

Belief Interval-Based Distance Measures in the Theory of Belief Functions

Deqiang Han, *Member, IEEE*, Jean Dezert, and Yi Yang

Abstract—In belief functions related fields, the distance measure is an important concept, which represents the degree of dissimilarity between bodies of evidence. Various distance measures of evidence have been proposed and widely used in diverse belief function related applications, especially in performance evaluation. Existing definitions of strict and nonstrict distance measures of evidence have their own pros and cons. In this paper, we propose two new strict distance measures of evidence (Euclidean and Chebyshev forms) between two basic belief assignments based on the Wasserstein distance between belief intervals of focal elements. Illustrative examples, simulations, applications, and related analyses are provided to show the rationality and efficiency of our proposed measures for distance of evidence.

Index Terms—Belief interval, dissimilarity, distance of evidence, evidence theory, the theory of belief functions.

I. INTRODUCTION

THE theory of belief functions, also called Dempster–Shafer evidence theory (DST) [1], is an important mathematical framework for uncertainty modeling and reasoning. It has been applied to information fusion [2], pattern recognition [3], [4], multiple-attribute decision making [5], fault diagnosis [6], etc. DST has some limitations (see [7]–[9] for discussions). Generalized or refined theories were proposed including transferable belief model [10] and Dezert–Smarandache theory [7], [11], etc.

In DST, the basic belief assignment (BBA) is a common way for modeling (epistemic) uncertainty. The distance of evidence is a crucial metric for measuring the distance between

two BBAs. It indicates a BBA is “far” from or “close” to another one. In many belief functions related applications, the distance of evidence is required. Such belief function-related applications can be categorized into two types. The first type is the performance evaluation or optimization [12]–[16]. For example, in the performance evaluation of BBA approximation [16], which aims to simplify the BBA to reduce the computational complexity, the distance of evidence is needed to measure accuracy of an approximated BBA (the one closer to the original BBA is better). Furthermore, some BBA approximation approach is directly based on the distance minimization [17], therefore, the distance of evidence is indispensable. The second type of applications is to determine the agreement between sources of information. For example, in the clustering analysis [4], [18], [19] and the determination of discounting factors [20], [21], the distance of evidence is required.

Since the distance of evidence is a very crucial concept in many applications, it has attracted increasing research interest recently in the belief functions community. Many definitions of distance (or dissimilarity) measures have been proposed in the past two decades [22]. Some of them are nonstrict distance metrics, although they are often called “distance.” In practice, Jousselme’s (strict) distance of evidence [13] and Tessem’s (nonstrict) betting commitment distance [23] (also called the pignistic probability distance) are most frequently used ones. A fuzzy set-based distance of evidence was also proposed in our previous work [24]. Jousselme and Maupin [22] provided an excellent survey on available works on the distance of evidence, where many definitions are introduced and compared.

Various types of distance of evidence have been proposed under the geometric interpretation [25] of the DST, where a BBA is considered as a vector of a Cartesian-alike space and each focal element is deemed as a base of the space [22]. However, all existing distances of evidence have their own limitations. First, a strict distance metric should satisfy the requirements including the non-negativity, nondegeneracy, symmetry, and triangular inequality. None of the existing distances of evidence except for Jousselme’s distance can satisfy all the requirements, i.e., they are not strict distance metrics. This is due to the switch between theoretical frameworks. For example, Tessem’s betting commitment distance [23] first transforms BBAs into pignistic probabilities, and fuzzy set-based distance of evidence [24] first transforms BBAs into fuzzy membership functions (FMFs). Such switches between different frameworks lead to the loss of information, thus

Manuscript received February 2, 2016; revised May 30, 2016 and September 16, 2016; accepted November 5, 2016. This work was supported by the Grant for State Key Program for Basic Research of China (973) under Grant 2013CB329405, in part by the National Natural Science Foundation under Grant 61573275 and Grant 61671370, in part by the Foundation for Innovative Research Groups of the National Natural Science Foundation of China under Grant 61221063, in part by the Science and Technology Project of Shaanxi Province under Grant 2013KJXX-46, in part by the Post-Doctoral Science Foundation of China under Grant 2016M592790, and in part by the Fundamental Research Funds for the Central Universities under Grant xjj2014122 and Grant xjj2016066. This paper was recommended by Associate Editor A. H. Tan.

D. Han is with the School of Electronic and Information Engineering, Xi’an Jiaotong University, Xi’an 710049, China (e-mail: deqhan@gmail.com).

J. Dezert is with ONERA, F-91761 Palaiseau, France (e-mail: jean.dezert@onera.fr).

Y. Yang is with the SKLSVMS, School of Aerospace, Xi’an Jiaotong University, Xi’an 710049, China (e-mail: jiafeiyi@mail.xjtu.edu.cn).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TSMC.2016.2628879

the distance between BBAs cannot be described precisely using these measures. Therefore, their strictness cannot be assured and they may encounter counter-intuitive results when measuring the distance between different BBAs. Although Jousselme's distance is a strict metric and performs well in many cases, it still has some unsatisfactory behaviors based on our experiments, e.g., the lack of discriminability in some cases and the maximum value problem as pointed out and analyzed in this paper. Due to the limitations of existing distance measures, we are motivated to propose better strict distance measures of evidence. We propose to use belief intervals [1] $[\text{Bel}(A), \text{Pl}(A)]$ of each focal element A to describe the closeness between BBAs, where $\text{Bel}(A)$ and $\text{Pl}(A)$ are, respectively, the belief and plausibility of a focal element A computed from the given BBA defined on a known frame of discernment (FOD). If we consider that a BBA is used to model the uncertainty as a whole for all focal elements, then the belief interval of each focal element in a BBA represents the uncertainty of the corresponding proposition. If we use all belief intervals of a BBA as a whole as its "feature" vector, then the distance between the feature vectors of different BBAs describes the difference between them. Since a belief interval is an interval number, the distance between the same focal element's two belief intervals in two BBAs can be calculated by Wasserstein's distance of interval numbers [26]. Based on all the distance values between belief intervals, we design a Euclidean-family distance using the sum of squares of all belief intervals' distance values, and a Chebyshev-family distance using the maximum of all belief intervals' distance values, respectively, to measure the distance between two different feature vectors of belief intervals, and thus to measure the distance between two BBAs. Our new definitions directly use the belief intervals defined in the DST, i.e., there is no switch between different theoretical frameworks. It can be proved that our new proposed measures of distance of evidence are strict distance metrics satisfying the requirements of non-negativity, nondegeneracy, symmetry, and triangle inequality. This paper extends our preliminary results in [27], where the basic idea of the belief interval-based distance is briefly introduced and a few illustrative examples are provided. In this paper, the limitations of existing distances are summarized more specifically, and the causes of these limitations are analyzed. More detailed formulations, proofs, and theoretical analyses of the new proposed distance measures are provided. More examples, simulations, and related analyses are provided for comparison between our proposed distances and the existing ones. An application of the proposed distances of evidence in the BBA approximation and an application of multiple criteria decision making (MCDM) using the proposed distance of evidence is also provided. These are all added values (contributions) of this paper.

The rest of this paper is organized as follows. Basics of the theory of belief functions are briefly introduced in Section II. The geometric interpretation and some commonly used distance measures of evidence are reviewed in Section III. Limitations of existing measures are explained based on illustrative examples in Section III. In Section IV, two new distance

metrics in DST are proposed based on the belief intervals and the distance between interval numbers. The proof of our proposed distance metrics' strictness, and the comparisons between our measures and distance bounds are also provided in Section IV. In Section V, examples, simulations, applications and related analyses are provided based on the comparison between new metrics and some existing ones from different aspects to show the rationality and efficiency of our new metrics. Section VI concludes this paper.

II. BASICS OF THEORY OF BELIEF FUNCTIONS

The theory of belief functions was first proposed by Dempster and then further developed by Shafer, therefore, it is usually called DST [1]. It has become an important theory and tool for uncertainty modeling and reasoning.

The basic concept of the theory of belief functions is the FOD, which represents the discourse domain of the problem we are interested in. Under the closed-world assumption, the FOD: $\Theta = \{\theta_1, \dots, \theta_n\}$ is defined as a set of n mutually exclusive and exhaustive elements. If a set function $m : 2^\Theta \rightarrow [0, 1]$, where 2^Θ is the powerset of Θ ,¹ satisfies

$$\sum_{A \subseteq \Theta} m(A) = 1, \quad m(\emptyset) = 0 \quad (1)$$

and if $m(A) \geq 0$ holds, then m is called a BBA (or mass function) over the FOD Θ . All the sets $A \in 2^\Theta$ satisfying $m(A) > 0$ are called the focal elements. Each focal element represents a proposition in the FOD. Given a BBA, a body of evidence (BOE) [1] can be determined, which is defined as the set of focal elements and their corresponding mass assignments.

A belief function over the FOD Θ , denoted by $\mathbf{Bel} : 2^\Theta \rightarrow [0, 1]$, is defined as

$$\text{Bel}(A) = \sum_{B \subseteq A} m(B), \quad \forall A \subseteq \Theta. \quad (2)$$

A plausibility function over the FOD Θ , denoted by $\mathbf{Pl} : 2^\Theta \rightarrow [0, 1]$, is defined as

$$\text{Pl}(A) = \sum_{B \cap A \neq \emptyset} m(B), \quad \forall A \subseteq \Theta. \quad (3)$$

The plausibility function and the belief function satisfy [1]

$$\text{Pl}(A) = 1 - \text{Bel}(\bar{A}) \quad (4)$$

where \bar{A} is the complementary proposition of $A \in 2^\Theta$. The plausibility $\text{Pl}(A)$ and the belief $\text{Bel}(A)$ constitute a belief interval $[\text{Bel}(A), \text{Pl}(A)]$. The length of the belief interval $[\text{Bel}(A), \text{Pl}(A)]$ represents the degree of imprecision for the proposition or focal element A . The non-null mass value assigned to Θ represents the degree of ignorance, i.e., the "unknown" state. Furthermore, in DST, different uncertainty measures have been proposed such as nonspecificity [28], ambiguity measure [29], aggregated uncertainty [30], and distance-based uncertainty measures [31].

The evidence combination rules are for uncertainty reasoning, e.g., Dempster's rule of combination is used to combine

¹The powerset is the set of all subsets of Θ including the empty set \emptyset .

different distinct BOEs. Suppose that there are two independent BBAs: \mathbf{m}_1 and \mathbf{m}_2 . The conflict coefficient [1] is defined as

$$K \triangleq \sum_{A_i \cap B_j = \emptyset} m_1(A_i)m_2(B_j). \quad (5)$$

If $K < 1$, then the combined BBA \mathbf{m} can be obtained using Dempster's rule of combination

$$m(A) = \begin{cases} 0, & A = \emptyset \\ \frac{\sum_{A_i \cap B_j = A} m_1(A_i)m_2(B_j)}{1 - \sum_{A_i \cap B_j = \emptyset} m_1(A_i)m_2(B_j)}, & A \neq \emptyset \end{cases} \quad (6)$$

where A_1, \dots, A_k and B_1, \dots, B_l are focal elements of \mathbf{m}_1 and \mathbf{m}_2 , respectively. Note that Dempster's rule of combination is both commutative and associative, i.e., symmetric.

The obtained BBA is in fact the orthogonal sum of the original BBAs. Dempster's rule of combination has been criticized for its counter-intuitive behaviors [9], [32], especially in high conflict cases. Accordingly, many alternative combination rules have emerged. See [7], [33] for details.

III. TRADITIONAL MEASURES OF DISTANCE OF EVIDENCE

How to measure the closeness between two BBAs? This is crucial for performance evaluation, algorithm optimization and other belief functions-based applications. The answer is the distance of evidence. The conflict coefficient K [defined in (5)] in Dempster's rule of combination was the only means to quantify the interaction between BBAs for about two decades (from 1967 to 1990). However, this coefficient K (denoted by d_C in the sequel) may be inappropriate to quantify the closeness between two BBAs as the conflict between two identical BBAs might not equal to 0.

Example 1: Suppose that the FOD is $\Theta = \{\theta_1, \dots, \theta_n\}$. Two BBAs defined on Θ are

$$\begin{aligned} m_1L(\{\theta_1\}) &= \dots = m_1(\{\theta_n\}) = 1/n \text{ and } m_2(\{\theta_1\}) \\ &= \dots = m_2(\{\theta_n\}) = 1/n. \end{aligned}$$

Obviously, they are two identical BBAs and $d_C = 1 - 1/n$. When n becomes large, d_C approximates to its upper bound (i.e., 1). If one considered d_C as a distance, such a result would be somewhat counter-intuitive.

A strict distance metric defined on the set $\mathcal{E} d(\cdot, \cdot) : \mathcal{E} \times \mathcal{E} \rightarrow \mathbb{R}$, $(x, y) \mapsto d(x, y)$ should satisfy the following.

- 1) *Non-Negativity:* $d(x, y) \geq 0$.
- 2) *Nondegeneracy:* $d(x, y) = 0 \Leftrightarrow x = y$.
- 3) *Symmetry:* $d(x, y) = d(y, x)$.
- 4) *Triangle Inequality:* $d(x, y) + d(y, z) \geq d(x, z)$, $\forall z \in \mathcal{E}$.

Obviously, d_C violates the nondegeneracy condition. It is not difficult to verify that d_C only satisfies the non-negativity and symmetry conditions. Therefore, it is not a strict distance metric.

Many other definitions² of distance of evidence were proposed in the past two decades as reported in Jousselme's

²To be rigorous, only those definitions satisfying the four requirements can be called distance. The ones that do not satisfy these four requirements can only be called "dissimilarity" or "closeness" measures. In the sequel, for the convenience, all dissimilarity definitions are called distance when no ambiguity should occur.

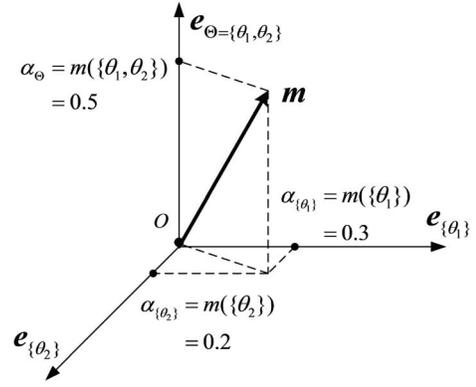


Fig. 1. Geometrical interpretation of a BBA.

survey [22]. Most of them can be considered as being established under the framework of the geometrical interpretation of the DST.

A. Geometric Interpretation of the Theory of Belief Functions

The geometrical interpretation of the DST [25] is as follows.

Suppose that the FOD is Θ with $|\Theta| = n$. Let \mathcal{E}_Θ be the 2^n -dimensional Cartesian space³ spanned by the set of column vectors $\{e_A, A \subseteq \Theta\}$. Each vector \mathbf{v} of \mathcal{E}_Θ could be rewritten as $\mathbf{v} = \sum_{A \subseteq \Theta} \alpha_A \cdot e_A$. Here $\alpha_A \in \mathbb{R}$ can be considered as the coordinate of \mathbf{v} along the direction of e_A .

A BBA \mathbf{m} is a vector of \mathcal{E}_Θ , which should satisfy $\sum_{A \subseteq \Theta} \alpha_A = 1$, $\alpha_\emptyset = 0$, with $\alpha_A \geq 0$ and $\alpha_A \triangleq m(A)$ due to the properties of unity and non-negativity for mass values, as illustrated in (1).

For example, suppose that the FOD $\Theta = \{\theta_1, \theta_2\}$. A BBA \mathbf{m} on Θ is $m(\{\theta_1\}) = 0.3$, $m(\{\theta_2\}) = 0.2$, $m(\{\theta_1, \theta_2\}) = 0.5$. Under the closed-world assumption, \mathbf{m} is illustrated in Fig. 1.

According to the geometrical interpretation of DST, two BBAs \mathbf{m}_1 and \mathbf{m}_2 are two vectors. That is, \mathbf{m}_1 and \mathbf{m}_2 are two "points" in the evidential Cartesian space. In the past 30 years, people use all kinds of distance for Cartesian space like Euclidean distances, Chebyshev distances, Minkowski distances, Manhattan distances, etc., to define the distance between BOEs according to the geometrical interpretation [22]. Note that many available definitions are nonstrict distance metrics [22]. A few typical measures are reviewed in detail in the following. Many other definitions can be found in Jousselme's survey [22].

B. Selected Existing Distance Measures of Evidence

The earliest distance of evidence is the Tessem's distance of betting commitment [23], which is proposed for the evaluation of BBA approximations.

1) *Tessem's Betting Commitment Distance:* The pignistic probability corresponding to a BBA \mathbf{m} is defined by [34]

$$\text{Bet}P(A) \triangleq \sum_{B \subseteq \Theta} \frac{|A \cap B|}{|B|} m(B) \quad (7)$$

³Note that whether the geometric interpretation of the DST satisfies the strict requirements or properties of the geometric space needs further justifications. Here we call it as the evidential Cartesian space.

which is a probabilistic transformation [35] from a BBA for the probabilistic decision making in DST. The betting commitment distance (or Tessem's distance) d_T is computed by [23]

$$d_T(\mathbf{m}_1, \mathbf{m}_2) \triangleq \max_{A \subseteq \Theta} \{ |\text{Bet}P_1(A) - \text{Bet}P_2(A)| \}. \quad (8)$$

It can be reformulated according to the evidential Cartesian space as

$$cd_T(\mathbf{m}_1, \mathbf{m}_2) = \max_{A \subseteq \Theta} \left\{ \begin{array}{c} |\text{Bet}P_1(A_1) - \text{Bet}P_2(A_1)| \\ |\text{Bet}P_1(A_2) - \text{Bet}P_2(A_2)| \\ \vdots \\ |\text{Bet}P_1(A_{2^n}) - \text{Bet}P_2(A_{2^n})| \end{array} \right\} \\ = \max_{A \subseteq \Theta} \{ |\mathbf{BetP}_1' \cdot \mathbf{e}_A - \mathbf{BetP}_2' \cdot \mathbf{e}_A| \} \quad (9)$$

where $\mathbf{BetP}_i = [\text{Bet}P_i(A_1), \text{Bet}P_i(A_2), \dots, \text{Bet}P_i(A_{2^n})]'$, $i = 1, 2$.

d_T is a Chebyshev L_∞ alike distance. From the definition of d_T , we can see that there is a switch from the DST framework to the probability framework when calculating this distance. The inconsistency between different theoretical frameworks leads to the loss of information and some unexpected results, therefore, it is not recommended.

Actually, due to the switch between different frameworks, Tessem's distance is not a strict distance metric [36]. It violates the nondegeneracy condition as shown in Example 2.

Example 2: Suppose that FOD is $\Theta = \{\theta_1, \dots, \theta_n\}$. There are two BBAs \mathbf{m}_1 and \mathbf{m}_2 defined on Θ including $m_1(\{\theta_1\}) = \dots = m_1(\{\theta_n\}) = 1/n$ and $m_2(\Theta) = 1$. Their corresponding pignistic probabilities are both $P(\theta_1) = \dots = P(\theta_n) = 1/n$. Therefore, $d_T(\mathbf{m}_1, \mathbf{m}_2) = 0$, although they are different BBAs. Thus, d_T does not satisfy the nondegeneracy condition. d_T also does not satisfy the triangle inequality and has other drawbacks. See details in [36].

2) *Fuzzy Membership Function-Based Dissimilarity:* First transform BBAs $\mathbf{m}_1(\cdot)$ and $\mathbf{m}_2(\cdot)$ into FMFs⁴: $\mu^{(1)}$ and $\mu^{(2)}$ as for $i = 1, 2$

$$\mu^{(i)} = [\mu^{(i)}(\theta_1), \mu^{(i)}(\theta_2), \dots, \mu^{(i)}(\theta_n)] \\ = [\text{Pl}^{(i)}(\theta_1), \text{Pl}^{(i)}(\theta_2), \dots, \text{Pl}^{(i)}(\theta_n)]. \quad (10)$$

According to the dissimilarity definition between FMFs, d_F is defined as [24]

$$d_F(\mathbf{m}_1, \mathbf{m}_2) = 1 - \frac{\sum_{i=1}^n (\mu^{(1)}(\theta_i) \wedge \mu^{(2)}(\theta_i))}{\sum_{i=1}^n (\mu^{(1)}(\theta_i) \vee \mu^{(2)}(\theta_i))}. \quad (11)$$

In (11), the operator \wedge represents the conjunction (min) and \vee represents the disjunction (max).

⁴The FMF quantifies the membership grade of the element to the fuzzy set. It is a generalization of the characteristic function in classical set and can take its values in the interval $[0, 1]$.

It can be reformulated according to the evidential Cartesian space as

$$d_F(\mathbf{m}_1, \mathbf{m}_2) \\ \triangleq 1 - \frac{\sum_{i=1}^n \min((\mathbf{Int} \cdot \mathbf{m}_1)' \cdot \mathbf{e}_{\theta_i}, (\mathbf{Int} \cdot \mathbf{m}_2)' \cdot \mathbf{e}_{\theta_i})}{\sum_{i=1}^n \max((\mathbf{Int} \cdot \mathbf{m}_1)' \cdot \mathbf{e}_{\theta_i}, (\mathbf{Int} \cdot \mathbf{m}_2)' \cdot \mathbf{e}_{\theta_i})} \quad (12)$$

where \mathbf{Int} is the intersection matrix, whose element is $\text{Int}(A, B) = 1$, if $A \cap B \neq \emptyset$; $\text{Int}(A, B) = 0$, if $A \cap B = \emptyset$. One has $\mathbf{Pl} = \mathbf{Int} \cdot \mathbf{m}$, where \mathbf{Pl} is the corresponding plausibility vector of \mathbf{m} .

d_F in fact indirectly represents the distance between two BBAs using the distance between their corresponding FMFs. Note that d_F is not a strict distance metric. First, d_F does not satisfy the nondegeneracy condition due to the switch from the DST framework to the fuzzy set framework. Given two different BBAs, their corresponding FMFs (singleton plausibility) might be the same as shown in Example 3.

Example 3: Suppose that FOD is $\Theta = \{\theta_1, \theta_2, \theta_3\}$. Two BBAs \mathbf{m}_1 and \mathbf{m}_2 defined on Θ are

$$m_1(\{\theta_1, \theta_3\}) = 0.3, m_1(\{\theta_1, \theta_2\}) = 0.7. \\ m_2(\{\theta_1\}) = 0.3, m_2(\{\theta_1, \theta_2\}) = 0.4, m_2(\Theta) = 0.3.$$

Their corresponding singleton plausibilities are the same

$$\mu^{(1)}(\theta_1) = \text{Pl}_1(\{\theta_1\}) = 1.0, \mu^{(1)}(\theta_2) = \text{Pl}_1(\{\theta_2\}) = 0.7 \\ \mu^{(1)}(\theta_3) = \text{Pl}_1(\{\theta_3\}) = 0.3. \\ \mu^{(2)}(\theta_1) = \text{Pl}_2(\{\theta_1\}) = 1.0, \mu^{(2)}(\theta_2) = \text{Pl}_2(\{\theta_2\}) = 0.7 \\ \mu^{(2)}(\theta_3) = \text{Pl}_2(\{\theta_3\}) = 0.3.$$

Therefore, $d_F(\mathbf{m}_1, \mathbf{m}_2) = 0$, although \mathbf{m}_1 and \mathbf{m}_2 are different BBAs.

3) *Jousselme's Distance:* By borrowing the L_2 Euclidean distance with weighting matrix in Cartesian space, Jousselme's distance [13] is defined as

$$d_J(\mathbf{m}_1, \mathbf{m}_2) \triangleq \sqrt{0.5 \cdot (\mathbf{m}_1 - \mathbf{m}_2)^T \mathbf{Jac} (\mathbf{m}_1 - \mathbf{m}_2)} \quad (13)$$

where the elements $\text{Jac}(A, B)$ of Jaccard's weighting matrix \mathbf{Jac} are defined as

$$\mathbf{Jac}(A, B) = \frac{|A \cap B|}{|A \cup B|}. \quad (14)$$

It has been proved to be a strict distance metric in [37] and has become the most commonly used one so far; however, it might cause some unsatisfactory results as shown in Example 4.

Example 4: Suppose that the FOD is $\Theta = \{\theta_1, \dots, \theta_6\}$. Three groups of BBAs are as follows:

$$\begin{cases} m_1(\{\theta_1\}) = 1 \\ m_2(\{\theta_2\}) = 1 \\ m_3(\{\theta_1, \theta_2\}) = 1 \\ m_4(\{\theta_3, \theta_4\}) = 1 \\ m_5(\{\theta_1, \theta_2, \theta_3\}) = 1 \\ m_6(\{\theta_4, \theta_5, \theta_6\}) = 1. \end{cases}$$

Using Jousselme's distance, one gets $d_J(\mathbf{m}_1, \mathbf{m}_2) = d_J(\mathbf{m}_3, \mathbf{m}_4) = d_J(\mathbf{m}_5, \mathbf{m}_6) = 1$, that is, they all reach the maximum value 1. The six BBAs here are all categorical BBAs.⁵ \mathbf{m}_1 and \mathbf{m}_2 each has a unique singleton focal element. The opinions of \mathbf{m}_1 and \mathbf{m}_2 are totally different, and both of them are specific, i.e., with no ambiguity. The opinions of \mathbf{m}_3 and \mathbf{m}_4 are totally different, however, both of them are not specific and with ambiguity. The BBAs \mathbf{m}_5 and \mathbf{m}_6 carry larger ambiguity. Intuitively, it makes sense that the distance between \mathbf{m}_1 and \mathbf{m}_2 should be larger than the distance between \mathbf{m}_3 and \mathbf{m}_4 ; also, the distance between \mathbf{m}_3 and \mathbf{m}_4 should be larger than the distance between \mathbf{m}_5 and \mathbf{m}_6 . Jousselme's distance does not provide this expected behavior.

Furthermore, Jousselme's distance is relatively insensitive to the change of BBA in some cases as shown in Example 5.

Example 5: Suppose that FOD is $\Theta = \{\theta_1, \dots, \theta_3\}$. Consider the following three BBAs \mathbf{m}_1 , \mathbf{m}_2 , and \mathbf{m}_3 :

$$\begin{aligned} m_1(\{\theta_1\}) &= m_1(\{\theta_2\}) = m_1(\{\theta_3\}) = 1/3 \\ m_2(\{\theta_1\}) &= m_2(\{\theta_2\}) = m_2(\{\theta_3\}) = 0.1, m_2(\Theta) = 0.7 \\ m_3(\{\theta_1\}) &= m_3(\{\theta_2\}) = 0.1, m_3(\{\theta_3\}) = 0.8. \end{aligned}$$

Since both \mathbf{m}_1 and \mathbf{m}_2 have no preference on any singleton $\{\theta_i\}$ and \mathbf{m}_3 commits more belief to $\{\theta_3\}$, it is intuitively expected that the distance between \mathbf{m}_1 and \mathbf{m}_2 should be smaller than that between \mathbf{m}_1 and \mathbf{m}_3 . However, Jousselme's distance leads to $d_J(\mathbf{m}_1, \mathbf{m}_2) = d_J(\mathbf{m}_1, \mathbf{m}_3) = 0.4041$, which shows that d_J does not discriminate them well.

In summary, many existing distance measures of evidence have evident limitations, even for the strict Jousselme's distance metric. For Tessem's distance and FMF-based distance, there exist the switches between different theoretical frameworks. With Tessem's distance, there is a switch from the framework of DST to the framework of the probability theory; with FMF-based distance, there is a switch from the framework of DST to the framework of the fuzzy sets theory. These switches bring the undesired loss of information, which should be avoided. Jousselme's distance borrows the distance metric from the traditional Cartesian space to the evidential Cartesian space. The strictness of the evidential Cartesian space, i.e., the geometrical interpretation of DST needs further verification. Therefore, it is not uncommon to obtain some unsatisfactory results when using Jousselme's distance.

Since traditional distances in DST have limitations (or unsatisfactory behaviors), we propose new strict distance measures of evidence with better behaviors.

IV. DISTANCE OF EVIDENCE USING BELIEF INTERVALS

As aforementioned, the limitations and nonstrictness of some existing distances of evidence are caused by the switches between theoretical frameworks, therefore in our design of the new distances, no such switch is allowed. In Jousselme's distance, there is no switch between different theoretical frameworks, where only the focal elements and the corresponding mass values are used. Given a BBA, the mass value for a proposition (or focal element) A represents the basic

belief assigned to A . Besides the mass value $m(A)$, other values, like $\text{Bel}(A)$ and $\text{Pl}(A)$, are optional. Furthermore, the belief interval $[\text{Bel}(A), \text{Pl}(A)]$ can be used to represent the degree of imprecision of A . Therefore, the belief interval $[\text{Bel}(A), \text{Pl}(A)]$ carries more information of a given proposition A than the mass value $m(A)$, which is a scalar. Therefore, we propose to use the belief interval (with more information) to replace the mass value for achieving better performance.

In DST, besides the BBA (\mathbf{m}), the belief function (**Bel**) and plausibility function (**Pl**), there also exist the doubt function (**Dou**) and the commonality function (**Q**) [1]. Given one function, it can be transformed to any other one of these five functions according to their definitions and the Möbius transformations [1]. That is, any one of the five functions has one-to-one correspondence to the other, therefore, one can also try to jointly use other functions like the commonality and doubt for designing new distance measures. In this paper, we choose the belief interval $[\text{Bel}(A), \text{Pl}(A)]$, $\forall A \subseteq \Theta$, since the belief and plausibility are more familiar to people and more widely used in practice than the doubt and the commonality. Furthermore, $[\text{Bel}(A), \text{Pl}(A)]$ has intuitive physical meaning, i.e., the degree of imprecision for the proposition A .

Suppose that two BBAs \mathbf{m}_1 and \mathbf{m}_2 are defined on $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$. For each focal element $A_i \subseteq \Theta$ ($i = 1, \dots, 2^n - 1$), we can calculate the belief interval of A_i for \mathbf{m}_1 and \mathbf{m}_2 , respectively, which are denoted by $[\text{Bel}_1(A_i), \text{Pl}_1(A_i)]$ and $[\text{Bel}_2(A_i), \text{Pl}_2(A_i)]$. That is, each BBA \mathbf{m}_j ($j = 1, 2$) can also equivalently be modeled by a matrix with the size of $(2^n - 1) \times 2$

$$\begin{bmatrix} [\text{Bel}_j(A_1), \text{Pl}_j(A_1)] \\ \vdots \\ [\text{Bel}_j(A_{2^n-1}), \text{Pl}_j(A_{2^n-1})] \end{bmatrix}.$$

A belief interval can be regarded as a classical interval number⁶ included in $[0, 1]$. Then the above matrix can be regarded as a vector of interval numbers (belief intervals)

$$\begin{aligned} \mathbf{Fe}_j &= \begin{bmatrix} [\text{Bel}_j(A_1), \text{Pl}_j(A_1)] \\ \vdots \\ [\text{Bel}_j(A_{2^n-1}), \text{Pl}_j(A_{2^n-1})] \end{bmatrix} \\ &= \begin{bmatrix} \text{BI}_j(A_1) \\ \vdots \\ \text{BI}_j(A_{2^n-1}) \end{bmatrix}. \end{aligned} \quad (15)$$

Here \mathbf{Fe}_j can be considered as a generalized feature vector describing the BBA \mathbf{m}_j . If we can define the distance between \mathbf{Fe}_1 and \mathbf{Fe}_2 , then the distance between \mathbf{m}_1 and \mathbf{m}_2 is readily obtained. Here \mathbf{Fe}_1 and \mathbf{Fe}_2 are two generalized vectors whose elements are intervals.⁷

⁶An interval number $[a, b]$ with $a \leq b$ is actually an interval with the lower bound a and the upper bound b , where $a, b \in \mathbb{R}$. When $a = b$, an interval number degenerates to a real number.

⁷The \mathbf{Fe}_j can be also considered in the evidential Cartesian-alike space, however, the coordinate of each direction \mathbf{e}_A is a generalized real number, i.e., an interval number.

⁵A categorical BBA is a BBA only has one focal element.

We can borrow the definition of the distance metric for the vectors in Cartesian space to define the distance of evidence here.

- 1) Define the distance between two feature vectors in each dimension.
- 2) Combine the distance value for each dimension into a scalar.

Therefore, in step 1, the distance in each dimension should be defined, i.e., we must define the distance between two interval numbers. Irpino and Verde [26] proposed a Wasserstein distance for interval numbers as briefly introduced below.

Suppose that F and G are the corresponding distribution functions of the random variables f and g , respectively, Wasserstein L_2 metric is defined as [26]

$$d_{\text{Wass}}(F, G) \triangleq \sqrt{\int_0^1 (F^{-1}(t) - G^{-1}(t))^2 dt}. \quad (16)$$

For a uniform distribution of points, an interval of reals $x_i = [a_i, b_i]$ can be expressed as a function of [26]

$$x_i(t) = [a_i, b_i] = a_i + t(b_i - a_i), \forall 0 \leq t \leq 1. \quad (17)$$

If one further considers a description of the interval using its midpoint $(a_i + b_i)/2$ and radius $(b_i - a_i)/2$, x_i can be rewritten as

$$x_i(t) = \frac{a_i + b_i}{2} + \frac{b_i - a_i}{2}(2t - 1), \forall 0 \leq t \leq 1. \quad (18)$$

Then, Euclidean distance between homologous points of two intervals $x_1 = [a_1, b_1]$ and $x_2 = [a_2, b_2]$ is defined as [26]

$$\begin{aligned} d_{\text{BI}}([a_1, b_1], [a_2, b_2]) &= d_{\text{Wass}}(x_1, x_2) \\ &= \sqrt{\int_0^1 [x_1(t) - x_2(t)]^2 dt} \\ &= \sqrt{\int_0^1 \left[\frac{a_1 + b_1}{2} - \frac{a_2 + b_2}{2} + \left(\frac{b_1 - a_1}{2} - \frac{b_2 - a_2}{2} \right) (2t - 1) \right]^2 dt} \\ &= \sqrt{\left[\frac{a_1 + b_1}{2} - \frac{a_2 + b_2}{2} \right]^2 + \frac{1}{3} \left[\frac{b_1 - a_1}{2} - \frac{b_2 - a_2}{2} \right]^2}. \end{aligned} \quad (19)$$

Note that there are also other types of distance between interval numbers [26]. We choose the Wasserstein distance in (19) to calculate the distance between belief intervals, because it is a strict distance metric, which is very crucial for defining distance measures of evidence. Furthermore, it has a simple form, and is easy to compute.

According to (19), the distance between two feature vectors \mathbf{Fe}_1 and \mathbf{Fe}_2 in terms of each dimension i ($i = 1, \dots, 2^n - 1$), i.e., the distance between two belief intervals $\text{BI}_1(A_i) : [\text{Bel}_1(A_i), \text{Pl}_1(A_i)]$ and $\text{BI}_2(A_i) : [\text{Bel}_2(A_i), \text{Pl}_2(A_i)]$ can be obtained. $d_{\text{BI}}(\text{BI}_1(A_i), \text{BI}_2(A_i))$ can be regarded as the distance between \mathbf{m}_1 and \mathbf{m}_2 when considering the focal element A_i only.

Therefore, we can obtain in total $2^n - 1$ belief interval distance values for all $A_i \subseteq \Theta$.

In step 2, we combine all the $2^n - 1$ distance values into one scalar, i.e., to get the total distance between \mathbf{Fe}_1 and \mathbf{Fe}_2 .

In Cartesian space, if we try to measure the distance between two points, we also calculate the dissimilarity between each dimension of the two points, and then use some way to combine the dissimilarity values of different dimensions to a scalar, i.e., the distance value. Euclidean family and Chebyshev family are two commonly used ways to generate such a scalar in the Cartesian space. We can borrow this idea to generate a scalar from the above mentioned $2^n - 1$ focal elements' corresponding dissimilarity values. Therefore, two commonly used distance definitions—the Euclidean family and the Chebyshev family—are used to combine the distance values of all dimensions into a scalar, i.e., the distance value. Two new distances of evidence are presented next.

A. Euclidean-Family Belief Interval-Based Distance d_{BI}^E

Given two BBAs \mathbf{m}_1 and \mathbf{m}_2 , our proposed Euclidean-family belief interval-based distance is a combination of each focal element's belief interval distance value. To be specific, it is a normalized root squared summation of the distance value between belief intervals in each dimension (focal element), as

$$d_{\text{BI}}^E(\mathbf{m}_1, \mathbf{m}_2) \triangleq \sqrt{N_c \cdot \sum_{i=1}^{2^n-1} [d_{\text{BI}}(\text{BI}_1(A_i), \text{BI}_2(A_i))]^2}. \quad (20)$$

Here N_c denotes the normalization factor to make $d_{\text{BI}}^E \in [0, 1]$. Equation (20) can be rewritten as

$$\begin{aligned} d_{\text{BI}}^E(\mathbf{m}_1, \mathbf{m}_2) &\triangleq \sqrt{N_c \cdot \mathbf{d}_{\text{BI}} \cdot \mathbf{I}^{(2^n-1)} \cdot \mathbf{d}_{\text{BI}}^T} \\ &= \sqrt{N_c \cdot \mathbf{d}_{\text{BI}} \cdot \mathbf{d}_{\text{BI}}^T} \end{aligned} \quad (21)$$

where $\mathbf{I}^{(2^n-1)}$ is an identity matrix with rank $2^n - 1$, and

$$\mathbf{d}_{\text{BI}} = \begin{bmatrix} d_{\text{BI}}(\text{BI}_1(A_1), \text{BI}_2(A_1)) \\ \vdots \\ d_{\text{BI}}(\text{BI}_1(A_{2^n-1}), \text{BI}_2(A_{2^n-1})) \end{bmatrix}.$$

Proposition 1: The normalization factor for Euclidean-family belief interval-based distance d_{BI}^E is $N_c = 1/2^{n-1}$.

Proof: Suppose that the FOD is $\{\theta_1, \theta_2, \dots, \theta_n\}$. \mathbf{m}_1 and \mathbf{m}_2 are two BOEs. $m_1(\{\theta_l\}) = 1, l \in \{1, \dots, n\}$ is a categorical BBA, which represents the most certain case, i.e., there is no uncertainty when assign the belief to the singleton proposition $\{\theta_l\}$. The two BBAs

$$m_1(\{\theta_i\}) = 1, m_2(\{\theta_j\}) = 1, \forall i \neq j, \quad i, j \in \{1, \dots, n\} \quad (22)$$

are two different and the most certain cases. They have no common part, i.e., they fully support different singletons, therefore, the dissimilarity (distance) between them reaches the maximum.

Assume that A is a focal element.

When $|A| = 1$, only two belief intervals have distance value d_{BI} of 1 (i.e., $d_{\text{BI}}(\text{BI}_1(\theta_i), \text{BI}_2(\theta_i)) = 1$ and $d_{\text{BI}}(\text{BI}_1(\theta_j), \text{BI}_2(\theta_j)) = 1$). The other values are 0.

When $|A| > 1$, $d_{\text{BI}} = 1$ for those focal elements including θ_i or θ_j (but not including both θ_i and θ_j) are 1. $d_{\text{BI}} = 0$ for the rest.

To be specific,

when $|A| = 2$, $d_{\text{BI}} = 1$ only for $2 \times C_{n-2}^1$ focal elements.⁸
 $d_{\text{BI}} = 0$ for the rest;

when $|A| = 3$, $d_{\text{BI}} = 1$ only for $2 \times C_{n-2}^2$ focal elements.
 $d_{\text{BI}} = 0$ for the rest;

⋮

when $|A| = n-1$, $d_{\text{BI}} = 1$ only for $2 \times C_{n-2}^{n-2}$ focal elements.
 $d_{\text{BI}} = 0$ for the rest;

when $|A| = n$, the d_{BI} value of unique focal element,
i.e., total set (Θ) is 0.

So, the summation S_c of all $(d_{\text{BI}})^2$ is

$$\begin{aligned} S_c &= 2 \times 1 + 2 \times C_{n-2}^1 + 2 \times C_{n-2}^2 + \cdots \\ &\quad + 2 \times C_{n-2}^{n-2} + 0 \\ &= 2 \times (C_{n-2}^0 + C_{n-2}^1 + C_{n-2}^2 + \cdots + C_{n-2}^{n-2}) \\ &= 2 \times 2^{n-2} = 2^{n-1}. \end{aligned} \quad (23)$$

So, the normalization factor $N_c = 1/S_c = 1/2^{n-1}$. ■

B. Chebyshev-Family Belief Interval-Based Distance d_{BI}^C

Given two BBAs \mathbf{m}_1 and \mathbf{m}_2 , our proposed Chebyshev-family belief interval-based distance is the maximum of all belief interval distance values

$$d_{\text{BI}}^C(\mathbf{m}_1, \mathbf{m}_2) \triangleq \max_{A_i \in \Theta, i=1, \dots, 2^n-1} \{d_{\text{BI}}(\text{BI}_1(A_i), \text{BI}_2(A_i))\}. \quad (24)$$

Actually, we use the distance of belief intervals for focal elements instead of their mass assignments to define the distances of evidence when compared with traditional definitions.

Proposition 2: Euclidean-family belief interval-based distance d_{BI}^E and Chebyshev-family belief interval-based distance d_{BI}^C are strict distance metrics.

Proof: d_{BI}^E and d_{BI}^C are defined over belief intervals. Given a BBA $(m(A_i), i = 1, \dots, 2^n-1)$, we can generate a set of belief intervals ($[\text{Bel}(A_i), \text{Pl}(A_i)]$). On the other hand, given a set of belief intervals ($[\text{Bel}(A_i), \text{Pl}(A_i)]$), according to the Möbius transformation, we can generate a unique BBA $(m(A_i), i = 1, \dots, 2^n-1)$ from $\text{Pl}(A_i), i = 1, \dots, 2^n-1$ or $\text{Bel}(A_i), i = 1, \dots, 2^n-1$. As we know [1], there is a one-to-one mapping between a set of belief intervals ($[\text{Bel}(A_i), \text{Pl}(A_i)]$) and a BBA $(m(A_i), i = 1, \dots, 2^n-1)$.

According to (20), (21), and (24), it is easy to verify that d_{BI}^E and d_{BI}^C satisfy non-negativity, nondegeneracy, and symmetry conditions. We need to prove the property of triangle inequality of d_{BI}^E .

Suppose that there are three BBAs $\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3$ defined on the same FOD with size of n . Because d_{BI} defined in (19) is a strict distance metric, so, for each A_i ($i = 1, \dots, s, s = 2^n-1$) there exists

$$\begin{aligned} d_{\text{BI}}^E(m_1(A_i), m_2(A_i)) + d_{\text{BI}}^E(m_2(A_i), m_3(A_i)) \\ \geq d_{\text{BI}}^E(m_1(A_i), m_3(A_i)). \end{aligned}$$

⁸Choose one element θ_k out of the $\Theta' = \Theta - \{\theta_i, \theta_j\}$ ($|\Theta'| = n-2$). Then, together with θ_i and θ_j , respectively, to constitute focal element $\{\theta_k, \theta_i\}$ and $\{\theta_k, \theta_j\}$, respectively. So, the number of focal elements with d_{BI} values of 1 is $2 \times C_{n-2}^1$. Similarly, we can obtain the values in other cases for $A > 1$.

Suppose that

$$\begin{aligned} d_{\text{BI}}^E(m_1(A_i), m_2(A_i)) &= a_i; \quad d_{\text{BI}}^E(m_2(A_i), m_3(A_i)) = b_i \\ d_{\text{BI}}^E(m_1(A_i), m_3(A_i)) &= c_i. \end{aligned}$$

One has

$$\begin{aligned} a_i + b_i \geq c_i &\Rightarrow (a_i + b_i)^2 \geq c_i^2 \\ &\Rightarrow a_i^2 + b_i^2 + 2a_i b_i \geq c_i^2 \\ &\Rightarrow \sum_{i=1}^s a_i^2 + \sum_{i=1}^s b_i^2 + 2 \sum_{i=1}^s a_i b_i \geq \sum_{i=1}^s c_i^2. \end{aligned} \quad (25)$$

According to Cauchy–Schwarz inequality

$$\sqrt{\sum_{i=1}^s a_i^2 \sum_{i=1}^s b_i^2} \geq \sum_{i=1}^s a_i b_i. \quad (26)$$

So

$$\begin{aligned} \sum_{i=1}^s a_i^2 + \sum_{i=1}^s b_i^2 + 2 \sqrt{\sum_{i=1}^s a_i^2 \sum_{i=1}^s b_i^2} \\ \geq \sum_{i=1}^s a_i^2 + \sum_{i=1}^s b_i^2 + 2 \sum_{i=1}^s a_i b_i \geq \sum_{i=1}^s c_i^2 \\ \Rightarrow \sum_{i=1}^s a_i^2 + \sum_{i=1}^s b_i^2 + 2 \sqrt{\sum_{i=1}^s a_i^2 \sum_{i=1}^s b_i^2} \geq \sum_{i=1}^s c_i^2. \end{aligned} \quad (27)$$

Therefore

$$\begin{aligned} \sum_{i=1}^s a_i^2 + \sum_{i=1}^s b_i^2 + 2 \sqrt{\sum_{i=1}^s a_i^2 \sum_{i=1}^s b_i^2} \\ = \left(\sqrt{\sum_{i=1}^s a_i^2} + \sqrt{\sum_{i=1}^s b_i^2} \right)^2 \\ = (d_{\text{BI}}^E(\mathbf{m}_1, \mathbf{m}_2) + d_{\text{BI}}^E(\mathbf{m}_2, \mathbf{m}_3))^2 \\ \Rightarrow (d_{\text{BI}}^E(\mathbf{m}_1, \mathbf{m}_2) + d_{\text{BI}}^E(\mathbf{m}_2, \mathbf{m}_3))^2 \geq (d_{\text{BI}}^E(\mathbf{m}_1, \mathbf{m}_3))^2 \\ \Rightarrow d_{\text{BI}}^E(\mathbf{m}_1, \mathbf{m}_2) + d_{\text{BI}}^E(\mathbf{m}_2, \mathbf{m}_3) \geq d_{\text{BI}}^E(\mathbf{m}_1, \mathbf{m}_3). \end{aligned} \quad (28)$$

So, the triangle inequality for d_{BI}^E is satisfied.

For d_{BI}^C , we have

$$\begin{aligned} d_{\text{BI}}^C(\mathbf{m}_1, \mathbf{m}_2) + d_{\text{BI}}^C(\mathbf{m}_2, \mathbf{m}_3) &= \max_{i=1, \dots, s} a_i + \max_{i=1, \dots, s} b_i \\ d_{\text{BI}}^C(\mathbf{m}_1, \mathbf{m}_3) &= \max_{i=1, \dots, s} c_i = a_k + b_k, \quad k = \arg \max_{i=1, \dots, s} c_i. \end{aligned} \quad (29)$$

There exists

$$a_k + b_k \leq \max_{i=1, \dots, s} a_i + \max_{i=1, \dots, s} b_i = d_{\text{BI}}^C(\mathbf{m}_1, \mathbf{m}_2) + d_{\text{BI}}^C(\mathbf{m}_2, \mathbf{m}_3) \quad (30)$$

that is $d_{\text{BI}}^C(\mathbf{m}_1, \mathbf{m}_2) + d_{\text{BI}}^C(\mathbf{m}_2, \mathbf{m}_3) \geq d_{\text{BI}}^C(\mathbf{m}_1, \mathbf{m}_3)$. Consequently, d_{BI}^C satisfies triangle inequality.

In summary, d_{BI}^E and d_{BI}^C are strict distance metrics. ■

In the traditional geometric interpretation of DST introduced in Section III, the coordinates of different bases are represented by mass values (real numbers), while for our new proposed distances, the coordinates are represented by belief intervals (interval numbers). Therefore, our new distances are under a generalized geometric interpretation of evidence theory.

C. Illustrative Example

Example 6: Suppose that the FOD is $\Theta = \{\theta_1, \theta_2, \theta_3\}$. Two BBAs $\mathbf{m}_1, \mathbf{m}_2$ over the FOD are

$$\begin{aligned} m_1(\{\theta_1\}) &= 0.1, m_1(\{\theta_2\}) = 0.1, m_1(\{\theta_3\}) = 0.05 \\ m_1(\{\theta_1, \theta_2\}) &= 0.1, m_1(\{\theta_1, \theta_3\}) = 0.05 \\ m_1(\{\theta_2, \theta_3\}) &= 0.1, m_1(\Theta) = 0.5 \\ m_2(\{\theta_1\}) &= 0.2, m_2(\{\theta_2\}) = 0.3, m_2(\{\theta_3\}) = 0.1 \\ m_2(\{\theta_1, \theta_2\}) &= 0.05, m_2(\{\theta_1, \theta_3\}) = 0.1 \\ m_2(\{\theta_2, \theta_3\}) &= 0.05, m_2(\Theta) = 0.2. \end{aligned}$$

First, the belief intervals are calculated for each focal element of \mathbf{m}_1 and \mathbf{m}_2 , respectively

$$\begin{aligned} \text{BI}_1(\{\theta_1\}) &: [0.10, 0.75] \\ \text{BI}_1(\{\theta_2\}) &: [0.10, 0.80] \\ \text{BI}_1(\{\theta_3\}) &: [0.05, 0.70] \\ \text{BI}_1(\{\theta_1, \theta_2\}) &: [0.30, 0.95] \\ \text{BI}_1(\{\theta_1, \theta_3\}) &: [0.20, 0.90] \\ \text{BI}_1(\{\theta_2, \theta_3\}) &: [0.25, 0.90] \\ \text{BI}_1(\Theta) &: [1.00, 1.00] \\ \text{BI}_2(\{\theta_1\}) &: [0.20, 0.55] \\ \text{BI}_2(\{\theta_2\}) &: [0.30, 0.60] \\ \text{BI}_2(\{\theta_3\}) &: [0.10, 0.45] \\ \text{BI}_2(\{\theta_1, \theta_2\}) &: [0.55, 0.90] \\ \text{BI}_2(\{\theta_1, \theta_3\}) &: [0.40, 0.70] \\ \text{BI}_2(\{\theta_2, \theta_3\}) &: [0.45, 0.80] \\ \text{BI}_2(\Theta) &: [1.00, 1.00]. \end{aligned}$$

Second, use (19) to compute the distance between belief intervals of each corresponding focal element in \mathbf{m}_1 and \mathbf{m}_2

$$d_{\text{BI}} = \begin{bmatrix} d_{\text{BI}}(\text{BI}_1(\{\theta_1\}), \text{BI}_2(\{\theta_1\})) \\ d_{\text{BI}}(\text{BI}_1(\{\theta_2\}), \text{BI}_2(\{\theta_2\})) \\ d_{\text{BI}}(\text{BI}_1(\{\theta_3\}), \text{BI}_2(\{\theta_3\})) \\ d_{\text{BI}}(\text{BI}_1(\{\theta_1, \theta_2\}), \text{BI}_2(\{\theta_1, \theta_2\})) \\ d_{\text{BI}}(\text{BI}_1(\{\theta_1, \theta_3\}), \text{BI}_2(\{\theta_1, \theta_3\})) \\ d_{\text{BI}}(\text{BI}_1(\{\theta_2, \theta_3\}), \text{BI}_2(\{\theta_2, \theta_3\})) \\ d_{\text{BI}}(\text{BI}_1(\Theta), \text{BI}_2(\Theta)) \end{bmatrix} = \begin{bmatrix} 0.1000 \\ 0.1155 \\ 0.1323 \\ 0.1323 \\ 0.1155 \\ 0.1000 \\ 0.0000 \end{bmatrix}.$$

Then, according to (20), $d_{\text{BI}}^E(\mathbf{m}_1, \mathbf{m}_2)$ is computed by

$$\begin{aligned} d_{\text{BI}}^E(\mathbf{m}_1, \mathbf{m}_2) &= \sqrt{2^{3-1} \times \left(0.1000^2 + 0.1155^2 + 0.1323^2 + 0.1323^2 + 0.1155^2 + 0.1000^2 + 0^2 \right)} \\ &= 0.1429. \end{aligned}$$

According to (24), $d_{\text{BI}}^C(\mathbf{m}_1, \mathbf{m}_2)$ is computed by

$$d_{\text{BI}}^C(\mathbf{m}_1, \mathbf{m}_2) = \max \left\{ \begin{matrix} 0.1000, & 0.1155, & 0.1323, \\ 0.1323, & 0.1155, & 0.1000, & 0 \end{matrix} \right\} = 0.1323.$$

D. On Distance Bounds

Here, the distance bounds are analyzed. In Antonucci's work [38], a lower bound and an upper bound of a distance

TABLE I
ALGORITHM 1: RANDOM BBA GENERATION—UNIFORM SAMPLING FROM ALL FOCAL ELEMENTS

Input: Θ : Frame of discernment;
N_{max} : Maximum number of focal elements
Output: \mathbf{m} : BBA
Generate $\mathcal{P}(\Theta)$, which is the power set of Θ ;
FOR each $1 \leq i \leq \mathcal{P}(\Theta) $ do
Generate a value according to the Gamma distribution $\mathcal{G}(1, 1) \rightarrow m_i$,
END
Normalize the vector $\mathbf{m} = [m_1, \dots, m_{ \mathcal{P}(\Theta) }] \rightarrow \mathbf{m}'$;
$\mathbf{m}(A_i) = m'_i$;

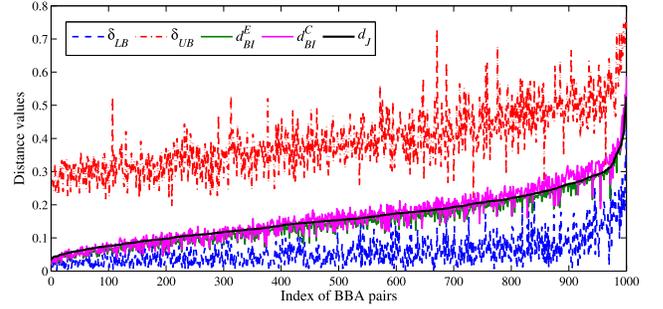


Fig. 2. Comparisons between bounds, d_J , d_{BI}^E , and d_{BI}^C .

of evidence were proposed based on the distance of consistent probabilities. For a BBA \mathbf{m} , its consistent set of probability mass functions (PMFs) is

$$K_{\mathbf{m}} = \left\{ P \mid \begin{matrix} \sum_{\theta \in \Theta} P(\theta) = 1 \\ \sum_{\theta \in A} P(\theta) \geq \mathbf{Bel}(A), \forall A \in 2^\Theta \end{matrix} \right\} \quad (31)$$

where P is a consistent PMF. Given two PMFs P_1 and P_2 , their Manhattan distance is

$$\delta(P_1, P_2) \triangleq \frac{1}{2} \cdot \sum_{\theta \in \Theta} |P_1(\theta) - P_2(\theta)|. \quad (32)$$

Given two BBAs \mathbf{m}_1 and \mathbf{m}_2 , the lower bound $\underline{\delta}$ and upper bound $\bar{\delta}$ are defined as

$$\begin{cases} \underline{\delta}(\mathbf{m}_1, \mathbf{m}_2) = \min_{P_1 \in K_{\mathbf{m}_1}, P_2 \in K_{\mathbf{m}_2}} \delta(P_1, P_2) \\ \bar{\delta}(\mathbf{m}_1, \mathbf{m}_2) = \max_{P_1 \in K_{\mathbf{m}_1}, P_2 \in K_{\mathbf{m}_2}} \delta(P_1, P_2). \end{cases} \quad (33)$$

We calculate d_{BI}^E , d_{BI}^C and the strict distance measure d_J together with the upper and lower bounds to check whether these measures are beyond the bounds or not. We set $|\Theta| = 3$ and randomly generate 1000 BBA pairs according to the BBA generation algorithm [39] in Table I.

The results are shown in Figs. 2 and 3 (zoom in around lower bound). Results are sorted by increasing values of d_J . It is experimentally shown that d_J and our proposed d_{BI}^E and d_{BI}^C are not beyond the lower and upper bounds as shown in Figs. 2 and 3 in this simulation.

In the next section, experiments and simulations are provided to show the rationalities of our proposed distance measures of evidence based on the comparisons with available measures.

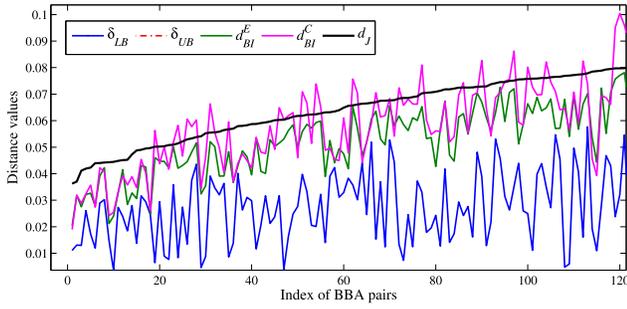


Fig. 3. Comparisons between bounds, d_J , d_{BI}^E , and d_{BI}^C (zoom in around the lower bound).

TABLE II
BBA m_1

Focal element	Mass assignment
$\{\theta_1\}$	0.1
$\{\theta_2\}$	0.8
$\{\theta_3\}$	0.1
$\{\theta_1, \theta_2\}$	0
$\{\theta_2, \theta_3\}$	0
$\{\theta_1, \theta_3\}$	0
$\{\theta_1, \theta_2, \theta_3\}$	0

TABLE III
BBAs m_i , $i = 2, \dots, 7$

Focal el. \ BBAs	m_2	m_3	m_4	m_5	m_6	m_7
θ_1	0.8	0	0	0	0	0
θ_2	0	0.8	0	0	0	0
θ_3	0	0	0.8	0	0	0
$\theta_1 \cup \theta_2$	0	0	0	0.8	0	0
$\theta_2 \cup \theta_3$	0	0	0	0	0.8	0
$\theta_1 \cup \theta_3$	0	0	0	0	0	0.8
$\theta_1 \cup \theta_2 \cup \theta_3$	0.2	0.2	0.2	0.2	0.2	0.2

V. EXPERIMENTS, SIMULATIONS, AND APPLICATIONS

To verify the rationality of the proposed distances, numerical examples, simulations and the applications related to BBA approximations and MCDM are provided.

In each example below, d_J , d_T , d_F , d_C , d_{BI}^E , and d_{BI}^C are compared.⁹

A. Example 7

Suppose that the FOD is $\Theta = \{\theta_1, \theta_2, \theta_3\}$. m_1 has relatively large mass value for $\{\theta_2\}$ as shown in Table II. Therefore, intuitively, for m_i , $i = 2, \dots, 7$ listed in Table III, if the mass assignment for $\{\theta_2\}$ is relative large, the distance between m_1 and m_i intuitively should be relatively small. For m_5 and m_6 , the mass of focal elements containing $\{\theta_2\}$ (i.e., $\{\theta_1, \theta_2\}$ and $\{\theta_2, \theta_3\}$) is 0.8. It makes more sense if the distance value with respect to m_5 and m_6 decreases.

Calculate the distance between m_1 and m_i , $i = 2, \dots, 7$ using different distance definitions as illustrated in Fig. 4. All the distance measures perform similarly in all seven cases and agree with the expected behavior as we can see in Fig. 4.

Examples 8–12 drawn from [13] are used for comparing our proposed measures and available ones.

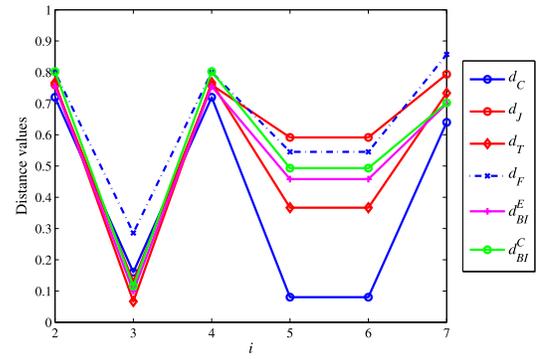


Fig. 4. Distance between m_1 and m_i , $i = 2, \dots, 7$.

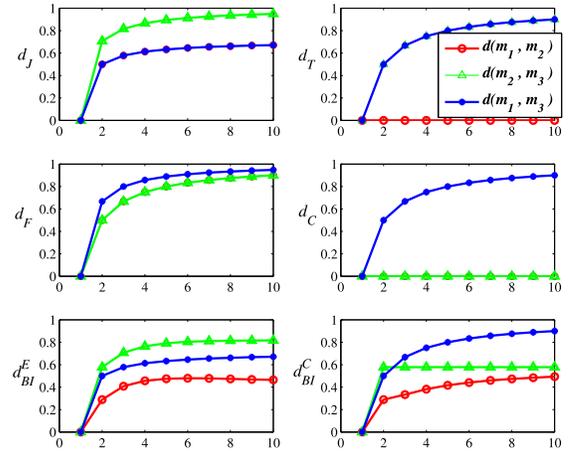


Fig. 5. Dissimilarities between m_1 , m_2 , and m_3 for Example 8.

B. Example 8

Suppose that three BBAs m_1 , m_2 , and m_3 are defined on the FOD $\Theta = \{\theta_1, \dots, \theta_n\}$ as follows:

$$m_1(\{\theta_1\}) = m_1(\{\theta_2\}) = \dots = m_1(\{\theta_n\}) = 1/n$$

$$m_2(\Theta) = 1$$

$$m_3(\{\theta_k\}) = 1, \text{ for some } k \in \{1, \dots, n\}.$$

The change of the distance values with the increase of the size n of FOD are illustrated in Fig. 5.

d_T provides undesired result, i.e., with the increase of n , there always exists $d_T(m_1, m_2) = 0$. d_C cannot discriminate m_1 and m_2 , and also m_2 and m_3 .

In this example, m_1 is a Bayesian BBA, which only has singleton focal elements; m_2 is a vacuous BBA, which only has the total set Θ as the unique focal element; and m_3 is a categorical BBA with one singleton focal element, which is absolutely confident in $\{\theta_k\}$.

m_1 represents the case with only discord and with zero non-specificity; m_3 represents the crispest case; and m_2 represents the most ambiguous case. So, the distance between m_2 and m_3 represents the dissimilarity between the most ambiguous case and the crispest case; the distance between m_1 and m_3 represents the dissimilarity between the case with zero non-specificity and the crispest case; the distance between m_1 and m_2 represents the dissimilarity between the case with zero non-specificity and the most ambiguous case.

⁹ d_C corresponds to the conflict coefficient K defined in (4).

TABLE IV
EXAMPLE 9: RESULTS BASED ON DIFFERENT DISTANCES OF EVIDENCE

Distance	d_J	d_T	d_F	d_C	d_{BI}^E	d_{BI}^C
$d(\mathbf{m}_1, \mathbf{m}_2)$	0.4041	0	0.5833	0.2000	0.2858	0.2333
$d(\mathbf{m}_1, \mathbf{m}_3)$	0.4041	0.4667	0.6364	0.6667	0.4041	0.4667

Therefore, intuitively, the distance between \mathbf{m}_2 and \mathbf{m}_3 should be the largest one. As we can see in Fig. 5, $d_{BI}^E(\mathbf{m}_2, \mathbf{m}_3)$ and $d_J(\mathbf{m}_2, \mathbf{m}_3)$ provide satisfactory results, that is

$$d_{BI}^E(\mathbf{m}_2, \mathbf{m}_3) = \max_{\substack{i,j \in \{1,2,3\} \\ i \neq j}} d_{BI}^E(\mathbf{m}_i, \mathbf{m}_j).$$

From the decision standpoint, \mathbf{m}_1 has no inclination to any choice θ_i ; \mathbf{m}_2 also has no inclination to any choice θ_i ; \mathbf{m}_3 has a clear inclination to the choice θ_k . Therefore, intuitively, the dissimilarity between \mathbf{m}_1 and \mathbf{m}_2 should be smaller than that between \mathbf{m}_1 and \mathbf{m}_3 . As shown in Fig. 5, our proposed d_{BI}^E , d_{BI}^C , and d_F provide satisfactory results according to this standpoint, i.e., $d_{BI}^E(\mathbf{m}_1, \mathbf{m}_2) < d_{BI}^E(\mathbf{m}_1, \mathbf{m}_3)$, $d_{BI}^C(\mathbf{m}_1, \mathbf{m}_2) < d_{BI}^C(\mathbf{m}_1, \mathbf{m}_3)$, and $d_F(\mathbf{m}_1, \mathbf{m}_2) < d_F(\mathbf{m}_1, \mathbf{m}_3)$.

As we observed, d_J cannot discriminate this since $d_J(\mathbf{m}_1, \mathbf{m}_2) = d_J(\mathbf{m}_1, \mathbf{m}_3)$ with the increase of n . This is because one has

$$d_J(\mathbf{m}_1, \mathbf{m}_2) = d_J(\mathbf{m}_1, \mathbf{m}_3) = \sqrt{\frac{1}{2} \left(1 - \frac{1}{n}\right)}$$

according to Jousselme's distance defined in (13).

Based on the analyses above, d_{BI}^E provides rational behaviors in this example.

C. Example 9 (Example 5 Revisited)

The values of the different distances between \mathbf{m}_1 and \mathbf{m}_2 , and between \mathbf{m}_1 and \mathbf{m}_3 are given in Table IV.

As aforementioned, both \mathbf{m}_1 and \mathbf{m}_2 have no preference on any singleton $\{\theta_i\}$ and \mathbf{m}_3 commits more belief to $\{\theta_3\}$, therefore, it is intuitively expected that the distance between \mathbf{m}_1 and \mathbf{m}_2 should be smaller than that between \mathbf{m}_1 and \mathbf{m}_3 .

Using Jousselme's distance, one obtains $d_J(\mathbf{m}_1, \mathbf{m}_2) = d_J(\mathbf{m}_1, \mathbf{m}_3) = 0.4041$ which is unsatisfactory for such a case. Table IV shows that when using d_T , d_C , d_F , d_{BI}^E , and d_{BI}^C , one obtains $d(\mathbf{m}_1, \mathbf{m}_2) < d(\mathbf{m}_1, \mathbf{m}_3)$, which is more reasonable. However, Tessem's distance leads to $d_T(\mathbf{m}_1, \mathbf{m}_2) = 0$, and it is counter-intuitive.

D. Example 10

Suppose that the FOD is $\Theta = \{\theta_1, \dots, \theta_{10}\}$. A BBA \mathbf{m}_t defined on Θ is

$$\begin{aligned} m_t(\Theta) &= 0.1, m_t(\{\theta_2, \theta_3, \theta_3\}) = 0.05, m_t(\{\theta_7\}) = 0.05 \\ m_t(A_t) &= 0.8 \end{aligned}$$

where A_t is a varying focal element from $\{\theta_1\}$ to Θ . One singleton $\{\theta_i\}$ is added at each step. All the A_t , $\forall t = 1, \dots, 10$ are as shown in Table V. The second BBA \mathbf{m}^* has only one focal element, and it is defined as

$$m^*(\{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5\}) = 1.$$

TABLE V
EXAMPLE 10: DISTANCE VALUE CHANGES WITH A_t

Step	A_t
1	$\{\theta_1\}$
2	$\{\theta_1, \theta_2\}$
3	$\{\theta_1, \theta_2, \theta_3\}$
4	$\{\theta_1, \theta_2, \theta_3, \theta_4\}$
5	$\{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5\}$
6	$\{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6\}$
\vdots	\vdots
10	$\{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \dots, \theta_{10}\}$

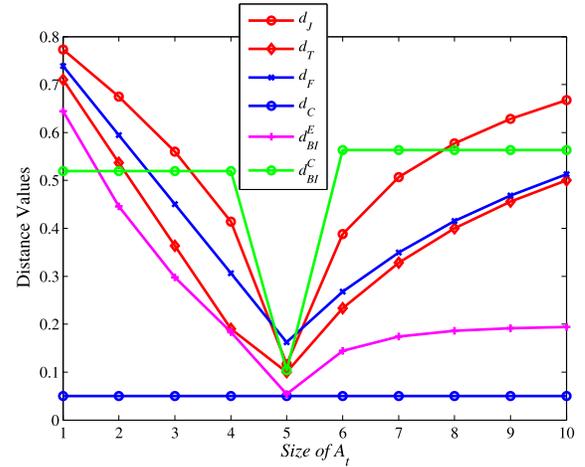


Fig. 6. Distance between \mathbf{m}_t and \mathbf{m}^* for Example 10.

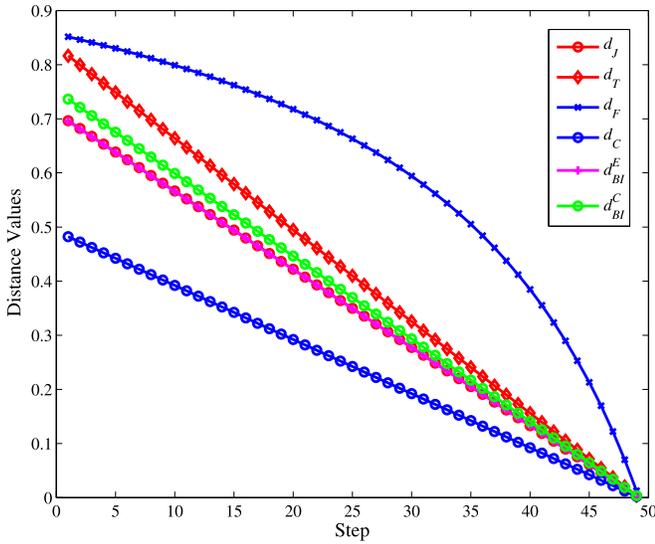
We use different distance measures including d_J , d_T , d_C , d_F , d_{BI}^E , and d_{BI}^C to calculate the distance between \mathbf{m}^* and \mathbf{m}_t . Their behaviors are shown in Fig. 6.

Intuitively, when A_t starts from the focal element $\{\theta_1\}$ to the focal element $\{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5\}$, the distance between \mathbf{m}_t and \mathbf{m}^* should become smaller. When $A_t = \{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5\}$, the distance should reach the minimum value. Then, when the size of A_t becomes larger and departs from $\{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5\}$, the distance value should become larger. As shown in Fig. 6, d_J , d_T , d_F and our proposed d_{BI}^E provide expected behaviors.

Since the conflict between \mathbf{m}_t and \mathbf{m}^* are fixed, i.e., $d_C(\mathbf{m}_t, \mathbf{m}^*) = m_t(\{\theta_7\}) \cdot m^*(\{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5\}) = 0.05$, the value of d_C is fixed to 0.05. Therefore, d_C is not a proper distance. As shown in Fig. 6, our proposed d_{BI}^E performs well; however, d_{BI}^C does not provide a satisfactory behavior. Although d_{BI}^C reaches its minimum value when $A_t = \{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5\}$, it cannot detect the change of A_t when the size of A_t is smaller than 5 or the size of A_t is larger than 5.

E. Example 11

Suppose that the FOD is $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6\}$. In each case of this example, we set a fixed BBA \mathbf{m}_2 , respectively, where $m_2(B) = 1$. B can be considered as a "desired" focal element. Another BBA \mathbf{m}_1 is also set, where $m_1(A) = 1/63, \forall A \subseteq \Theta$. Let \mathbf{m}_1 approximate to \mathbf{m}_2 in some way. To implement this, at each step, we increase $m_1(B)$'s value of $\Delta = 0.02$ and the mass value of other focal elements ($A \neq B, \forall A \subseteq \Theta$) is decreased of $\Delta/62$.


 Fig. 7. Distance between m_1 and m_2 for Example 11—case A.

We also let m_1 go away from m_2 . To implement this, at each step, $m_1(C)$, $C \neq B, \forall C \subseteq \Theta$ has an increase of $\Delta = 0.02$ and the mass value of other focal elements ($A \neq C, \forall A \subseteq \Theta$) has a decrease of $\Delta/62$. Therefore, C can be considered as an “undesired” focal element.

We use different distances between m_1 and m_2 at each step. Their behaviors with varying m_1 are analyzed.

1) *Case A*: Here $B = \{\theta_2\}$, i.e., the desired focal element B is a singleton. With the change of $m_1(B)$, m_1 is gradually close to m_2 . Therefore, if a distance measure becomes smaller with the change of $m_1(B)$, then it behaves as intuitively expected.

The changes of the different distance measures in the above procedure are shown in Fig. 7. All the distance measures used here provide expected behaviors.

2) *Case B*: Here $|B| > 1$, e.g., $B = \{\theta_1, \theta_2\}$ or $B = \{\theta_1, \theta_2, \theta_3\}$. That is, the desired focal element B is a compound focal element. With the change of $m_1(B)$, m_1 is gradually close to m_2 .

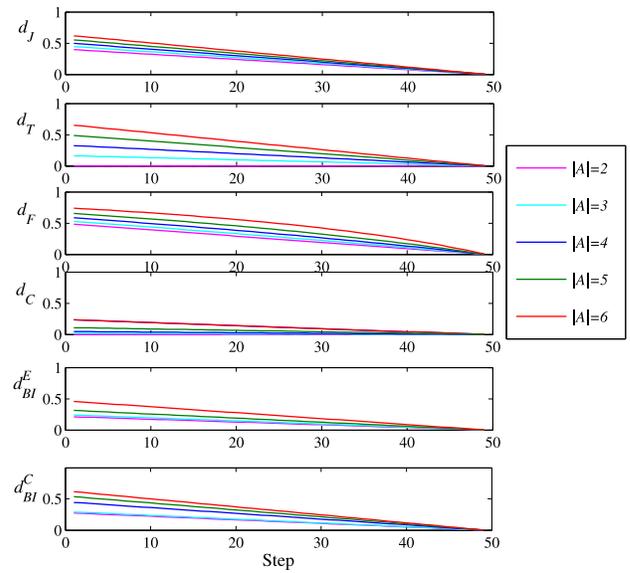
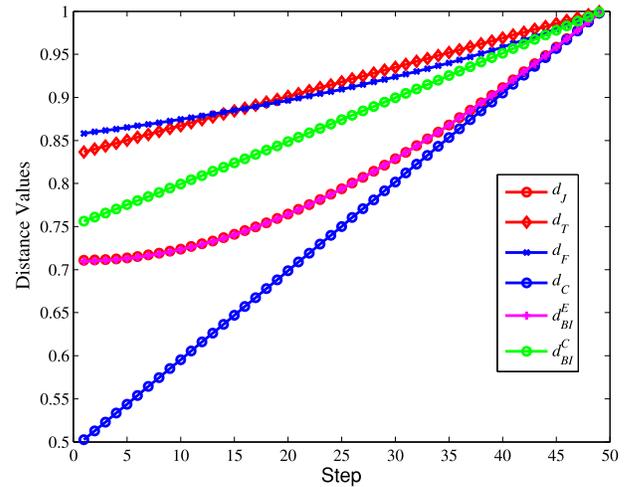
Given different $|B|$, the changes of different distances in the above procedure are shown in Fig. 8, where all distances used here provide expected behaviors when B is a compound focal element.

3) *Case C*: Here $B = \{\theta_4\}$ and $C = \{\theta_5\}$, where the undesired focal element C is a singleton.

With the change of $m_1(C)$, m_1 is gradually away from m_2 . If a distance measure becomes larger with the change of $m_1(C)$, then it behaves reasonably (i.e., as intuitively expected).

The changes of the different distance measures in the above procedure are shown in Fig. 9, where all the distance measures tested here provide expected behaviors.

4) *Case D*: Here $B = \{\theta_6\}$ and the undesired focal element $C = \Theta$. With the change of $m_1(C)$, m_1 is gradually away from m_2 . If a distance measure becomes larger with the change of $m_1(C)$, then it behaves as intuitively expected. Fig. 10 shows the changes of the different distance measures in the above procedure.


 Fig. 8. Distance between m_1 and m_2 for Example 11—case B.

 Fig. 9. Distance between m_1 and m_2 for Example 11—case C.

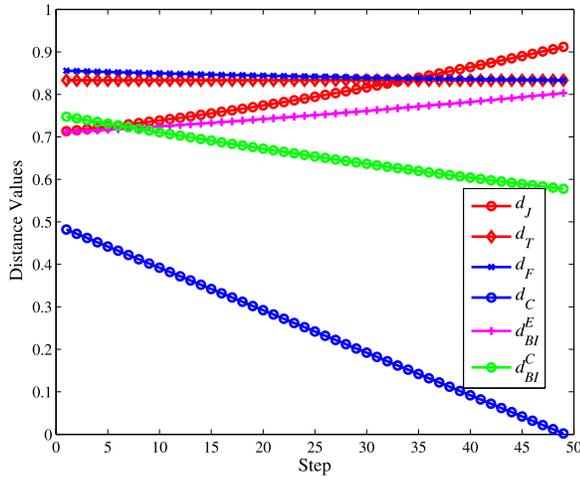
As seen in Fig. 10, only d_{BI}^E and d_J behave as expected. d_T never changes in the whole procedure, because the corresponding pignistic probability never change with the increase of $m_1(\Theta)$. d_C diminishes significantly with the increase of $m_1(\Theta)$, because as the mass assignment is increasing for the total set, the conflict between m_1 and m_2 becomes smaller. Therefore, d_C is only a conflict degree and must not be used as a proper distance measure.

F. Example 12

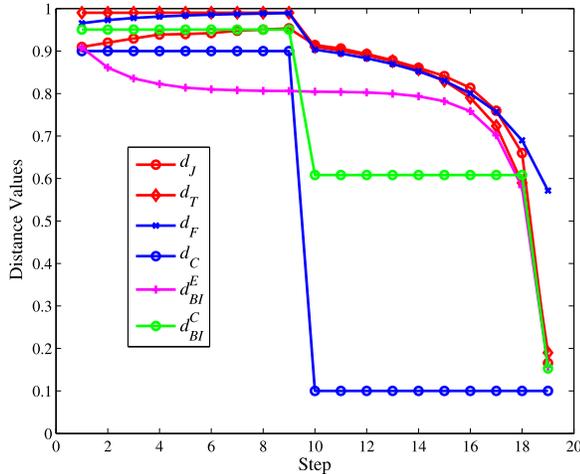
Suppose that the FOD is $\Theta = \{\theta_1, \dots, \theta_{10}\}$. A BBA m_t defined on Θ is

$$m_t(\Theta) = 0.1, m_t(\{\theta_2, \theta_3, \theta_3\}) = 0.05, m_t(\{\theta_7\}) = 0.05 \\ m_t(B_t) = 0.8$$

where B_t is a varying focal element from $\{\theta_1\}$ to Θ . One singleton θ_i is added at each step (steps 1–10). $B_t, \forall t = 1, \dots, 10$ equals to A_t as shown in Table V in Example 10.

Fig. 10. Distance between m_1 and m_2 for Example 11—case D.TABLE VI
EXAMPLE 10: DISTANCE VALUE CHANGES WITH B_t

Step	B_t
1-10	A_t
11	$\{\theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_7, \theta_8, \theta_9, \theta_{10}\}$
12	$\{\theta_3, \theta_4, \theta_5, \theta_6, \theta_7, \theta_8, \theta_9, \theta_{10}\}$
13	$\{\theta_4, \theta_5, \theta_6, \theta_7, \theta_8, \theta_9, \theta_{10}\}$
14	$\{\theta_5, \theta_6, \theta_7, \theta_8, \theta_9, \theta_{10}\}$
15	$\{\theta_6, \theta_7, \theta_8, \theta_9, \theta_{10}\}$
⋮	⋮
19	$\{\theta_{10}\}$

Fig. 11. Distance between m_t and m^* for Example 12.

From steps 11–19, B_t is pruned from its first element until attaining the singleton $\{\theta_{10}\}$ at step 19. All the B_t at different steps are shown in Table VI. The second BBA m^* is $m^*(\{\theta_{10}\}) = 1$. We test different distance measures including d_J , d_T , d_C , d_F , d_{BI}^E , and d_{BI}^C to calculate the distance between m^* and m_t . Their behaviors are shown in Fig. 11.

From steps 1–9, B_t does not include $\{\theta_{10}\}$. At the step 10, $B_t = \{\theta_1, \dots, \theta_{10}\}$, which first includes $\{\theta_{10}\}$. After step 10, all distance values diminish to reach their minimum values when $B_t = \{\theta_{10}\}$. This is what we expect intuitively.

At the first stage (steps 1–9), d_C does not change when B_t changes, because the conflict between m_t and m^* never changes before the step 10, where

$$d_C(m_t, m^*) = (m_t(\{\theta_1, \theta_2, \theta_3\} + m_t(\{\theta_7\}) + m_t(B_t)) \cdot m^*(\{\theta_{10}\}) = 0.9).$$

At the second stage (steps 10–19), d_C does not change with the change of B_t . Although with the emergence of $\{\theta_{10}\}$, d_C diminishes, however, its value is fixed up to the final step, because the degree of conflict never changes after the decreasing at step 10, where

$$d_C(m_t, m^*) = (m_t(\{\theta_1, \theta_2, \theta_3\} + m_t(\{\theta_7\})) \cdot m^*(\{\theta_{10}\}) = 0.1).$$

Therefore, d_C must not be used as a proper distance measure. It is just a degree of conflict between two BBAs.

At the first stage, d_F provides unsatisfactory behavior. It slightly increases with the change of B_t , that is, it is insensitive to the change of B_t in the first stage. At the second stage, d_F provides an expected behavior, i.e., it decreases and reaches its minimum value at the final step 19.

d_{BI}^C is insensitive to the change of B_t in both first and second stages. Its value never changes in the first stage and after a decreasing at the step 10, it remains unchanged in the second stage.

The major difference between the behaviors of d_J and d_{BI}^E is in the first stage, where d_J increases while d_{BI}^E decreases. We think that the decrease makes more sense in fact, and the reason is as follows. In the first stage, the size of B_t becomes larger, and thus, the degree of uncertainty, i.e., the ambiguity of m_t increases. For two focal elements $\{\theta_1\}$ and $\{\theta_1, \theta_2\}$, although they both do not include $\{\theta_{10}\}$, the distance from $\{\theta_{10}\}$ to a more ambiguous case, i.e., $\{\theta_1, \theta_2\}$ intuitively should be smaller than the distance from $\{\theta_{10}\}$ to a more specific case. We can make an analogy here. $\{\theta_{10}\}$ is our desired result, while $\{\theta_1\}$ and $\{\theta_1, \theta_2\}$ are two undesired results. A more ambiguous undesired result should be more preferred than a clear undesired result, i.e., the distance from the desired result to the more ambiguous undesired result should be intuitively smaller.

With the increase of $|B_t|$, such a distance should intuitively further decrease. Therefore, d_{BI}^E provides the correct expected behavior in this example.

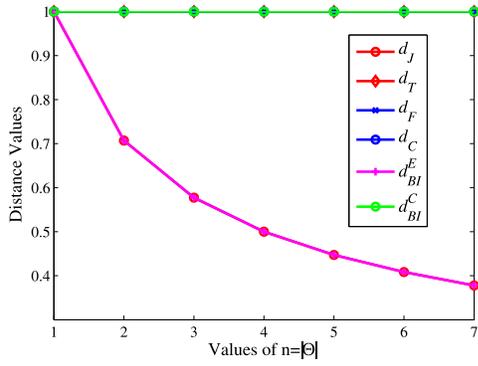
G. Example 13

Suppose that the FOD is $\Theta = \{\theta_1, \theta_2, \dots, \theta_{2n}\}$. Two BBAs defined on Θ are

$$m_1 : m_1(\{\theta_1\}) = m_1(\{\theta_2\}) = \dots = m_1(\{\theta_n\}) = 1/n$$

$$m_2 : m_2(\{\theta_{n+1}\}) = m_2(\{\theta_{n+2}\}) = \dots = m_2(\{\theta_{2n}\}) = 1/n.$$

In this example, we set n from 1 to 7, i.e., the size of FOD is from 2 to 14. We use d_J , d_T , d_C , d_F , d_{BI}^E , and d_{BI}^C to calculate the distance between m_1 and m_2 given different values of n . The distance values are shown in Fig. 12. As we can see in Fig. 12, all the distance measures except for our proposed d_{BI}^E remain unchanged with the increase of n . Our proposed d_{BI}^E decreases with the increase of n , which intuitively makes sense. The reason is as follows. With the increase of n from $k-1$ to k , the cardinality of the FOD, i.e., $|\Theta| = 2(k-1)$

Fig. 12. Distance between m_1 and m_2 for Example 13.

also increases to $2k$. Then, the number of all possible “focal” elements¹⁰ increases from $2^{2(k-1)}$ to 2^k .

Note that for each BBA, there are only n focal elements with nonzero mass assignment. When $n = k - 1$, for each BBA, there are $2(k - 1)$ focal elements in total with nonzero value; when $n = 2k$, for each BBA, there are $2(k - 1)$ focal elements in total with nonzero mass assignment. So, the number of focal elements with nonzero mass assignment increases from $2(k - 1)$ to $2k$, i.e., only two more focal elements with nonzero mass assignment are added.

On the other hand, when $n = k - 1$, for each BBA, there are $2^{2(k-1)} - (k - 1) - 1$ focal elements in total with zero mass assignment; when $n = k$, for each BBA, there are $2^{2k} - k - 1$ focal elements in total with zero mass assignment. That is, with the increase of n from $k - 1$ to k , there are $2^{2k} - k - 1 - (2^{2(k-1)} - (k - 1) - 1) = 3 \times 4^{k-1} + 1$ more focal elements with zero mass values.

The common part (focal elements with zero mass assignment) between m_1 and m_2 is significantly enlarged. At the same time, their different parts (those focal elements with nonzero values) only slightly increases of 2. Therefore, their distance should decrease. So, our proposed d_{BI}^E also behaves as expected in this case.

H. Brief Summary

According to above examples, our proposed d_{BI}^E behaves as expected in all the cases, in contrary to other measures compared. d_J also behaves well in many cases, however, in some special cases, it provides counter-intuitive behaviors. Our proposed d_{BI}^C behaves as expected in many cases, however, it is insensitive to the change of BBA due to the L_∞ norm used in its definition. Other measures like d_C , d_F , and d_T are not strict distance metrics. They generate counter-intuitive behaviors in some cases, although they can be used to describe the dissimilarity between BOEs in particular cases.

Note that the results of the above examples can only show that our proposed distance measures behave as expected in those cases in the examples. Whether the rationality of our proposed measures has more generalized meaning needs further theoretical analysis besides the testing based on examples.

¹⁰Here focal elements refer to all the subsets of the FOD Θ . They could have nonzero mass values or zero mass values.

In the following part, simulation results based on random experiments are presented.

I. Simulation

In this section, different measures are compared based on random simulations.

Relationship analyses are helpful for the joint use of multiple distance measures. Almost all the available distance measures have their own pros and cons. If one does not trust any single distance measure, one can use two distance measures together to construct a 2-D measure to describe the dissimilarity between two BOEs, e.g., Liu’s 2-D measure [40]. Then how to describe such a complementarity between members in a 2-D measure? As referred in Jousselme’s survey [22], a low correlation (close to 0) between two measures means that they quantify two distinct (and possibly complementary) aspects of the distance between two belief functions, while a high correlation means that they are redundant. Hence, weakly correlated pairs of distances could be good candidates for 2-D measures.

The relationships between different measures are described using scatter plots and the correlation coefficient. The basic procedure of the simulations is as follows.

Let \mathcal{D} denote the set of distance measures used here, which includes d_J , d_T , d_C , d_F , d_{BI}^E , and d_{BI}^C . Here, we calculate the correlation between different distance measures as follows.

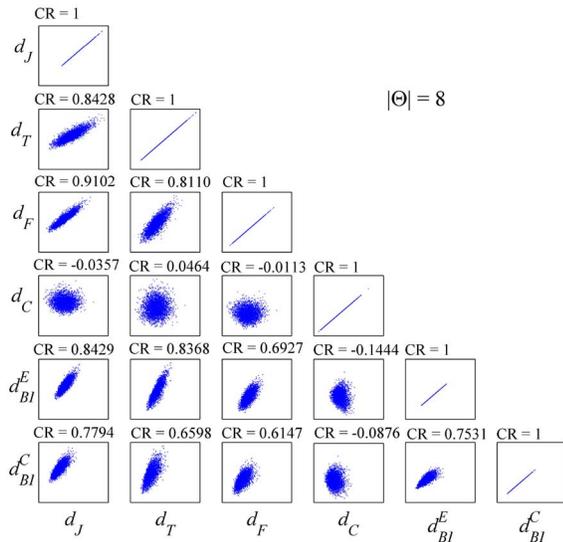
- 1) Set the size of FOD to $|\Theta|$ and generate N_s BBAs: m^s ($s = 1, \dots, N_s$) according to Algorithm 1 [39] in Table I.
- 2) Generate a reference BBA m^r according to Algorithm 1.
- 3) Pick up a distance pair d_x and d_y , where $d_x, d_y \in \mathcal{D}$ and calculate $(d_x(m^r, m^s), d_y(m^r, m^s))$ for all $s = 1, \dots, N_s$.
- 4) Draw the scatter plot for $(d_x(m^r, m^s), d_y(m^r, m^s))$ ($s = 1, \dots, N_s$) to show the correlation between d_x and d_y .
- 5) Compute the correlation coefficient [22] for d_x and d_y .

$$CR(d_x, d_y) = \frac{\sum_{s=1}^{N_s} (d_x^s - \bar{d}_x)(d_y^s - \bar{d}_y)}{\sqrt{\sum_{s=1}^{N_s} (d_x^s - \bar{d}_x)^2} \sqrt{\sum_{s=1}^{N_s} (d_y^s - \bar{d}_y)^2}} \quad (34)$$

where d_x^s denotes $d_x(m^r, m^s)$, d_y^s denotes $d_y(m^r, m^s)$, \bar{d}_x denotes the mean of d_x^s , $s = 1, \dots, N_s$, and \bar{d}_y denotes the mean of d_y^s , $s = 1, \dots, N_s$. For each pair d_x and d_y in \mathcal{D} , we calculate their correlation coefficient, to obtain a correlation matrix CR.

In simulations, we generate five types of BBAs.

- 1) *Complete BBA*: A BBA with $2^{|\Theta|} - 1$ focal elements with nonzero mass assignment.
- 2) *Fixed Length BBA*: A BBA with a fixed number of focal elements.
- 3) *Simple Support BBA*: $m(A) = a, m(\Theta) = 1 - a$, where $A \subset \Theta$ and $a \in [0, 1]$.
- 4) *Dichotomous BBA*: $m(A) = a, m(\bar{A}) = b, m(\Theta) = 1 - a - b$, where $A \subset \Theta$, \bar{A} is the complementary set of $A \in \Theta$, $a, b \in [0, 1]$ and $a + b \leq 1$.
- 5) *Consonant Support BBA*: A BBA with nested focal elements, e.g., $\{\theta_1\}$, $\{\theta_1, \theta_2\}$, and $\{\theta_1, \theta_2, \theta_3\}$.

Fig. 13. Scatter plots for $|\Theta| = 8$ using complete BBAs.

One can just make minor modifications to Algorithm 1 to randomly generate the above types of BBA.

Case A: Here we set $|\Theta| = 8$. Randomly generate 4000 complete BBAs m^s , $s = 1, \dots, 4000$. The reference BBA (complete) m^r is also randomly generated. According to the above simulation steps, we can obtain the scatter plots between each pair of distance measures in $\mathcal{D} = \{d_J, d_T, d_C, d_F, d_{BI}^E, d_{BI}^C\}$ as shown in Fig. 13, where their corresponding correlation coefficients are also provided for convenience.

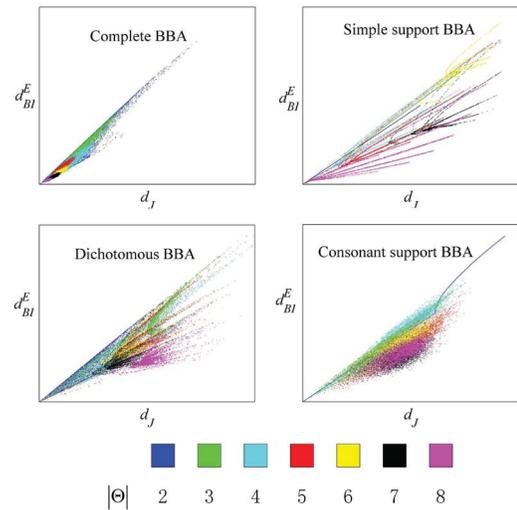
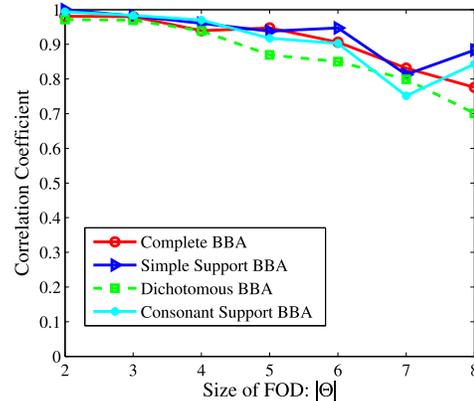
As we can see in Fig. 13, our proposed d_{BI}^E and d_{BI}^C have high correlation with Joussemle's distance d_J , which is a strict distance metric and performs well in many cases as demonstrated in the previous section.

d_C always has low correlation with other measures, since it is actually a degree of conflict, which is different to the distance.

If a 2-D or 3-D measure to jointly use multiple distance measures is desired, we can refer to the scatter plots and corresponding CR values in Fig. 13. As aforementioned, the members in the 2-D measure should better have low correlation (close to 0), thus, they could be possibly complementary. As shown in Fig. 13, our proposed d_{BI}^E and d_{BI}^C have relatively low correlation with d_C and d_F , therefore, d_C and d_F are more proper to be selected to construct 2-D measures. The focus of this paper is not the 2-D measures. We mention 2-D measure just to show our motivation of the correlation analysis between different 1-D measures. If one is interested in the construction and applications of 2-D measures, one can refer to Liu's work [40], where d_T and d_C are used jointly as a 2-D measure.

As shown in the previous section, d_{BI}^E and d_J are two very appealing measures when compared with others, and they seem highly correlated to each other. Therefore, in the sequel, we will discuss the relationship between d_J and our proposed d_{BI}^E in detail.

Case B: Although in Case A, the high correlation between d_{BI}^E and d_J has already been verified, with different FOD

Fig. 14. Scatter plots for $|\Theta| = 2, \dots, 8$ using different types of BBAs.Fig. 15. Evolution of the correlation coefficient between d_J and d_{BI}^E using different types of BBAs.

size $|\Theta|$, the correlation degree can be different. Here we use different FOD size $|\Theta|$ to check whether the correlation between d_{BI}^E and d_J is greatly affected by $|\Theta|$ or not, and to obtain the influence trend with the change of $|\Theta|$.

In this case, we set the size of the FOD to $|\Theta| = 2, 3, 4, 5, 6, 7, 8$, respectively. First, randomly generate 4000 complete BBAs, 4000 simple support BBAs, 4000 dichotomous BBAs, and 4000 consonant support BBAs. Their corresponding reference BBAs (complete, simple support, dichotomous, consonant support) m^r 's are also randomly generated.

Following the above steps and under different sizes of FOD, we can obtain the scatter plots between each pair of distance measures in $\mathcal{D} = \{d_J, d_{BI}^E\}$ for the 4000 complete BBAs, 4000 simple support BBAs, 4000 dichotomous BBAs and 4000 consonant support BBAs, respectively, as shown in Fig. 14.

With the increase of $|\Theta|$, the evolution of the correlation coefficient between d_J and d_{BI}^E for four different types of BBAs including complete, simple support, dichotomous, and consonant support are shown in Fig. 15.

As seen in Figs. 14 and 15, the increase of $|\Theta|$ leads to the decrease of the correlation coefficient for all types of BBAs. No matter using which types of BBA, d_J and d_{BI}^E are highly correlated, although the correlation coefficient decreases with

the increase of $|\Theta|$. As aforementioned, this to some extent shows the rationalities of our proposed new measure d_{BI}^E .

J. Application of Distance in BBA Approximation Evaluation

Here we provide an application of different distance measures of evidence in BBA approximations. The BBA approximation [23], [41], [42] aims to obtain a simpler BBA by removing some focal elements and thus to reduce the computational cost in the evidence combination and other operations in DST [1], [43]. A good BBA approximation should have little loss of information when simplifying the BBA. If the BBA obtained using an approximation is closer to the original BBA, such an approximation has less loss of information and thus, is more desired. Therefore, we can use the distance of evidence to evaluate BBA approximations.

Here three types of BBA approximations are compared including $k-l-x$ [23], D1 [41], and summarization (sum) [42]. Using $k-l-x$, the approximated BBA is obtained by the following.

- 1) Keeping no less than k focal elements.
- 2) Keeping no more than l focal elements.
- 3) Deleting the masses which are no larger than x .

Sum method [42] also keeps focal elements with the largest mass values as in $k-l-x$. The masses of removed focal elements are accumulated and assigned to their union set.

D1 method [41] is to keep some focal elements with the largest mass values in the original BBA and to reassign the mass assignments of the other focal elements to those kept focal elements according to a well-designed criterion. See more details in [23], [41], and [42].

$k-l-x$ has a coarse way of renormalization, and Sum method reassigns the masses of removed focal elements to their union set. D1 has a more subtle way to reassign the mass, therefore, D1 should be a better method. Here we provide a simulation with distance of evidence as the evaluation criterion to check if the evaluation results agree with the analysis.

In our simulation, $|\Theta| = 4$. A complete BBA \mathbf{m} (i.e., with $2^4 - 1 = 15$ nonempty focal elements) can be randomly generated according to Algorithm 1 in Table I. We use the distance of evidence (d_J , d_T , d_F , d_{BI}^E , and d_{BI}^C , respectively) between the approximated BBA $\hat{\mathbf{m}}$ and the original one \mathbf{m} in average as the performance evaluation criterion.

Our comparative analyses have 1000 Monte Carlo runs (i.e., totally 1000 complete BBAs are randomly generated). The number of remaining focal elements r for the approaches used here are set to from 14 down to 2 (decrease by 1). Then, different approximation results in each run can be obtained using the different approximations given a number r . The average (over 1000 runs) distance values between the original BBA \mathbf{m} and the approximated BBA $\hat{\mathbf{m}}$ obtained using different approaches given different remaining focal elements number are shown in Fig. 16(a)–(e).

Here the parameter in $k-l-x$ is set as $k = l = r$ and $x = 0.5$. As shown in Fig. 16(a)–(e), using different distances, the distance values are different; however, the changing trends are the same, i.e., with the decrease of the number remaining focal elements, the distance value increases. This represents

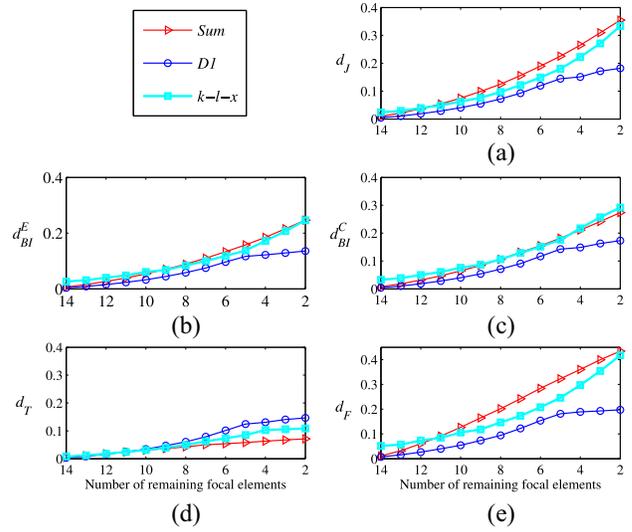


Fig. 16. Evaluation of BBA approximations using different distances. (a) d_J , (b) d_{BI}^E , (c) d_{BI}^C , (d) d_T , and (e) d_F .

more loss of information. Based on all the distances of evidence used here except for d_T , the BBA obtained by D1 is usually closer to the original BBA. It is experimentally shown that when using the distances of evidence including our new proposed ones, D1 is a better BBA approximation when compared with others. This is accordant to the analyses above, therefore, our proposed distances of evidence can be well used in performance evaluation in belief function related applications.

K. Application of d_{BI}^E in Multiple Criteria Decision Making

Here we provide an MCDM application of our developed measure d_{BI}^E , which usually performs well in the previous examples and simulations.

Let us consider a selection problem in the car purchase. Four cars $\{A_1, A_2, A_3, A_4\}$ are considered.

- 1) $A_1 =$ TOYOTA YARIS 69 VVT-i Tendence.
- 2) $A_2 =$ SUZUKI SWIFT MY15 1.2 VVT So'City.
- 3) $A_3 =$ VOLKSWAGEN POLO 1.0 60 Confortline.
- 4) $A_4 =$ OPEL CORSA 1.4 Turbo 100 ch Start/Stop Ed.

Following criteria are for selecting the best car to purchase.

- 1) C_1 is the price (in €).
- 2) C_2 is fuel consumption (in L/km).
- 3) C_3 is the CO₂ emission (in g/km).
- 4) C_4 is the fuel tank volume (in L).
- 5) C_5 is the trunk volume (in L).

From information extracted from car-makers technical characteristics available on the Internet,¹¹ we can build the score matrix $\mathbf{S} = [S_{ij}]$ for the above four cars as

$$\mathbf{S} = \begin{matrix} & \begin{matrix} C_1 & C_2 & C_3 & C_4 & C_5 \end{matrix} \\ \begin{matrix} A_1 \\ A_2 \\ A_3 \\ A_4 \end{matrix} & \begin{bmatrix} 15000 & 4.3 & 99 & 42 & 737 \\ 15290 & 5.0 & 116 & 42 & 892 \\ 15350 & 5.0 & 114 & 45 & 952 \\ 15490 & 5.3 & 123 & 45 & 1120 \end{bmatrix} \end{matrix}.$$

When we use criteria C_1 , C_2 , and C_3 , smaller is better. For criteria C_4 and C_5 , larger is better. We multiply values

¹¹<http://www.choisir-sa-voiture.com>

of columns C_1 , C_2 , and C_3 by -1 to generate a modified score matrix S' in order that the MCDM problem here is with homogeneous preference order ("larger is better") for each column

$$S' = \begin{matrix} & C_1 & C_2 & C_3 & C_4 & C_5 \\ \begin{matrix} A_1 \\ A_2 \\ A_3 \\ A_4 \end{matrix} & \begin{bmatrix} -15000 & -4.3 & -99 & 42 & 737 \\ -15290 & -5.0 & -116 & 42 & 892 \\ -15350 & -5.0 & -114 & 45 & 952 \\ -15490 & -5.3 & -123 & 45 & 1120 \end{bmatrix} \end{matrix}.$$

For simplicity, the importance $\text{imp}(C_j)$ of each criteria C_j takes a value in $\{1, 2, 3, 4, 5\}$, where 1 means *the least important*, and 5 means *the most important*. Here, $\text{imp}(C_1) = 5$, $\text{imp}(C_2) = 4$, $\text{imp}(C_3) = 4$, $\text{imp}(C_4) = 1$, and $\text{imp}(C_5) = 3$ are adopted, which means that the price (criteria C_1) is the most important one and the volume of fuel tank (criteria C_4) is the least important one. According to these importance values and after the normalization, we obtain the following vector of relative weights of criteria: $\mathbf{w} = [(5/17) (4/17) (4/17) (1/17) (3/17)]$.

We use the belief function-based technique for order preference by similarity to ideal solution (BF-TOPSIS) approach [44] with our d_{BI}^E to solve the MCDM problem above.

First, from the score matrix S' , generate BBAs $m_{i,j}(A_i)$, $m_{i,j}(\bar{A}_i)$, and $m_{i,j}(A_i \cup \bar{A}_i)$ ¹² according to the BBA generation approach proposed in [44] as

$$\begin{aligned} m_{1,1}(A_1) &= 0.9859, m_{1,1}(A_2 \cup A_3 \cup A_4) = 0.0047 \\ m_{1,1}(\Theta) &= 0.0094; m_{2,1}(A_2) = 1.0 \\ m_{2,1}(A_1 \cup A_3 \cup A_4) &= 0, m_{2,1}(\Theta) = 0 \\ m_{3,1}(A_3) &= 0.0022, m_{3,1}(A_1 \cup A_2 \cup A_4) = 0.9932 \\ m_{3,1}(\Theta) &= 0.0046; m_{4,1}(A_4) = 1.0 \\ m_{4,1}(A_1 \cup A_2 \cup A_3) &= 0, m_{4,1}(\Theta) = 0 \\ m_{1,2}(A_1) &= 1.0, m_{1,2}(A_2 \cup A_3 \cup A_4) = 0, m_{1,2}(\Theta) = 0 \\ m_{2,2}(A_2) &= 0.1250, m_{2,2}(A_1 \cup A_3 \cup A_4) = 0.4375 \\ m_{2,2}(\Theta) &= 0.4375 \\ m_{3,2}(A_3) &= 0.1250, m_{3,2}(A_1 \cup A_2 \cup A_4) = 0.4375 \\ m_{3,2}(\Theta) &= 0.4375 \\ m_{4,2}(A_4) &= 1.0, m_{4,2}(A_1 \cup A_2 \cup A_3) = 0, m_{4,2}(\Theta) = 0 \\ m_{1,3}(A_1) &= 1.0, m_{1,3}(A_2 \cup A_3 \cup A_4) = 0, m_{1,3}(\Theta) = 0 \\ m_{2,3}(A_2) &= 0.1250, m_{2,3}(A_1 \cup A_3 \cup A_4) = 0.4375 \\ m_{2,3}(\Theta) &= 0.4375 \\ m_{3,3}(A_3) &= 0.1964, m_{3,3}(A_1 \cup A_2 \cup A_4) = 0.3750 \\ m_{3,3}(\Theta) &= 0.4286 \\ m_{4,3}(A_4) &= 1.0, m_{4,3}(A_1 \cup A_2 \cup A_3) = 0, m_{4,3}(\Theta) = 0 \\ m_{1,4}(A_1) &= 0, m_{1,4}(A_2 \cup A_3 \cup A_4) = 1, m_{1,4}(\Theta) = 0 \\ m_{2,4}(A_2) &= 0, m_{2,4}(A_1 \cup A_3 \cup A_4) = 1, m_{2,4}(\Theta) = 0 \\ m_{3,4}(A_3) &= 1.0, m_{3,4}(A_1 \cup A_2 \cup A_4) = 0, m_{3,4}(\Theta) = 0 \\ m_{4,4}(A_4) &= 1.0, m_{4,4}(A_1 \cup A_2 \cup A_3) = 0, m_{4,4}(\Theta) = 0 \end{aligned}$$

¹² $i = 1, \dots, 4$ denotes the index of the alternative; $j = 1, \dots, 5$ denotes the index of the criterion.

$$\begin{aligned} m_{1,5}(A_1) &= 0, m_{1,5}(A_2 \cup A_3 \cup A_4) = 1, m_{1,5}(\Theta) = 0 \\ m_{2,5}(A_2) &= 0.1990, m_{2,5}(A_1 \cup A_3 \cup A_4) = 0.3825 \\ m_{2,5}(\Theta) &= 0.4185 \\ m_{3,5}(A_3) &= 0.3530, m_{3,5}(A_1 \cup A_2 \cup A_4) = 0.2231 \\ m_{3,5}(\Theta) &= 0.4239 \\ m_{4,5}(A_4) &= 1.0, m_{4,5}(A_1 \cup A_2 \cup A_3) = 0, m_{4,5}(\Theta) = 0. \end{aligned}$$

Second, for each alternative A_i , compute the $d_{\text{BI}}^E(\mathbf{m}_{i,j}, \mathbf{m}_{i,j}^{\text{best}})$ between $\mathbf{m}_{i,j}$ and the best ideal BBA defined by $\mathbf{m}_{i,j}^{\text{best}}(A_i) \triangleq 1$, and the distances $d_{\text{BI}}^E(\mathbf{m}_{i,j}, \mathbf{m}_{i,j}^{\text{worst}})$ between $\mathbf{m}_{i,j}$ and the worst ideal BBA defined by $\mathbf{m}_{i,j}^{\text{worst}}(\bar{A}_i) \triangleq 1$. Then, two distance matrices¹³ are obtained

$$D_{\text{BI}}^{\text{best}} = \begin{bmatrix} 0 & 0 & 0 & 0.8660 & 0.8660 \\ 0.6151 & 0.7032 & 0.7071 & 0.8660 & 0.6419 \\ 0.7100 & 0.7032 & 0.6430 & 0 & 0.5102 \\ 0.8660 & 0.8660 & 0.8660 & 0 & 0 \end{bmatrix}$$

and

$$D_{\text{BI}}^{\text{worst}} = \begin{bmatrix} 0.8660 & 0.8660 & 0.8660 & 0 & 0 \\ 0.2804 & 0.2033 & 0.1938 & 0 & 0.2552 \\ 0.1885 & 0.2033 & 0.2555 & 0.8660 & 0.3819 \\ 0 & 0 & 0 & 0.8660 & 0.8660 \end{bmatrix}.$$

Here, the element $D_{\text{BI}}^{\text{best}}(i, j) = d_{\text{BI}}^E(\mathbf{m}_{i,j}, \mathbf{m}_{i,j}^{\text{best}})$ and $D_{\text{BI}}^{\text{worst}}(i, j) = d_{\text{BI}}^E(\mathbf{m}_{i,j}, \mathbf{m}_{i,j}^{\text{worst}})$.

Third, compute the weighted average of $d_{\text{BI}}^E(\mathbf{m}_{i,j}, \mathbf{m}_{i,j}^{\text{best}})$ values with relative importance weighting factor w_j of criteria C_j . Similarly, compute the weighted average of $d_{\text{BI}}^E(\mathbf{m}_{i,j}, \mathbf{m}_{i,j}^{\text{worst}})$ values. More specifically, compute

$$d^{\text{best}}(A_i) \triangleq \sum_{j=1}^5 w_j \cdot d_{\text{BI}}^E(\mathbf{m}_{i,j}, \mathbf{m}_{i,j}^{\text{best}}) \quad (35)$$

$$d^{\text{worst}}(A_i) \triangleq \sum_{j=1}^5 w_j \cdot d_{\text{BI}}^E(\mathbf{m}_{i,j}, \mathbf{m}_{i,j}^{\text{worst}}). \quad (36)$$

The relative closeness of the alternative A_i with respect to ideal best solution A^{best} is then defined by

$$Cl(A_i, A^{\text{best}}) \triangleq \frac{d^{\text{worst}}(A_i)}{d^{\text{worst}}(A_i) + d^{\text{best}}(A_i)}. \quad (37)$$

Since $d^{\text{best}}(A_i) \geq 0$ and $d^{\text{worst}}(A_i) \geq 0$, then $Cl(A_i, A^{\text{best}}) \in [0, 1]$. If $d^{\text{best}}(A_i) = 0$, it means that the alternative A_i coincides with the ideal best solution and thus $Cl(A_i, A^{\text{best}}) = 1$ (the relative closeness of A_i with respect to A^{best} is maximal). Contrariwise, if $d^{\text{worst}}(A_i) = 0$, it means that the alternative A_i coincides with the ideal worst solution and thus $Cl(A_i, A^{\text{best}}) = 0$ (the relative closeness of A_i with respect to A^{best} is minimal).

In the final, the set of alternatives is ranked according to the descending order of $Cl(A_i, A^{\text{best}}) \in [0, 1]$, where a larger $Cl(A_i, A^{\text{best}})$ value means a better alternative (or a higher preference).

Based on the score matrix S' and importance of criteria, A_1 tends to be the best car to buy, since the three most important

¹³One can also try to use other distance measures for belief functions as referred above. Here we only use d_{BI}^E for illustration.

criteria clearly take their best values for car A_1 . When using the classical TOPSIS [45] method with the Euclidean distance, we obtain the preference order $A_4 > A_1 > A_3 > A_2$, where A_4 is the best choice and A_2 is the worst one. When we use the BF-TOPSIS method based on our proposed d_{BI}^E , we obtain a more satisfactory preference order $A_1 > A_3 > A_2 > A_4$.

As shown in this application example, our proposed distance measure can be well used in the multiple criterion decision making. d_{BI}^E has also been used successfully in other kind of applications related to risk management and for protecting housing areas against torrential floods in France [46], [47].

VI. CONCLUSION

Two novel distance measures of evidence have been proposed based on the distance measures between belief intervals. According to the comparisons between our proposed measures and the existing ones based on examples and simulations, it is shown that our proposed distances well describe the degree of closeness between different BOEs. Our results demonstrate that Euclidean distance based on belief intervals works better than the Chebyshev distance based on belief intervals.

Besides their good behaviors, the main interest of our proposed distances of evidence is that they have been established directly in the belief functions framework, contrary to most of other distance measures that switch from belief functions to probabilistic or fuzzy set framework, which leads to loss of information and bad behaviors in general.

Note that in this paper, many justifications or verifications of our new proposed distance measures are based on numerical examples and simulations. Numerical examples in belief functions related fields are usually designed according to the subjective intuitions, which lack objective criteria and the standard testing data. Furthermore, the results and conclusions only based on examples are usually incomplete. Therefore, more thorough justifications including theoretical analysis and more examples in special cases are needed to further examine our new measures. However, the theoretical evaluation or justification in belief functions related fields is still premature.

Therefore, our future work will focus on the theoretical and the objective evaluation and analysis of the belief functions related fields. We will try to establish the standard testing BBAs for the distance measures in the theory of belief functions. Our proposed distance measures will also be tested based on more experiments and simulations to find the possible counter-intuitive examples and analyze the reasons for the possible counter-intuitive behaviors. Our new distance measures will be applied to more belief functions related applications, e.g., the performance evaluations, for the further verification.

Furthermore, all the distance measures including ours are under the closed-world assumption. That is, when the mass assignment for the emptyset is positive, they cannot be used to measure the closeness between BOEs. Therefore, generalizing our new distance metrics to the open-world assumption is one of our future research directions.

REFERENCES

- [1] G. Shafer, *A Mathematical Theory of Evidence*, vol. 1. Princeton, NJ, USA: Princeton Univ. Press, 1976.
- [2] G. Lin, J. Liang, and Y. Qian, "An information fusion approach by combining multigranulation rough sets and evidence theory," *Inf. Sci.*, vol. 314, pp. 184–199, Sep. 2015.
- [3] Z.-G. Liu, Q. Pan, and J. Dezert, "Evidential classifier for imprecise data based on belief functions," *Knowl. Based Syst.*, vol. 52, pp. 246–257, Nov. 2013.
- [4] Z.-G. Liu, Q. Pan, J. Dezert, and G. Mercier, "Credal c-means clustering method based on belief functions," *Knowl. Based Syst.*, vol. 74, pp. 119–132, Jan. 2015.
- [5] D. Han, J. Dezert, J.-M. Tacnet, and C. Han, "A fuzzy-cautious OWA approach with evidential reasoning," in *Proc. 15th Int. Conf. Inf. Fusion (FUSION)*, Singapore, Jul. 2012, pp. 278–285.
- [6] X. Xu, Z. Zhang, D.-L. Xu, and Y.-W. Chen, "Interval-valued evidence updating with reliability and sensitivity analysis for fault diagnosis," *Int. J. Comput. Intell. Syst.*, vol. 9, no. 3, pp. 396–415, 2016. [Online]. Available: <http://dx.doi.org/10.1080/18756891.2016.1175808>
- [7] F. Smarandache and J. Dezert, *Advances and Applications of DSMT for Information Fusion: Collected Works-Volume 3*. Rehoboth, DE, USA: Amer. Res. Press, 2009.
- [8] P. Wang, "A defect in Dempster-Shafer theory," in *Proc. 10th Int. Conf. Uncertainty Artif. Intell.*, Seattle, WA, USA, Jul. 1994, pp. 560–566.
- [9] J. Dezert and A. Tchamova, "On the validity of Dempster's fusion rule and its interpretation as a generalization of Bayesian fusion rule," *Int. J. Intell. Syst.*, vol. 29, no. 3, pp. 223–252, 2014.
- [10] P. Smets and R. Kennes, "The transferable belief model," *Artif. Intell.*, vol. 66, no. 2, pp. 191–234, 1994.
- [11] F. Smarandache and J. Dezert, *Advances and Applications of DSMT for Information Fusion: Collected Works-Volume 4*. Rehoboth, DE, USA: Amer. Res. Press, 2015.
- [12] T. Denceux, "A neural network classifier based on Dempster-Shafer theory," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 30, no. 2, pp. 131–150, Mar. 2000.
- [13] A.-L. Jousselme, D. Grenier, and É. Bossé, "A new distance between two bodies of evidence," *Inf. Fusion*, vol. 2, no. 2, pp. 91–101, 2001.
- [14] D. Han, J. Dezert, and C. Han, "New basic belief assignment approximations based on optimization," in *Proc. 15th Int. Conf. Inf. Fusion*, Singapore, Jul. 2012, pp. 286–293.
- [15] F. Cuzzolin, "Lp consonant approximations of belief functions," *IEEE Trans. Fuzzy Syst.*, vol. 22, no. 2, pp. 420–436, Apr. 2014.
- [16] D. Han, J. Dezert, and Y. Yang, "Two novel methods for BBA approximation based on focal element redundancy," in *Proc. 18th Int. Conf. Inf. Fusion*, Washington, DC, USA, Jul. 2015, pp. 428–434.
- [17] Y. Yang and Y. Liu, "Iterative approximation of basic belief assignment based on distance of evidence," *PLoS One*, vol. 11, no. 2, pp. 1–27, Feb. 2016.
- [18] S. B. Hariz, Z. Elouedi, and K. Mellouli, "Clustering approach using belief function theory," in *Proc. 12th Int. Conf. Artif. Intell. Methodol. Syst. Appl.*, Varna, Bulgaria, Sep. 2006, pp. 162–171.
- [19] J. Schubert, "Clustering decomposed belief functions using generalized weights of conflict," *Int. J. Approx. Reason.*, vol. 48, no. 2, pp. 466–480, Jun. 2008.
- [20] D. Mercier, B. Quost, and T. Denceux, "Refined modeling of sensor reliability in the belief function framework using contextual discounting," *Inf. Fusion*, vol. 9, no. 2, pp. 246–258, Apr. 2008.
- [21] A. Martin, A.-L. Jousselme, and C. Osswald, "Conflict measure for the discounting operation on belief functions," in *Proc. 11th Int. Conf. Inf. Fusion*, Cologne, Germany, Jun. 2008, pp. 1–8.
- [22] A.-L. Jousselme and P. Maupin, "Distances in evidence theory: Comprehensive survey and generalizations," *Int. J. Approx. Reason.*, vol. 53, no. 2, pp. 118–145, 2012.
- [23] B. Tessem, "Approximations for efficient computation in the theory of evidence," *Artif. Intell.*, vol. 61, no. 2, pp. 315–329, 1993.
- [24] D. Han, J. D. Onera, C. Han, and Y. Yang, "New dissimilarity measures in evidence theory," in *Proc. 14th Int. Conf. Inf. Fusion*, Chicago, IL, USA, Jul. 2011, pp. 1–7.
- [25] F. Cuzzolin, "A geometric approach to the theory of evidence," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 38, no. 4, pp. 522–534, Jul. 2008.
- [26] A. Irpino and R. Verde, "Dynamic clustering of interval data using a Wasserstein-based distance," *Pattern Recognit. Lett.*, vol. 29, no. 11, pp. 1648–1658, 2008.

- [27] D. Han, J. Dezert, and Y. Yang, "New distance measures of evidence based on belief intervals," in *Proc. 3rd Int. Conf. Belief Funct.*, Oxford, U.K., Sep. 2014, pp. 432–441.
- [28] R. R. Yager, "Entropy and specificity in a mathematical theory of evidence," *Int. J. Gen. Syst.*, vol. 9, no. 4, pp. 249–260, 1983.
- [29] A.-L. Jousselme, C. Liu, D. Grenier, and É. Bossé, "Measuring ambiguity in the evidence theory," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 36, no. 5, pp. 890–903, Sep. 2006.
- [30] D. Harmanec and G. J. Klir, "Measuring total uncertainty in Dempster–Shafer theory: A novel approach," *Int. J. Gen. Syst.*, vol. 22, no. 4, pp. 405–419, 1994.
- [31] Y. Yang and D. Han, "A new distance-based total uncertainty measure in the theory of belief functions," *Knowl. Based Syst.*, vol. 94, pp. 114–123, Feb. 2016.
- [32] L. A. Zadeh, "A simple view of the Dempster–Shafer theory of evidence and its implication for the rule of combination," *AI Mag.*, vol. 7, no. 2, pp. 85–90, 1986.
- [33] X. Deng, D. Han, J. Dezert, Y. Deng, and Y. Shyr, "Evidence combination from an evolutionary game theory perspective," *IEEE Trans. Cybern.*, vol. 46, no. 9, pp. 2070–2082, Sep. 2016.
- [34] P. Smets, "Decision making in the TBM: The necessity of the pignistic transformation," *Int. J. Approx. Reason.*, vol. 38, no. 2, pp. 133–147, 2005.
- [35] D. Han, J. Dezert, and Z. Duan, "Evaluation of probability transformations of belief functions for decision making," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 46, no. 1, pp. 93–108, Jan. 2016.
- [36] D. Han, Y. Deng, C. Han, and Y. Yang, "Some notes on betting commitment distance in evidence theory," *Sci. China Inf. Sci.*, vol. 55, no. 3, pp. 558–565, 2012.
- [37] M. Bouchard, A.-L. Jousselme, and P.-E. Doré, "A proof for the positive definiteness of the Jaccard index matrix," *Int. J. Approx. Reason.*, vol. 54, no. 5, pp. 615–626, 2013.
- [38] A. Antonucci, "An interval-valued dissimilarity measure for belief functions based on credal semantics," in *Proc. 2nd Int. Conf. Belief Funct.*, Compiègne, France, May 2012, pp. 37–44.
- [39] T. Burger and S. Destercke, "Random generation of mass functions: A short howto," in *Proc. 2nd Int. Conf. Belief Funct.*, Compiègne, France, May 2012, pp. 145–152.
- [40] W. Liu, "Analyzing the degree of conflict among belief functions," *Artif. Intell.*, vol. 170, no. 11, pp. 909–924, 2006.
- [41] M. Bauer, "Approximations for decision making in the Dempster–Shafer theory of evidence," in *Proc. 12th Int. Conf. Uncertainty Artif. Intell.*, Portland, OR, USA, Aug. 1996, pp. 73–80.
- [42] J. D. Lowrance, T. D. Garvey, and T. M. Strat, "A framework for evidential-reasoning systems," in *Proc. 5th Nat. Conf. Artif. Intell. (AAAI)*, Philadelphia, PA, USA, Aug. 1986, pp. 896–901.
- [43] P. Smets, "Practical uses of belief functions," in *Proc. 15th Conf. Uncertainty Artif. Intell.*, vol. 15. Stockholm, Sweden, 1999, pp. 612–621.
- [44] J. Dezert, D. Han, and H. Yin, "A new belief function based approach for multi-criteria decision-making support," in *Proc. 19th Int. Conf. Inf. Fusion*, Heidelberg, Germany, Jul. 2016, pp. 782–789.
- [45] C.-L. Hwang and K.-S. Yoon, *Multiple Attribute Decision Making: Methods and Applications*. Berlin Heidelberg, German: Springer-Verlag, 1981.
- [46] S. Carladous, J.-M. Tacnet, J. Dezert, D. Han, and M. Batton-Hubert, "Evaluation of efficiency of torrential protective structures with new BF-TOPSIS methods," in *Proc. 19th Int. Conf. Inf. Fusion*, Heidelberg, Germany, Jul. 2016, pp. 2267–2274.
- [47] S. Carladous, J.-M. Tacnet, J. Dezert, D. Han, and M. Batton-Hubert, "Applying ER-MCDA and BF-TOPSIS to decide on effectiveness of torrent protection," in *Proc. 4th Int. Conf. Belief Funct.*, Prague, Czech Republic, Sep. 2016, pp. 56–65.



Deqiang Han (S'08–M'09) was born in Xi'an, China, in 1980. He received the bachelor's degree in communication and control engineering and the master's and Ph.D. degrees in control science and engineering from Xi'an Jiaotong University, Xi'an, in 2001, 2004, and 2008, respectively.

He is currently an Associate Professor with Xi'an Jiaotong University. His research interests include evidence theory, information fusion, and pattern classification.

Dr. Han is an International Society of Information Fusion Member and a Technical Program Committee Member for the 1st–7th Chinese Conference on Information Fusion in 2009–2015.



Jean Dezert was born in l'Hay les Roses, France in 1962. He received the electrical engineering degree from the École Française de Radioélectricité Électronique et Informatique, Paris, France, in 1985, the D.E.A. degree from University Paris VII (Jussieu), Paris, in 1986, and the Ph.D. degree from University Paris XI, Orsay, France, in 1990, all in automatic control and signal processing.

He has been a Senior Research Scientist with the Information and Fusion Systems Research Unit, Information and Modeling and Processing Department, ONERA, Palaiseau, France, since 1993. He gave several invited plenary talks and seminars on information fusion in Europe, America, Australia, and China, during latest years. His current research interests include autonomous navigation, estimation, stochastic systems and their applications to multisensor-multitarget tracking, information fusion, and plausible reasoning.

Dr. Dezert has served as a Local Arrangements Organizer for the 2000 3rd International Conference on Information Fusion (Fusion) in Paris, and a Secretary, an Executive Vice-President, and the President for the International Society of Information Fusion (ISIF) in 2001, 2004, and 2016, respectively. He has been involved in the Technical Program Committee of Fusion 2001–2016 International Conferences. He was a Board Member of the ISIF.



Yi Yang was born in Xi'an, China, in 1980. She received the bachelor's degree in automation from the Xi'an University of Technology, Xi'an, China, in 2002, and the master's and Ph.D. degrees in control science and engineering from Xi'an Jiaotong University, Xi'an, in 2005 and 2010, respectively.

She is currently a Lecturer with the School of Aerospace, Xi'an Jiaotong University. Her research interests include evidence theory, pattern classification, and image processing.