Computational and Applied Mathematics (2022) 41:292
https://doi.org/10.1007/s40314-022-02000-3

Earth Mover's divergence of belief function

Peilin Liu[1,2] · **Fuyuan Xiao**[2,3]

Received: 15 September 2021 / Revised: 5 March 2022 / Accepted: 29 June 2022

Abstract
Divergence is used to measure the difference of two systems, and it is widely applied in many fields. To solve this problem more efficiently, Dempster–Shafer evidence theory has been proposed, different from the traditional probability distribution, and because of its processing advantages of uncertainty, has been widely used in many aspects of reality. In this paper, a new method of belief divergence measure of mass functions is proposed, named as Earth Mover's divergence of belief function, which is a generalization of Earth Mover's distance (Wasserstein distance). Compared with other existing methods of divergence measuring, the EM divergence can show good performance in the presence of higher degrees of uncertainty and more conflicts. Numerical examples help have a better understanding of the Earth Mover's divergence of belief function. Based on the new method of belief divergence measure, there is a combination model proposed to address the problem of data fusion. Application in target recognition is used to show the efficiency of the proposed method of divergence measure.

Keywords Dempster–Shafer theory · Kullback–Leibler divergence · Jessen–Shannon divergence · Earth Mover's divergence · Target recognition · Data fusion

Mathematics Subject Classification 03E75

Communicated by Graçaliz Pereira Dimuro.

✉ Fuyuan Xiao
xiaofuyuan@cqu.edu.cn ; doctorxiaofy@hotmail.com

Peilin Liu
liupeilinswu@163.com

1 School of Computer Science and Software, Southwest University of China, Chongqing 400715, China

2 School of Big Data and Software Engineering, Chongqing University, Chongqing 401331, China

3 National Engineering Laboratory for Integrated Aero-Space-Ground-Ocean Big Data Application Technology, Xi'an, China

1 Introduction

Dempster–Shafer theory (Dempster 1967; Shafer 1976), was first proposed by Dempster and then Shafer promoted it. It can deal with this uncertainty caused by ignorance. Evidence theory satisfies a weaker axiom system than traditional probability theory. Because D–S theory can deal with uncertainty in many different situations, it has applications in many expert areas, such as: pattern recognition (Xiao 2022), active learning (Zhang et al. 2020), decision-making (Garg and Chen 2020; Fei et al. 2020; Tang et al. 2020), possibility measurement (Garg 2021), fuzzy sets processing (Liu and Gao 2019; Ye et al. 2021), information volume (Deng and Deng 2021; Deng 2020), classification of sources (Liu et al. 2020a, b; Xu et al. 2017), fusing multi-sources (Xie et al. 2022; Xiao 2020), risk evaluation (Mo 2021; Zhou et al. 2020; Fu et al. 2020; Wang et al. 2021), prediction (Xiao 2022), entropy measure (Deng 2020; Liao et al. 2020; Yager 2018), reliability evaluation (Meng et al. 2020; Zhou et al. 2020), uncertainty processing (Yager et al. 2017; Feng et al. 2016, 2020; Tian et al. 2020), social model design and analysis (Ni et al. 2021; Cao et al. 2019; Meng et al. 2021a, b), complex networks (Wen and Cheong 2021; Lai et al. 2020), etc.

However, in evidence theory, when combined with highly conflicting evidence, it produces a counter-intuitive result. To handle this question, researchers have proposed different method to measure the difference(or divergence) among the evidence from different sources, such as: Jousselme et al.'s distance (Jousselme et al. 2001), divergence of belief function, Belief Jensen–Shannon (BJS) divergence (Xiao 2019), Kullback–Leibler divergence (Kullback 1997), belief interval-based distance (Han et al. 2016; Cheng and Xiao 2021), a new distance to measure basic probability assignment (BPA) Gao et al. (2021) based on information volume (Gao et al. 2021) and method of combining dependent evidences (Yong et al. 2004; Fei et al. 2019; Song and Xiao 2022). It is proved by many experiments and data analysis that the higher the correlation degree, the better fusion result of two bodies of evidence.

In Generative Adversarial Networks field, Wasserstein has proposed a new distance named as Wasserstein distance (Earth Mover's distance) Arjovsky et al. (2017), which is to evaluate the difference degree between two multi-dimensional distributions, where the distance between a single feature is measured, which we call the ground distance. The distance from the Earth Mover's distance (EMD) individual features "upgrade" to a full distribution.

In this paper, to give a new definition of divergence(or distance) among different evidence that can solve the issue of application in evidence fusion more efficiently. A new method of evidence divergence measurement is proposed, which is a generalization of Wasserstein distance(EMD) Arjovsky et al. (2017). The proposed divergence method meets the symmetry and it also can measure the difference degree of two basic probability assignments (BPAs)more accurately and more efficiently. Together with the proposed divergence measure method, a combination model based on the EM divergence is proposed to handle the problem of data fusion. Then, an application in target recognition shows how the proposed divergence method can get the highest support degree on the correct target.

The organization of this paper is as follows. In Sect. 2, the basic definitions of the Dempster–Shafer theory, the Dempster combination rule and different methods of divergence and distance measure are given. In Sect. 3, the definition of Earth Mover's divergence of belief function between two bodies of evidence is proposed and numerical examples to help have a better understanding of this method. In Sect. 4, the accuracy and efficiency of the Earth Mover's divergence of belief function in evidence fusion and target recognition are verified. In Sect. 5, there is a brief conclusion of this paper.

Table 1 Notation list

Notation	Definition
EMD	Earth Mover's distance
BPA	basic probability assignment
FOD	frame of discernment
BOE	body of evidence
$Div_{EM}(A, B)$	EM divergence between A and B
T	conversion matrix
D	divergence matrix
$<,>_F$	the Frobenius inner product
inf	infimum
Π	the set of all different transport plans
$Sim(A, B)$	similarity between A and B
SMM	similarity measure matrix
$Sup(A)$	support degree of A
$W(A)$	weight(or credibility degree) of A
$WAE(A)$	weight value of the evidence A

2 Preliminaries

2.1 Notation

A notation list is given as follows:

2.2 Dempster–Shafer theory

Uncertainty problem has attracted many attention (Deng and Jiang 2020; Zhou et al. 2020), where a lot of method have been proposed, such as extend Z-number (Jiang et al. 2020), evidence theory (Chang et al. 2021; Fujita and Ko 2020), complex-valued model (Xiao 2021), and others. We will briefly introduce Dempster–Shafer theory below.

Let Ω be an N number set of all possible hypotheses. N hypotheses in set Ω are mutually exclusive and exhaustive, called the frame of discernment (FOD), which is denoted as follows (Dempster 1967; Shafer 1976) :

$$\Omega = \{H_1, H_2, H_3, \cdots, H_N\}. \tag{1}$$

The power set of Ω composed with all possible subsets of Ω, denoted as 2^{Ω}:

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

$$2^{\Omega} = \{\emptyset, \{H_1\}, \ldots, \{H_N\}, \{H_1, H_2\}, \ldots, \{H_1, \ldots, H_N\}. \tag{3}$$

BPA is the critical part of the D–S theory and has a wide application, such as: information volume (Deng 2020; Deng and Deng 2021) and similarity measure (Deng 2020; Fan and Deng 2021). The mass function of BPA meets the following properties:

$$\sum_{H \in 2^{\Omega}} m(H) = 1, \tag{4}$$

$$m(\emptyset) = 0 \tag{5}$$

Mass function can also be extended to random permutation set (Deng 2022).

2.3 Combination rule of D–S theory

Two BPAs m_1 and m_2 are from two different sources. The combination rule is to combine m_1 and m_2 based on orthogonal sum. The result of combination is denoted as m, which is expressed as follows (Dempster 1967; Shafer 1976):

$$m(\emptyset) = 0, \tag{6}$$

$$m(A) = \frac{\sum_{B\cap C=A} m_1(B)\cdot m_2(C)}{1-K};\ A \neq \emptyset, \tag{7}$$

where the conflict of m_1 and m_2 is denoted as K, which is defined as Shafer (1976):

$$K = \sum_{B\cap C=\emptyset} m_1(B)\cdot m_2(C). \tag{8}$$

2.4 Different methods of divergence measure

2.4.1 Kullback–Leibler divergence

K–L divergence (Kullback 1997) is proposed to measure the divergence of two probability distribution and also is widely applied in information field.

Let P and Q be two probability distribution,

$P = \{P(x_1), P(x_2), \ldots, P(x_n)\}$ and $Q = \{Q(x_1), Q(x_2), \ldots, Q(x_n)\}$.

The K–L divergence is defined as Kullback (1997):

$$D_{KL}(P, Q) = \sum_{i=1}^{n} P(x_i) \log \frac{P(x_i)}{Q(x_i)}. \tag{9}$$

One thing that needs to be emphasized is Kullback–Leibler divergence is a kind of asymmetric distance, which means (Kullback 1997)

$$D_{KL}(P, Q) \neq D_{KL}(Q, P).$$

Moreover, when calculating the K–L divergence of two probability distribution, to avoid the denominator to be zero, 10^{-8}, which is a very small value near zero, is usually used to replace zero.

2.4.2 Jessen–Shannon divergence

The aim of Jessen–Shannon divergence is to make the K–L divergence symmetric. Given two probability distribution P and Q, the Jessen–Shannon divergence is defined as Xiao (2019)

$$D_{BJS}(P, Q) = D_{BJS}(Q, P) = \frac{D_{KL}(P, Q) + D_{KL}(Q, P)}{2}. \tag{10}$$

2.4.3 Earth Mover's distance

In mathematics, the EMD is also called as the Wasserstein distance. To have a better understanding of the definition of EMD, we take two arbitrary discrete distributions P_r and P_θ as example(Fig.1) Herrmann (2017)

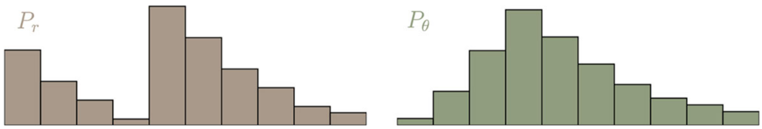
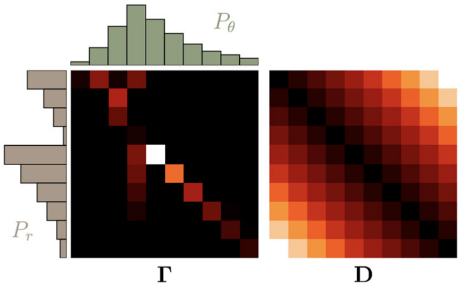


Fig. 1 Probability distribution P_r and P_θ each with ten states

Fig. 2 Transport plan T with P_r and P_θ, and distance D



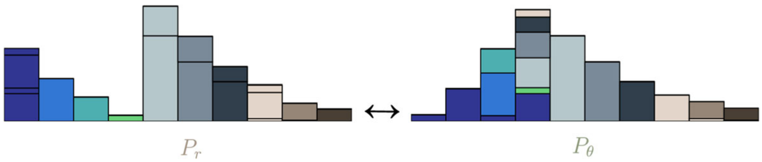


Fig. 3 Optimal transport plan between P_r and P_θ

Before calculating the EMD of this question, it is not hard to know that there are plenty of ways to transport the earth from one to the other side. But EMD is based on the optimal transport plan(Fig.3). Given the definition of the EMD (Arjovsky et al. 2017)

$$EMD(P_r, P_\theta) = \inf_{\gamma \in \Pi} < D, T >_F, \tag{11}$$

where $<, >_F$ is the Frobenius inner product. Π is the set of all distributions and its margins are P_r and P_θ, respectively. $T = \gamma(a, b)$ and $\gamma(a, b)$ simply states that the amount of earth that should be moved from place a to place b and it should satisfy $\sum_a \gamma(a, b) = P_r(b)$ as well as $\sum_b \gamma(a, b) = P_\theta(a)$. $D = ||a - b||$ is the distance between place a and b. (Figs.1–3 are obtained from a blog that discuss Wasserstein GAN and the Kantorovich–Rubinstein Duality (Herrmann 2017))

Another important thing should be mentioned that the EMD is independent of the order of the elements in the probability distributions.

Suppose that $\Gamma \sim U[0, 1]$ is uniformly distributed on the unit interval. P_r is the distribution of $(0, \Gamma) \in R^2$, which is uniformly distributed on a straight vertical line passing through the origin. $f_\theta(\gamma) = (\theta, \gamma)$ with a single parameter. In this case, we can easily make a comparison among the different distances

From the table, it can be seen that the change in EMD is smoother, while K–L divergence and BJS divergence both have sudden change.

Table 2 The comparison among different distances(2 discrete probability distributions P_r and P_θ)

Name	Formula	$\theta = 0$	$\theta \neq 0$
$D_{KL}(P_r, P_\theta)$	$\sum_{i=1}^{n} P(x_i) \log \frac{P(x_i)}{Q(x_i)}$	0	∞
$D_{BJS}(P_r, P_\theta)$	$\frac{1}{2}(D_{KL}(P, Q) + D_{KL}(Q, P))$	0	$\log 2$
$EMD(P_r, P_\theta)$	$\inf_{\gamma \in \Pi} < D, T >_F$	$\|\theta\|$	$\|\theta\|$

2.5 EM belief divergence

EM belief divergence is a new belief divergence measuring method generalized from EMD (Arjovsky et al. 2017). EM divergence retains the advantages of EMD and modifies some of the definition and algorithm of EMD for better application to evidence theory. The new belief divergence satisfies the symmetry and can measure the difference between two BPAs more accurately and efficiently. Its advantages can be roughly divided into the following points:

- signatures are greater compact, and the value of transferring “earth” displays the thinking of nearness properly, besides the quantization issues of most different measures.
- “lifts” this distance from individual features to full distributions.
- match focal elements in a simple one-to-one correspondence.
- the definition of ground distance can be modified according to individual wishes, and you can change the different distance measuring methods which means you can choose a optimal distance for your problem.

Moreover, it has a good performance on evidence fusion and target recognition with an example application in the later part of this paper.

3 A new divergence of belief function based on EMD

3.1 Earth Mover’s divergence of belief function

Different from the classical EMD, the EM divergence of belief function on BPAs requires that the focal elements of two BPAs must correspond to each other.

Give an easy example here:

$$[\{x_1\} : 0.5, \{x_2\} : 0.3, \{x_1, x_2\} : 0.2]; [\{x_1\} : 0.5, \{x_2\} : 0.2, \{x_1, x_2\} : 0.3].$$

If these two sets were taken as traditional probability distributions, the EMD between them is zero, because the EMD does not depend on the order of elements which also means it treats each of the focal elements as the same. However, if these two sets were taken as two BPAs on the same FOD, we can easily know the information that these two BPAs give us is totally different, so the divergence or distance between these two BPAs is definitely not zero. This easy example can intuitively reflect the differences between EMD and EM divergence. EM divergence gives each focal element the inner meaning makes each of them different. Later, there are some numerical examples in section 4, and we can see the difference between EM belief divergence and traditional EMD.

Suppose the frame of discernment Θ contain N number of hypotheses which are mutually exclusive and exhaustive and there are two BPAs m_1 and m_2 on the FOD Θ.

The power set of Θ is indicated by 2^{Θ}, denoted as:

$$2^{\Theta} = \{\emptyset, \{A_1\}, \ldots, \{A_N\}, \{A_1, A_2\}, \ldots, \{A_1, \ldots, A_N\}, \ldots\}. \tag{12}$$

Give the definition for EM divergence of belief function here:

$$Div_{EM}(m_1, m_2) = \inf_{T \in \Pi} < T, D >_F, \tag{13}$$

where T is the conversion matrix which dimension is $2^N \times 2^N$ and $T_{i,j}$ represents how much mass should be moved from A_i to A_j. D is divergence matrix which dimension is also $2^N \times 2^N$ and $D_{a,b} = ||A_a, A_b||$ represents the distance from A_a to A_b ($A_a, A_b \in \Theta$). Generally speaking, the BPA vector is just like a one-dimensional array, so the distance can then be the difference between the ordinates of the two focal elements. Of course, you can also use the Euclidean distance and other definitions of distance as the distance here. Moreover, $<, >_F$ is the Frobenius inner product (sum of all the element-wise products) of T and D. The result of $< T, D >_F$ can be taken as the total work of moving these mass. inf stands for the infimum and Π is the set of all different kinds of transport plans.

There is another way to show the Div_{EM}:

$$Div_{EM} = \min WORK(m_1, m_2). \tag{14}$$

The meaning of this kind of expression is to do the least amount of work to change m_1 to m_2. And the definition of work can be described figuratively as the mass to be moved multiplied by the distance moved, which is just like the definition of $WORK$ in the physics.

After getting the divergence of two BPAs, now the similarity of them is expressed as:

$$Sim(m_1, m_2) = 1 - Div_{EM}(m_1, m_2). \tag{15}$$

3.2 Numerical examples

Example 1 Assume that BPAs m_1 and m_2 are on the frame of discernment $\Theta = \{x_1, x_2\}$.

$$m_1 = [\{x_1\} : 0.6, \{x_2\} : 0.4]; m_2 = [\{x_1\} : 0.6, \{x_2\} : 0.4]$$

$$Div_{EM}(m_1, m_2) = 0.$$

It is easy to find that the each element in matrix T and D are zero. So, the divergence between m_1 and m_2 is zero and $Sim(m_1, m_2)$=1-$Div_{EM}(m_1, m_2)$=1 which means m_1 and m_2 are totally same. It is the same as using Jousselme et al.'s distance (Jousselme et al. 2001) to calculate which is also equal to 0.

Example 2 Same conditions as the previous example.

$$m_1 = [\{x_1\} : 1, \{x_2\} : 0]; m_2 = [\{x_1\} : 0, \{x_2\} : 1],$$

$$D = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad T = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix},$$

$$Div_{EM}(m_1, m_2) =< T, D >_F= 1.$$

There is only one kind of transport plan in this case which is move $1 \in \{x_1\}$ to $\{x_2\}$ if we want to change m_1 to m_2. So, we can calculate the EM divergence of them which is 1. That means m_1 and m_2 get the biggest divergence on this frame of discernment, in another way, they have the biggest conflict.

$$Sim(m_1, m_2) = 1 - Div_{EM}(m_1, m_2) = 0.$$

If we take these two BPAs as normal probability distribution, the EMD of them is zero because in traditional EMD, it is independent of the order of set, which means each element in the set is the same. It is the same as using Jousselme et al.'s distance (Jousselme et al. 2001) to calculate, which is also equal to 1.

Example 3 Change the m_2 in Example.1 to $m_2 = [\{x_1\} : 0.4, \{x_2\} : 0.6]$ and then continue to calculate the divergence between m_1 and m_2.

Then, we can get the matrix T and D:

$$D = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad T = \begin{pmatrix} 0 & 0.2 \\ 0 & 0 \end{pmatrix},$$

$$Div_{EM}(m_1, m_2) =< T, D >_F = 0.2.$$

From matrix T, we can see that the transport plan that we choose is move $0.2 \in \{x_1\}$ to $\{x_2\}$ so that they can be totally same. The EM belief divergence of belief function of m_1 and m_2 is 0.2. It is the same as using Jousselme et al.'s distance (Jousselme et al. 2001) to calculate which is also equal to 0.2.

Example 4 (mathematical property) This example is to show some of the mathematical property of the EM divergence, which is inspired by Jousselme et al.'s distance (Jousselme et al. 2001).

Given a simple case of set which has 5 objects $\Theta = \{1, 2, 3, 4, 5\}$. $P(\Theta)$ has 32 subsets of Θ. At the beginning, we set 31 non-empty subset with equal distribution, and the mass of the one empty set($\{\emptyset\}$) is 0. The process of this algorithm is to gradually increase $m(A*)(\{A*\}$ contains one object) from 0 to 1 with each increase a δ from other 31 subsets, which means31 subset each steps get a decrease of $\delta/31$. (δ is fixed to 0.02)

And we can see the algorithm converges right to the solution, when it contains several objects in Fig 4, compared with the Jousselme et al's.

The distance result calculated in various way may be slightly different, but when compared with in the same way, the EM divergence of belief function has its accuracy and efficiency. Also, the EM divergence of belief function has a good performance on evidence combination and there is a subsequent example application to show its benefits.

4 Application in evidence target recognition

In this part, the effectiveness and accuracy of the EM divergence of belief function is illustrated with an application of target recognition. First, a brief introduction of the evidence combination algorithm is given and then the result and comparison with other methods are shown.

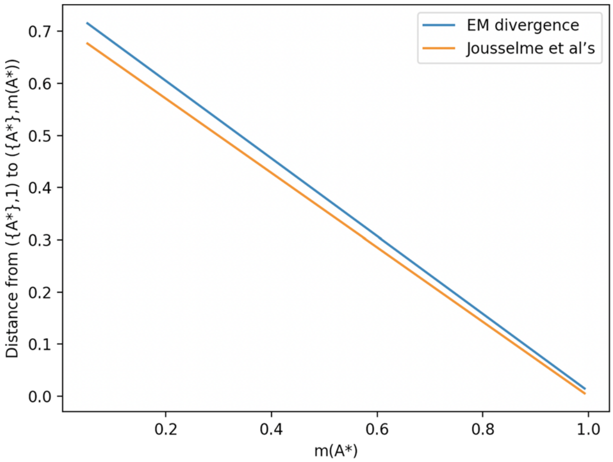


Fig. 4 Convergence to the right solution

4.1 Preparation

The input requirement for this algorithm are two BPAs and the output is a new BPA after the fusion of two BPAs. The programming language used in the experiment is Python 3.9, and the experimental platform is :

PyCharm 2020.3.5 (Community Edition)
Runtime version: 11.0.10+8-b1145.96 x86_64
CPU: 1.4 GHz 4Core Intel Core i5
RAM: 8 GB 2133 MHz LPDDR3
macOS 10.16

4.2 Evidence combination

Suppose m_1 and m_2 are two BPAs on the same FOD $\Omega = \{A, B, C\}$ and the divergence of m_1 and m_2 can be calculated by the proposed method. Then, we can get the similarity degree of two bodies of evidence by $Eq.$(15).

$Step.$1.

Assume that the number of BOE (Body Of Evidence) is n. After we get all the similarity degree of each two bodies of evidence, then the matrix of similarity measure(SMM) can be built.

$$SMM = \begin{pmatrix} 1 & \dots & S_{1k} & \dots & S_{1n} \\ \vdots & \ddots & \vdots & & \vdots \\ S_{k1} & \cdots & 1 & \cdots & S_{kn} \\ \vdots & & \vdots & \ddots & \vdots \\ S_{1n} & \cdots & S_{nk} & \cdots & 1 \end{pmatrix}, \tag{16}$$

where S_{ij} is the similarity degree of m_i and m_j.

$Step.$2.

Table 3 The sensor collection (Song and Deng 2019)

Sensors\Targets	{M}	{N}	{P}	{M,P}	{M,N,P}
$S_1 : m_1$	0.00	0.08	0.12	0.10	0.70
$S_2 : m_2$	0.30	0.05	0.05	0.20	0.40
$S_3 : m_3$	0.15	0.10	0.10	0.15	0.50
$S_4 : m_4$	0.30	0.05	0.10	0.10	0.45

Get the support degree of BOE m_i(i ∈ [1,n]):

$$Sup(m_i) = \sum_{j=1, j \neq i}^{n} Sim(m_i, m_j). \tag{17}$$

Step.3.
Obtain the weight (or the credibility degree) of each BOE by using support degree:

$$W(m_i) = \frac{Sup(m_i)}{\sum Sup(m_j)}, \tag{18}$$

where the weight of each BOE indicates the relative importance of the each evidence source and $\sum W_i = 1$ should be satisfied.
Step.4.
Get the average weight value of the evidence WAE (or the modified average value).

$$WAE(A) = \sum_{i=1}^{n} W(m_i) \times m_i(A), \tag{19}$$

where A is the focal element and WAE can be taken as a new weighted BPA of these BOE.
Step.5.
Combine the WAE using the Dempster combination rule $n - 1$ times.

4.3 Numerical example

Let $\Omega = \{M, N, P\}$ be the frame of discernment. There is a sensor collection in Table 3 Song and Deng (2019). Then, find the fusion result step by step.

Step.1: Obtain the similarity degree of each two bodies of evidence and construct the SMM.

$$SMM = \begin{pmatrix} 1.00 & 0.60 & 0.78 & 0.68 \\ 0.60 & 1.00 & 0.80 & 0.90 \\ 0.78 & 0.80 & 1.00 & 0.80 \\ 0.68 & 0.90 & 0.90 & 1.00 \end{pmatrix}.$$

Step.2: Obtain the degree of support of each BPA from the similarity measure matrix.

$$Sup(m_1) = \sum_{k \neq 1}^{4} Sim(m_1, m_k) = 2.06,$$

$$Sup(m_2) = \sum_{k=1, k \neq 2}^{4} Sim(m_2, m_k) = 2.30,$$

$$Sup(m_3) = \sum_{k=1, k \neq 3}^{4} Sim(m_3, m_k) = 2.38,$$

$$Sup(m_4) = \sum_{k=1, k \neq 4}^{4} Sim(m_4, m_k) = 2.38.$$

Step.3: Obtain the weight of each BPA.

$$W(m_1) = \frac{Sup(m_1)}{\sum Sup(m_k)} = 0.226,$$

$$W(m_2) = \frac{Sup(m_2)}{\sum Sup(m_k)} = 0.252,$$

$$W(m_3) = \frac{Sup(m_3)}{\sum Sup(m_k)} = 0.261,$$

$$W(m_4) = \frac{Sup(m_4)}{\sum Sup(m_k)} = 0.261.$$

Step.4: Obtain the weight average of evidence WAE (The new BPA).

$$WAE(M) = \sum_{i=1}^{n} W(m_i) \times m_i(M) = 0.193,$$

$$WAE(N) = \sum_{i=1}^{n} W(m_i) \times m_i(N) = 0.070,$$

$$WAE(P) = \sum_{i=1}^{n} W(m_i) \times m_i(P) = 0.091,$$

$$WAE(MP) = \sum_{i=1}^{n} W(m_i) \times m_i(M, P) = 0.138,$$

$$WAE(MNP) = \sum_{i=1}^{n} W(m_i) \times m_i(M, N, P) = 0.507.$$

Step.5: Combine the WAE using the Dempster rule 3 times and the result of fusion is shown in Table.3.

To verify the efficiency of it, there are three typical fusion models included in this example to compare with the proposed method, which are Deng et al.'s method (Yong et al. 2004; Gao and Deng 2020; Li and Deng 2018; Cui et al. 2022) and Belief Jessen–Shannon divergence (Xiao 2019). The conflicting data processing method of Deng et al.'s is one of the most typical and most effective methods. Belief Jessen–Shannon divergence (Xiao 2019) is a measure of the difference between the evidence and the degree of conflict, which is widely applied

Table 4 The fusion result of sensors

Method Name	{M}	{N}	{P}	{M,P}	{M,N,P}
BJS Divergence Xiao (2019)	0.4744	0.0666	0.1850	0.1691	0.1049
Deng *et al*. Yong et al. (2004)	0.4790	0.0686	0.1868	0.1637	0.1019
EM divergence of belief function	0.4861	0.0679	0.1829	0.1627	0.1004

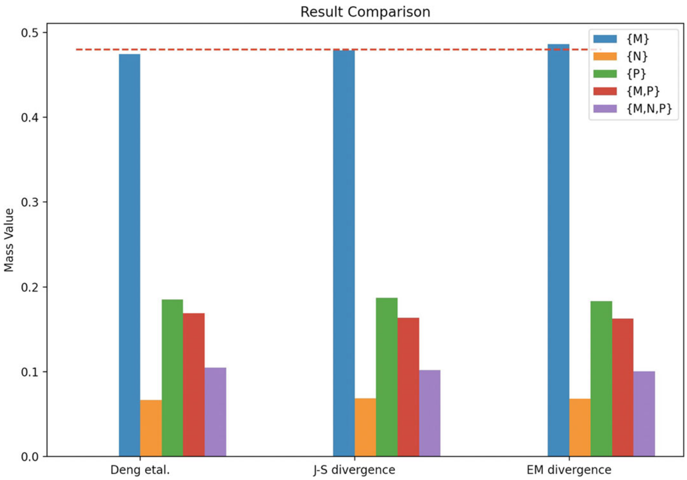


Fig. 5 Result comparison

in multi-sensor data fusion. Compared with these three methods, the efficiency of the EM divergence of belief function can be explained convincingly.

From the result in Table.3 and Fig.5, it can intuitively reflect the comparison of result values of different methods. The reliability for target {M}(0.4861) is the highest and the reliability for target {N}(0.0679) and {M,N,P}(0.1004) is the least, which shows that it is not only effective in determining the target, but also has a great performance in reducing the uncertainty of the final target. The proposed method has larger support degree on correct target{M}. Both JS divergence and Deng et al.'s has a support degree on the correct target which is lower than 0.4800, but the proposed divergence can get a higher support degree on the correct one({M}). It is not hard to know that the difference among these methods is not large, but we can see that under such an uncertain environment, EM divergence has a better performance. At least, it is not worse than these two methods.

5 Conclusion

So as to resolve the issue of divergence measure among the evidence from different sources, the contributions of this paper are that a new method of measuring the evidence divergence of BPAs is proposed, named as Earth Mover's divergence of belief function. This divergence method is the generalization of Earth Mover's distance(Wasserstein distance) Arjovsky et al. (2017) which can still provide a meaningful and smooth representation of the distance between two probability assignment in lower dimensional manifolds without overlaps. There is a significant difference between EM divergence and EMD because it has been improved on the basis of EMD to be more suitable for the evidence measuring. EMD is independent of the order of the elements in the probability distributions and EM divergence require that the focal elements of two BPAs must corresponding to each other. Moreover, compared with other existing methods of divergence measuring, the EM divergence can show good performance in the presence of higher degrees of uncertainty and more conflicts. The example application of target recognition which is on the basis of data fusion has illustrated that the proposed divergence can get the highest reliability to the correct target.

Acknowledgements This research is supported by the National Natural Science Foundation of China (No. 62003280), Chongqing Talents: Exceptional Young Talents Project (No. cstc2022ycjh-bgzxm0070), and Chongqing Overseas Scholars Innovation Program (No. cx2022024). The authors greatly appreciate the reviewers' suggestions and editor's encouragement.

References

Arjovsky M, Chintala S, Bottou L (2017) Wasserstein generative adversarial networks. Int Conf Mach Learn. PMLR,214–223

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

Chang L, Zhang L, Fu C, Chen Y-W (2021) Transparent digital twin for output control using belief rule base. IEEE Trans Cybern. https://doi.org/10.1109/TCYB.2021.3063285,

Cheng C, Xiao F (2021) A distance for belief functions of orderable set. Pattern Recogn Lett 145:165–170

Cui H, Zhou L, Li Y, Kang B (2022) Belief entropy-of-entropy and its application in the cardiac interbeat interval time series analysis. Chaos Solitons Fractals 155:111736

Dempster AP (1967) Upper and lower probabilities induced by a multivalued mapping. Ann Math Stat 382:325–339 (**04**)

Deng Y (2020) Information volume of mass function. Int J Comput Commun Control 15(6):3983

Deng Y (2020) Uncertainty measure in evidence theory. SCIENCE CHINA Inf Sci 63(11):210201

Deng Y (2022) Random permutation set. Int J Comput Commun Control 17(1):4542

Deng J, Deng Y (2021) Information volume of fuzzy membership function. Int J Comput Commun Control 16(1):4106

Deng X, Jiang W (2020) On the negation of a Dempster-Shafer belief structure based on maximum uncertainty allocation. Inf Sci 516:346–352

Fan L, Deng Y (2021) Determine the number of unknown targets in Open World based on Elbow method. IEEE Trans Fuzzy Syst 29(5):986–995

Fei L, Feng Y, Liu L (2019) Evidence combination using OWA-based soft likelihood functions. Int J Intell Syst 34(9):2269–2290

Fei L, Lu J, Feng Y (2020) An extended best-worst multi-criteria decision-making method by belief functions and its applications in hospital service evaluation. Comput Ind Eng 142:106355

Feng F, Cho J, Pedrycz W, Fujita H, Herawan T (2016) Soft set based association rule mining. Knowl-Based Syst 111:268–282

Feng F, Xu Z, Fujita H, Liang M (2020) Enhancing PROMETHEE method with intuitionistic fuzzy soft sets. Int J Intell Syst 35:1071–1104

Fu C, Xue M, Chang W, Xu D, Yang S (2020) An evidential reasoning approach based on risk attitude and criterion reliability. Knowl-Based Syst 199:105947

Fujita H, Ko Y-C (2020) A heuristic representation learning based on evidential memberships: Case study of UCI-SPECTF. Int J Approx Reason 120

Gao X, Deng Y (2020) The pseudo-pascal triangle of maximum Deng entropy. Int J Comput Commun Control 15(1):1006

Gao X, Pan L, Deng Y (2021) A generalized divergence of information volume and its applications. Eng Appl Artif Intell 108:104584

Gao Q, Wen T, Deng Y (2021) Information volume fractal dimension. Fractals 29(8):2150263

Garg H (2021) A new possibility degree measure for interval-valued q-rung orthopair fuzzy sets in decision-making. Int J Intell Syst 36(1):526–557

Garg H, Chen S (2020) Multiattribute group decision making based on neutrality aggregation operators of q-rung orthopair fuzzy sets. Inf Sci 517:427–447

Han D, Dezert J, Yang Y (2016) Belief interval-based distance measures in the theory of belief functions. IEEE Trans Syst Man Cybern Syst 48(6):833–850

Herrmann V (2017) Wasserstein gan and the kantorovich-rubinstein duality. URL: https://vincentherrmann.github. io/blog/wasserstein

Jiang W, Cao Y, Deng X (2020) A novel Z-network model based on Bayesian network and Z-number. IEEE Trans Fuzzy Syst 28(8):1585–1599

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

Kullback S (1997) Information theory and statistics. Courier Corporation

Lai JW, Chang J, Ang L, Cheong KH (2020) Multi-level information fusion to alleviate network congestion. Inf Fusion 63:248–255

Li Y, Deng Y (2018) Generalized ordered propositions fusion based on belief entropy. Int J Comput Commun Control 13(5):792–807

Liao H, Ren Z, Fang R (2020) A Deng-entropy-based evidential reasoning approach for multi-expert multi-criterion decision-making with uncertainty. Int J Comput Commun Control 13(1):1281–1294

Liu Z-G, Huang L-Q, Zhou K, Denoeux T(2020) Combination of transferable classification with multisource domain adaptation based on evidential reasoning. IEEE Trans Neural Netw Learn Syst. https://doi.org/10.1109/TNNLS.2020.2995862,

Liu P, Gao H (2019) Some intuitionistic fuzzy power bonferroni mean operators in the framework of Dempster-Shafer theory and their application to multicriteria decision making. Appl Soft Comput 85:105790

Liu Z, Liu Y, Dezert J, Cuzzolin F (2020) Evidence combination based on credal belief redistribution for pattern classification. IEEE Trans Fuzzy Syst 28(4):618–631

Meng D, Li Y, He C, Guo J, Lv Z, Wu P (2021) Multidisciplinary design for structural integrity using a collaborative optimization method based on adaptive surrogate modelling. Mater Des 206:109789

Meng D, Hu Z, Wu P, Zhu S-P, Correia JA, De Jesus AM (2020) “Reliability-based optimisation for offshore structures using saddlepoint approximation. In: Proceedings of the Institution of Civil Engineers-Maritime Engineering, vol. 173, no. 2.Thomas Telford Ltd, 33–42

Meng D, Xie T, Wu P, He C, Hu Z, Lv Z (2021) An uncertainty-based design optimization strategy with random and interval variables for multidisciplinary engineering systems. In: Structures, vol. 32.Elsevier, pp 997–1004

Mo H (2021) A SWOT method to evaluate safety risks in life cycle of wind turbine extended by D number theory. J Intell Fuzzy Syst 40(3):4439–4452

Ni L, Chen Y-W, de Brujin O (2021) Towards understanding socially influenced vaccination decision making: an integrated model of multiple criteria belief modelling and social network analysis. Eur J Oper Res 293(1):276–289

Shafer G (1976) A mathematical theory of evidence. Princeton university press Princeton,1

Song Y, Deng Y (2019) Divergence measure of belief function and its application in data fusion. IEEE Access 7:107465–107472

Song X, Xiao F (2022) Combining time-series evidence: A complex network model based on a visibility graph and belief entropy. Appl Intell. https://doi.org/10.1007/s10489-021-02956-5

Tang M, Liao H, Mi X, Xu X, Herrera F (2020) Dynamic subgroup-quality-based consensus in managing consistency, nearness, and evenness quality indices for large-scale group decision making under hesitant environment. J Oper Res Soc 1–14

Tian Y, Liu L, Mi X, Kang B (2020) ZSLF: A new soft likelihood function based on Z-numbers and its application in expert decision system,. IEEE Trans Fuzzy Syst. https://doi.org/10.1109/TFUZZ.2020.2997328,

Wang H, Fang Y-P, Zio E (2021) Risk assessment of an electrical power system considering the influence of traffic congestion on a hypothetical scenario of electrified transportation system in new york state. IEEE Trans Intell Transp Syst 22(1):142–155

Wen T, Cheong KH (2021) The fractal dimension of complex networks: a review. Inf Fusion 73:87–102

Xiao F (2022) CEQD: A complex mass function to predict interference effects. IEEE Trans Cybern 52(8): 7402–7414

Xiao F (2021) On the maximum entropy negation of a complex-valued distribution. IEEE Trans Fuzzy Syst 29(11): 3259–3269

Xiao F (2022) CaFtR: A fuzzy complex event processing method. Int J Fuzzy Syst 24(2): 1098–1111

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

Xiao F (2020) Evidence combination based on prospect theory for multi-sensor data fusion. ISA Trans 106:253–261
Xie D, Xiao F, Pedrycz W (2022) Information quality for intuitionistic fuzzy values with its application in decision making. Eng Appl Artif Intell. https://doi.org/10.1016/j.engappai.2021.104568,
Xu X, Zheng J, Yang J-B, Xu D-L, Chen Y-W (2017) Data classification using evidence reasoning rule. Knowl-Based Syst 116:144–151
Yager RR (2018) Interval valued entropies for Dempster-Shafer structures. Knowl-Based Syst 161:390–397
Yager RR, Elmore P, Petry F (2017) Soft likelihood functions in combining evidence. Inf Fusion 36:185–190
Ye J, Zhan J, Ding W, Fujita H (2021) A novel fuzzy rough set model with fuzzy neighborhood operators. Inf Sci 544:266–297
Yong D, Wenkang S, Zhenfu Z, Qi L (2004) Combining belief functions based on distance of evidence. Decis Support Syst 38(3):489–493
Zhang S, Han D, Yang Y (2020) Active learning based on belief functions. SCIENCE CHINA Inf Sci 63(11):210205
Zhou J, Su X, Qian H (2020) Risk assessment on offshore photovoltaic power generation projects in china using D umbers and anp. IEEE Access 99:1–1
Zhou Z, Feng Z, Hu C, Hu G, He W, Han X (2020) Aeronautical relay health state assessment model based on belief rule base with attribute reliability. Knowl-Based Syst 197:105869
Zhou M, Liu X-B, Chen Y-W, Qian X-F, Yang J-B, Wu J (2020) Assignment of attribute weights with belief distributions for MADM under uncertainties. Knowl-Based Syst 189:105110