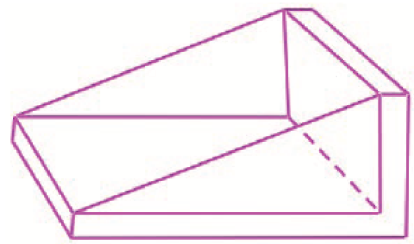
They are Pyramid, L-shape, Handle and Cylinder, and we will call them P, L, H and C in the following for convenience, respectively. Each classification can been seen as an attribute of a mass function in the frame of discernment. So, the knowledge base can be denoted as $FF = \{P, L, H, C\}$. For an object, it can be observed from three aspects: faces, edges and vertices. For example, there are six faces for a cube considering from the point of face. Thus, if focusing on the face only, a triad can be used to represent a 3-D object as $(number, type, curvature)$. Based on the proposed frame of discernment, we make the following provisions:

- The number of faces satisfies: $number \in [1, 8]$.
- The type of faces satisfies: $type \in \{triangle(T), square(S), rectangle(R), bracket(B), circle(C), disk(D)\}$.
- The curvature of faces satisfies: $curvature \in \{planar(P), curved(Q)\}$.

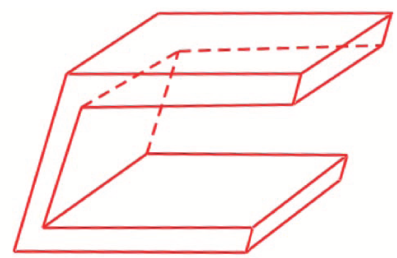
For above mentioned cube, it can be denoted using a triad as $(6, S, P)$. So, the four classifications in FF could include $8 \times 6 \times 2 = 96$ kinds of way to describe all the combination. That is, there should be 96 triads to represent all the possibility in the frame of discernment FF. In the knowledge base, each classification includes its prototype objects, and denoted by mass functions. There are three prototypical objects for Pyramid and L-Shape, two for Handle and one for Cylinder. The mass



(a) *Pyramid*



(b) $L - Shape$



(c) *Handle*

(d) *Cylinder*

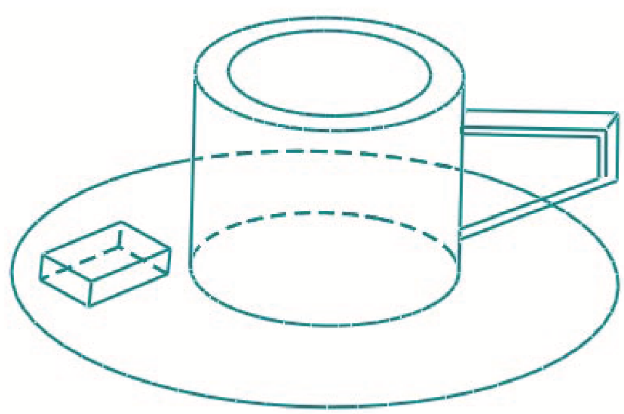


Fig. 3. Graspable objects in complex environment

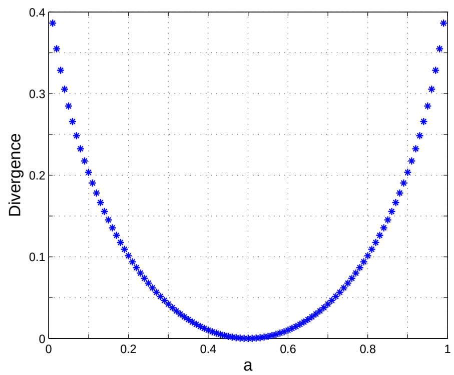


Fig. 1. The divergence with changing parameter a

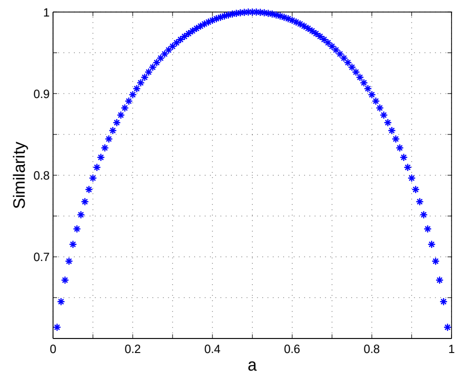


Fig. 2. The similarity with changing parameter a

functions corresponding by prototypical objects of four classifications in FF are shown as Table I.

Each prototypical object is denoted by a list of these attributes weighted based on the degree to which they characterize the object. The weight distributed to FF can identify the incompleteness of the list of the attributes.

In Table I, (4,T,P) represents for four planar triangular faces; (1, S, P) for one planar square face; (1, R, P) for one planar rectangular face; (2, B, P) for two planar bracket faces; (2, C, P) for two planar circle faces; (1, R, C) for one curved rectangular face, etc.

For the proposed pattern classification method, the input is a mass function(BPA) in frame of discernment FF which can reflect the uncertainty derived from the imperfections of sensors. In this example, there are four sensors observe shape primitives at the same time, then provide four mass functions(BPAs) shown as follows.

$$\begin{aligned} Y_1 &= [\{(1,R,P),FF\},\{0.8,0.2\}], \\ Y_2 &= [\{(1,S,P),FF\},\{0.6,0.4\}], \\ Y_3 &= [\{(2,S,P),FF\},\{0.7,0.3\}], \\ Y_4 &= [\{(2,T,P),FF\},\{0.5,0.5\}]. \end{aligned}$$

That is, the first evidence supports that there exists a planar rectangular face in the observed object with the probability of 0.8; The second supports exists a planar square face with the probability of 0.6; The third has the belief that two planar square faces in with the credibility of 0.7; The last one determine two planar triangular faces with half of the trust.

B. The proposed algorithm for pattern classification

Here, we will introduce how to recognize the right classification using the above proposed divergence and similarity.

The matching algorithm for one input evidence with the knowledge base $FF = \{\Gamma_1, ..., \Gamma_n\}$ is listed as follows.

Step 1:

1) Calculating the divergence $div(Y, X_{ij})$ between Y and each element X_{ij} in Γ_i.
2) Calculating the corresponding similarity $sim(Y, X_{ij})$ based on Eq. (10).
3) Selecting the maximal similarity between Y and X_{ij} as the similarity of Y and Γ_i.
$$sim(Y,\Gamma_i) = max_j sim(Y, X_{ij}) \tag{11}$$
4) Assigning basic probability to each classification in the knowledge base. The basic probability for universal set FF is:
$$M(FF) = 1 - max_i sim(Y,\Gamma_i) \tag{12}$$
The remaining n classifications is shown as follows.
$$\begin{aligned} M(\Gamma_i) &= (1 - M(FF)) \cdot \frac{sim(Y,\Gamma_i)}{\sum_i sim(Y,\Gamma_i)} \\ &= max_i sim(Y,\Gamma_i) \cdot \frac{sim(Y,\Gamma_i)}{\sum_i sim(Y,\Gamma_i)} \end{aligned} \tag{13}$$
5) Determining the matching degree between the input evidence Y and each classification based on the obtained mass function(BPA).

Using the above steps, multiple mass functions(BPAs) can be obtained based on corresponding input evidences. The flow chart of the proposed recognition algorithm is shown in Figure 4.

C. Experimental results and analysis

Based on above algorithm and calculation flow, at first, calculating the divergences between all the prototypical objects in knowledge base and the input evidences from Y_1 to Y_4, and the results are shown in Table II.

Obtaining the similarities between all the prototypical objects in knowledge base and the input evidences from Y_1 to Y_4 based on Eq. (10), and the results are shown in Table III.

Selecting the maximum similarity between Y_i and prototypical objects of each classification as the similarity between Y_i and corresponding classification, and the results are shown in Table IV.

Determining four mass functions with respect to all the classifications based on the results in Table IV and Eqs. (12) and (13). The first evidence Y_1 is token as an example,

TABLE I
KNOWLEDGE BASE

Classification	Prototypical Object
Pyramid	$P_1 = [\{(4,T,P),(1,S,P),F\};\{0.5,0.3,0.2\}]$
	$P_2 = [\{(4,T,P),(1,R,P),F\};\{0.5,0.25,0.25\}]$
	$P_3 = [\{(4,T,P),F\};\{0.7,0.3\}]$
L-Shape	$L_1 = [\{(4,T,P),(2,L,P),(5,R,P)(1,S,P),F\};\{0.2,0.4,0.1,0.2,0.1\}]$
	$L_2 = [\{(2,T,P),(1,S,P),(5,R,P),F\};\{0.4,0.2,0.1,0.3\}]$
	$L_3 = [\{(6,R,P),F\};\{0.5,0.5\}]$
Handle	$H_1 = [\{(2,B,P),(6,R,P),(2,S,P),F\};\{0.4,0.2,0.2,0.2\}]$
	$H_2 = [\{(2,C,P),(1,R,Q),F\};\{0.4,0.4,0.2\}]$
Cylinder	$C_1 = [\{(2,C,P),(1,R,Q),F\};\{0.4,0.4,0.2\}]$

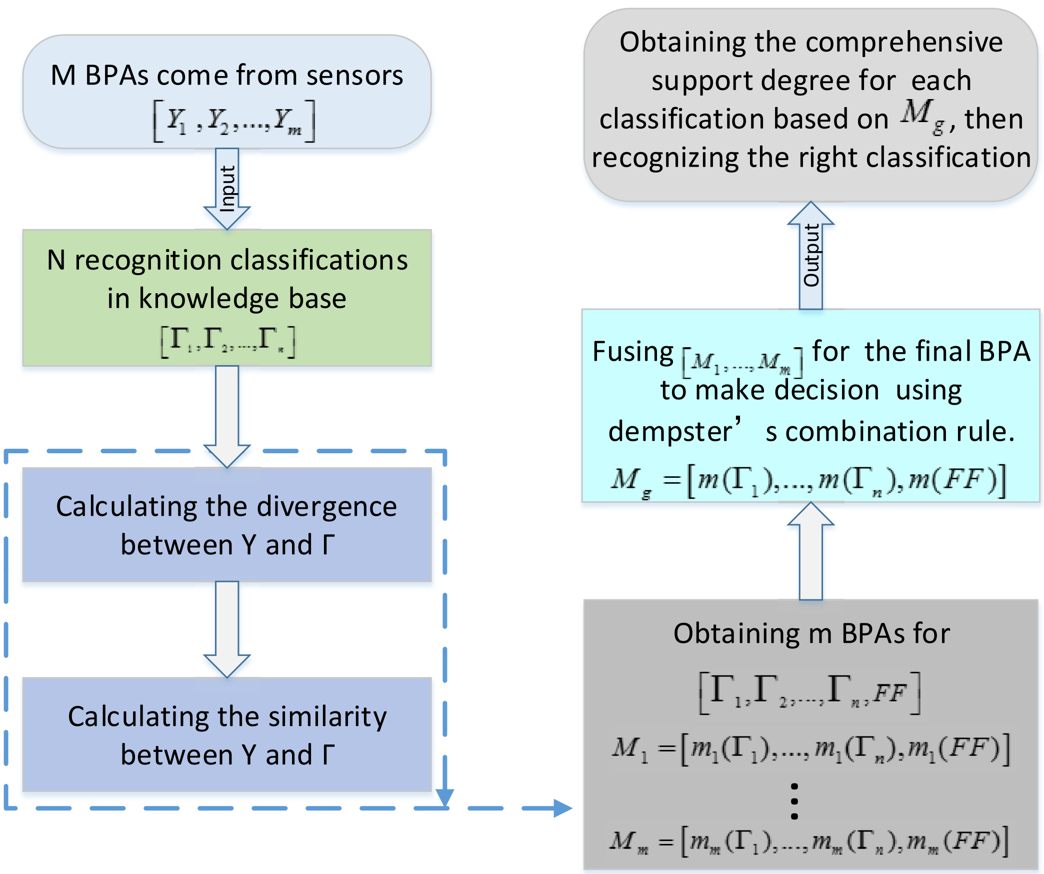


Fig. 4. The flow chart of the proposed algorithm for pattern recognition

TABLE II
DIVERGENCES BETWEEN Y_i AND ALL THE PROTOTYPICAL OBJECTS

	P			L			H		C
	P_1	P_2	P_3	L_1	L_2	L_3	H_1	H_2	C_1
Y_1	0.5545	0.2504	0.5249	0.5977	0.5249	0.4838	0.5545	0.5545	0.5545
Y_2	0.2158	0.4766	0.4541	0.3431	0.2292	0.3840	0.5022	0.5022	0.5022
Y_3	0.5249	0.5037	0.4852	0.5807	0.4852	0.4285	0.2865	0.5249	0.5249
Y_4	0.4838	0.4545	0.4285	0.5580	0.1194	0.3466	0.4838	0.4838	0.4838

TABLE III
SIMILARITIES BETWEEN Y_i AND ALL THE PROTOTYPICAL OBJECTS

	P			L			H		C
	P_1	P_2	P_3	L_1	L_2	L_3	H_1	H_2	C_1
Y_1	0.4455	0.7496	0.4751	0.4023	0.4751	0.5162	0.4455	0.4455	0.4455
Y_2	0.7842	0.5234	0.5459	0.6569	0.7708	0.6160	0.4978	0.4978	0.4978
Y_3	0.4751	0.4963	0.5148	0.4193	0.5148	0.5715	0.7135	0.4751	0.4751
Y_4	0.5162	0.5455	0.5715	0.4420	0.8806	0.6534	0.5162	0.5162	0.5162

TABLE IV
THE SIMILARITY BETWEEN Y_i AND EACH CLASSIFICATION

	P	L	H	C
Y_1	0.7496	0.5162	0.4455	0.4455
Y_2	0.7842	0.7788	0.4978	0.4978
Y_3	0.5148	0.5715	0.7135	0.4751
Y_4	0.5715	0.8806	0.5162	0.5162

$$M_1(FF) = 1 - max_i sim(Y_1, \Gamma_i) = 1 - 0.7496 = 0.2504$$

Next, the probability of $M_1(P)$, $M_1(L)$, $M_1(H)$ and $M_1(C)$ are determined from the remaining 0.7496 based on Eq. (13). The result is shown as follows.

$$\Theta = \{P, L, H, C, PLHC\},$$
$$M_1 = \{0.2605, 0.1795, 0.1548, 0.1548, 0.2504\}.$$

Consistent with the above method, the others mass function can been obtained as follows.

$$M_2 = \{0.2411, 0.2370, 0.1531, 0.1531, 0.2157\},$$
$$M_3 = \{0.1615, 0.1792, 0.2238, 0.1490, 0.2865\},$$
$$M_4 = \{0.2026, 0.3121, 0.1830, 0.1830, 0.1193\}.$$

Obtaining the final mass function for making decision using the Dempster's combination rule(Eq. (4)).

$$\Theta = \{P, L, H, C, PLHC\},$$
$$M_g = \{0.2893, 0.3388, 0.1929, 0.1622, 0.0168\}.$$

Considering the similarity overall between the input evidence Y_1, Y_2, Y_3 and Y_4 and each classification in knowledge base, a conclusion can be drawn from the mass function M_g that the classification $L - Shape$ gets the highest support, and then $Pyramid$, $Handle$ and $Cylinder$, respectively. That is, the decision making based on the proposed algorithm determining the observation target is a L-Shape. What's more, the result is in agreement with [2]. In other word, the evidential input has the largest correlation with class L, and is therefore classified as L-shape.

V. CONCLUSION

Under extremely uncertain environments, information fusion is an effective and significant method to handle such as pattern recognition problems. Dempster-Shafer evidence theory is very important in the field of information fusion and applied widely in many processes because of its powerful features to handle the uncertainty. However, the counter-intuitive results are often obtained if there exist the larger conflict between different evidences. To resolve this serious problem effectively, in this paper, a new divergence is proposed based on Kullback-Leibler divergence to measure the difference between BPAs, then a similarity measure is introduced based on the new proposed divergence. Numerical examples are used to illustrate the efficiency of the new divergence and similarity. A brief introduction about pattern recognition is given and a new algorithm for classification under the extremely uncertain environment is presented based on the proposed divergence and similarity. And the results of an application for classification demonstrate the effectiveness of the presented algorithm. In the future study, the conceptual framework of the presented pattern classification methodology based on the proposed divergence and similarity can be improved further. And in order to verify the effectiveness of the proposed algorithm, more other applications should be conducted by it.

ACKNOWLEDGMENTS

The work is partially supported by National Natural Science Foundation of China (Grant Nos. 61174022, 61573290, 61503237).

REFERENCES

[1] W. H. D., "Deep sea explorations," *The American Naturalist*, vol. 4, no. 12, pp. 744–746, 1871.

[2] A. M. Erkmen and H. E. Stephanou, "Information fractals for evidential pattern classification," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 20, no. 5, pp. 1103–1114, 1990.

[3] J.-R. Ruiz-Sarmiento, C. Galindo, and J. Gonzalez-Jimenez, "Exploiting semantic knowledge for robot object recognition," *Knowledge-Based Systems*, vol. 86, pp. 131–142, 2015.

[4] X. Deng, Q. Liu, Y. Deng, and S. Mahadevan, "An improved method to construct basic probability assignment based on the confusion matrix for classification problem," *Information Sciences*, vol. s 340341, pp. 250–261, 2016.

[5] D. Han, W. Liu, J. Dezert, and Y. Yang, "A novel approach to pre-extracting support vectors based on the theory of belief functions," *Knowledge-Based Systems*, vol. 110, pp. 210–223, 2016.

[6] R. R. Yager and N. Alajlan, "Decision making with ordinal payoffs under dempstershafer type uncertainty," *International Journal of Intelligent Systems*, vol. 28, no. 11, pp. 1039–1053, 2013.

[7] A. P. Dempster, "Upper and lower probabilities induced by a multivalued mapping," *Annals of Mathematics and Statistics*, vol. 38, no. 2, pp. 325–339, 1967.

[8] G. Shafer, *A Mathematical Theory of Evidence*. Princeton: Princeton University Press, 1976.

[9] V.-N. Huynh, T. T. Nguyen, and C. A. Le, "Adaptively entropy-based weighting classifiers in combination using dempster–shafer theory for word sense disambiguation," *Computer Speech & Language*, vol. 24, no. 3, pp. 461–473, 2010.

[10] L. V. Utkin, "A new ranking procedure by incomplete pairwise comparisons using preference subsets," *Intelligent Data Analysis*, vol. 13, no. 2, pp. 229–241, 2009.

[11] X. Su, S. Mahadevan, P. Xu, and Y. Deng, "Dependence assessment in Human Reliability Analysis using evidence theory and AHP," *Risk Analysis*, vol. 35, pp. 1296–1316, 2015.

[12] Y. Deng, "Generalized evidence theory," *Applied Intelligence*, vol. 43, no. 3, pp. 530–543, 2015.

[13] J.-B. Yang and D.-L. Xu, "Evidential reasoning rule for evidence combination," *Artificial Intelligence*, vol. 205, pp. 1–29, 2013.

[14] X. Deng, Y. Hu, F. T. Chan, S. Mahadevan, and Y. Deng, "Parameter estimation based on interval-valued belief structures," *European Journal of Operational Research*, vol. 241, no. 2, pp. 579–582, 2015.

[15] X. Deng, Y. Deng, and F. T. Chan, "An improved operator of combination with adapted conflict," *Annals of Operations Research*, vol. 223, no. 1, pp. 451–459, 2014.

[16] F. Cuzzolin, "A geometric approach to the theory of evidence," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 38, no. 4, pp. 522–534, 2008.

[17] J. Schubert, "Conflict management in Dempster-Shafer theory using the degree of falsity," *International Journal of Approximate Reasoning*, vol. 52, no. 3, pp. 449–460, 2011.

[18] L. A. Zadeh, "A simple view of the dempster-shafer theory of evidence and its implication for the rule of combination," *AI magazine*, vol. 7, no. 2, p. 85, 1986.

[19] W. Liu, "Analyzing the degree of conflict among belief functions," *Artificial Intelligence*, vol. 170, no. 11, pp. 909–924, 2006.

[20] L. A. Zadeh, *A simple view of the Dempster-Shafer theory of evidence*. Berkeley Cognitive Science Program, Institute of Cognitive Studies, University of California, Berkeley, 1984.

[21] J. Ma, "Measuring divergences among mass functions," in *Proceedings on the International Conference on Artificial Intelligence (ICAI)*. The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), 2013, p. 1.

[22] X. Wang, J. Zhu, Y. Song, and L. Lei, "Combination of unreliable evidence sources in intuitionistic fuzzy mcdm framework," *Knowledge-Based Systems*, vol. 97, pp. 24–39, 2016.

[23] E. Lefèvre and Z. Elouedi, "How to preserve the conflict as an alarm in the combination of belief functions?" *Decision Support Systems*, vol. 56, pp. 326–333, 2013.

[24] Y. Deng, W. Shi, Z. Zhu, and Q. Liu, "Combining belief functions based on distance of evidence," *Decision support systems*, vol. 38, no. 3, pp. 489–493, 2004.

[25] Y.-M. Xiong, Z.-P. Yang, and X.-F. Qu, "Novel combination method of conflict evidence based on evidential model modification," *Control and Decision*, vol. 26, no. 6, pp. 883–887, 2011.

[26] R. R. Yager, "On the dempster-shafer framework and new combination rules," *Information sciences*, vol. 41, no. 2, pp. 93–137, 1987.

[27] L. Xinde, W. Fengyu, X. Yefan, and C. Zongxiong, "A method of evidence of combination based on the improved similarity measure," in *Information and Automation (ICIA), 2014 IEEE International Conference on*. IEEE, 2014, pp. 461–466.

[28] C. K. Murphy, "Combining belief functions when evidence conflicts," *Decision support systems*, vol. 29, no. 1, pp. 1–9, 2000.

[29] K. Yamada, "A new combination of evidence based on compromise," *Fuzzy sets and Systems*, vol. 159, no. 13, pp. 1689–1708, 2008.

[30] D. Han, Y. Deng, and C. Han, "Sequential weighted combination for unreliable evidence based on evidence variance," *Decision Support Systems*, vol. 56, pp. 387–393, 2013.

[31] A. Martin, A.-L. Jousselme, and C. Osswald, "Conflict measure for the discounting operation on belief functions," in *Information Fusion, 2008 11th International Conference on*. IEEE, 2008, pp. 1–8.

[32] Y. Song, X. Wang, and H. Zhang, "A distance measure between intuitionistic fuzzy belief functions," *Knowledge-Based Systems*, vol. 86, pp. 288–298, 2015.

[33] E. Lefevre, O. Colot, and P. Vannoorenberghe, "Belief function combination and conflict management," *Information fusion*, vol. 3, no. 2, pp. 149–162, 2002.

[34] Y. Anzai, *Pattern Recognition & Machine Learning*. Elsevier, 2012.

[35] Z.-g. Liu, Q. Pan, J. Dezert, and A. Martin, "Adaptive imputation of missing values for incomplete pattern classification," *Pattern Recognition*, vol. 52, pp. 85–95, 2016.

[36] L. E. Peterson, "K-nearest neighbor," *Scholarpedia*, vol. 4, no. 2, p. 1883, 2009.

[37] P. Langley and S. Sage, "Induction of selective bayesian classifiers," in *Proceedings of the Tenth international conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., 1994, pp. 399–406.

[38] I. Jolliffe, *Principal component analysis*. John Wiley & Sons, Ltd, 2014.

[39] A. J. Izenman, "Linear discriminant analysis," in *Modern Multivariate Statistical Techniques*. Springer, 2008, pp. 237–280.

[40] P. Smets, "Decision making in the tbm: the necessity of the pignistic transformation," *International Journal of Approximate Reasoning*, vol. 38, no. 2, pp. 133–147, 2005.

[41] S. Kullback and R. A. Leibler, "On information and sufficiency," *The annals of mathematical statistics*, vol. 22, no. 1, pp. 79–86, 1951.

[42] S. Kullback, *Information theory and statistics*. Courier Corporation, 1968.

[43] J. W. Gibbs, *A method of geometrical representation of the thermodynamic properties of substances by means of surfaces*. Connecticut Academy of Arts and Sciences, 1873.

[44] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Advances in Neural Information Processing Systems 13*. MIT Press, 2001, pp. 556–562.