



A new distance-based total uncertainty measure in the theory of belief functions



Yi Yang^a, Deqiang Han^{b,*}

^a SKLSVMS, School of Aerospace, Xi'an Jiaotong University, Xi'an 710049, China

^b Institute of Integrated Automation, Xi'an Jiaotong University, Xi'an 710049, China

ARTICLE INFO

Article history:

Received 27 July 2015

Revised 22 October 2015

Accepted 20 November 2015

Available online 27 November 2015

Keywords:

Belief functions

Evidence theory

Uncertainty measure

Belief interval

Distance of interval

ABSTRACT

The theory of belief functions is a very important and effective tool for uncertainty modeling and reasoning, where measures of uncertainty are very crucial for evaluating the degree of uncertainty in a body of evidence. Several uncertainty measures in the theory of belief functions have been proposed. However, existing measures are generalizations of measures in the probabilistic framework. The inconsistency between different frameworks causes limitations to existing measures. To avoid these limitations, in this paper, a new total uncertainty measure is proposed directly in the framework of belief functions theory without changing the theoretical frameworks. The average distance between the belief interval of each singleton and the most uncertain case is used to represent the total uncertainty degree of the given body of evidence. Numerical examples, simulations, applications and related analyses are provided to verify the rationality of our new measure.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Dempster–Shafer evidence theory (DST) [1], also called the theory of belief functions, has been widely used in many applications related to uncertainty modeling and reasoning, e.g., information fusion [2], pattern classification [3,4], clustering analysis [5], fault diagnosis [6], and multiple attribute decision-making (MADM) [7,8].

In the theory of belief functions, there are two types of uncertainty including the discord (or conflict or randomness) [9] and the non-specificity [10], hence the ambiguity [11]. Various kinds of measures for these two types of uncertainty and the total uncertainty including both two types were proposed. The measures of discord include the discord measure [9], the strife [9], the confusion [12], etc; the measures of non-specificity include Dubois and Prade's definition [10] generalized from the Hartley entropy [13] in the classical set theory, Yager's definition [14], and Korner's definition [15], etc. The most representative total uncertainty measures are the aggregated measure (AU) [16] and the ambiguity measure (AM) [11].

In essential, no matter AU or AM, they are the generalization of Shannon entropy [17] in probability theory, but not the direct definition in the framework of the theory of belief functions. That is, they pick up a probability according to some criteria or constraints

established based on the given body of evidence (BOE), and then calculate the corresponding Shannon entropy of the probability to indirectly depict the degree of uncertainty for the given body of evidence. As mentioned in the related references [11,18], AU and AM have their own limitations. For example, they are insensitive to the change of BBA. These limitations to some extent are related to the inconsistency [19,20] between the framework of the theory of belief functions and that of the probability theory. Therefore, to avoid the limitations of the traditional definitions for the uncertainty measure, in our work, a new total uncertainty measure is proposed directly in the framework of the theory of belief functions without the switching between different frameworks. We analyze the belief interval and conclude that belief intervals carry both the randomness part and the imprecision part (non-specificity) in the uncertainty incorporated in a BOE. Thus, it is feasible to define a total uncertainty measure for a BOE. Given a BOE, the distance between the belief interval of each singleton and the most uncertain interval [0, 1] is used for constructing the degree of total uncertainty. The larger the average distance, the smaller the degree of uncertainty. Since there is no switch of theoretical frameworks, our new definition has desired properties including the ideal value range and the monotonicity. Furthermore, the uncertainty measure can provide more rational results when compared with the traditional ones, which can be supported by experimental results and related analyses.

The rest of this paper is organized as follows. Section 2 provides the brief introduction of the theory of belief functions. Commonly used uncertainty measures in the theory of belief functions are

* Corresponding author. Tel.: +86 13119115975.

E-mail addresses: jiafeiyi@mail.xjtu.edu.cn (Y. Yang), deqhan@gmail.com, deqhan@mail.xjtu.edu.cn (D. Han).

introduced in Section 3. Some drawbacks of the available total uncertainty measures including AM and AU are also pointed out in Section 3. In Section 4, a new total uncertainty measure is proposed. Some desired properties and related analyses on the new proposed definition are also provided. Experiments and simulations are provided in Section 5 to show the rationality of our proposed total uncertainty measure. An application of the new total uncertainty measure on feature evaluation is provided in Section 6. Section 7 concludes this work.

2. Basics of the theory of belief functions

The basic conception in the theory of belief functions [1] is the frame of discernment (FOD), whose elements are mutually exclusive and exhaustive, representing what we concern. $m : 2^\Theta \rightarrow [0, 1]$ is called a basic belief assignment (BBA) defined over an FOD Θ if it satisfies

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

When $m(A) > 0$, A is called a focal element. A BBA is also called a mass function. The set of all the focal elements denoted by \mathfrak{F} and their corresponding mass assignments constitute a body of evidence (BOE) (\mathfrak{F}, m) .

The belief function (Bel) and plausibility function (Pl) are defined as

$$Bel(A) = \sum_{B \subseteq A} m(B) \quad (2)$$

$$Pl(A) = \sum_{A \cap B \neq \emptyset} m(B) \quad (3)$$

The belief function $Bel(A)$ represents the justified specific support for the focal element (or proposition) A , while the plausibility function $Pl(A)$ represents the potential specific support for A . The length of the belief interval $[Bel(A), Pl(A)]$ is used to represent the degree of imprecision for A .

Two independent BBAs $m_1(\cdot)$ and $m_2(\cdot)$ can be combined using Dempster's rule of combination as [1]

$$m(A) = \begin{cases} 0, & \text{if } A = \emptyset \\ \frac{\sum_{A_i \cap B_j = A} m_1(A_i) m_2(B_j)}{1 - \sum_{A_i \cap B_j = \emptyset} m_1(A_i) m_2(B_j)}, & \text{if } A \neq \emptyset \end{cases} \quad (4)$$

There are still some other alternative combination rules. See [21] for details. The theory of belief functions has been criticized for its validity [19,20,22–25]. It is not a successful generalization of the probability theory, i.e., there exists inconsistency [20] between the framework of the theory of belief functions and that of the probability theory.

3. Uncertainty measures in the theory of belief functions

In the theory of belief functions, there are two types of uncertainty including the discord (or the conflict or the randomness) and the non-specificity, hence ambiguity [11].

3.1. Measure of discord in belief function

Measures of discord are for describing the randomness (or discord or conflict) in a BOE [11]. Available definitions are listed below. Although with various names, they are all for the discord part of the uncertainty in a BOE.

(1) Confusion measure [12]

$$Conf(m) = - \sum_{A \subseteq \Theta} m(A) \log_2 (Bel(A)) \quad (5)$$

(2) Dissonance measure [14]

$$Diss(m) = - \sum_{A \subseteq \Theta} m(A) \log_2 (Pl(A)) \quad (6)$$

(3) Discord measure [26]

$$Disc(m) = - \sum_{A \subseteq \Theta} m(A) \log_2 \left[1 - \sum_{B \subseteq \Theta} m(B) \frac{|B - A|}{|B|} \right] \quad (7)$$

(4) Strife measure [9]

$$Stri(m) = - \sum_{A \subseteq \Theta} m(A) \log_2 \left[1 - \sum_{B \subseteq \Theta} m(B) \frac{|A - B|}{|A|} \right] \quad (8)$$

Note that all these definitions can be considered as Shannon entropy-like measures. More detailed information on these measures can be found in [9].

3.2. Measures for non-specificity in belief function

Non-specificity [15,27] means two or more alternatives are left unspecified and represents an imprecision degree. It only focuses on those focal elements with cardinality larger than one. Non-specificity is a special uncertainty type in the framework of belief functions theory when compared with the probabilistic framework. Some non-specificity measures [10,14,15] were proposed. The most commonly used non-specificity definition is [10]

$$NS(m) = \sum_{A \subseteq \Theta} m(A) \log_2 |A| \quad (9)$$

It can be regarded as a generalized Hartley measure [13] from the classical set theory. When the BBA $m(\cdot)$ is a Bayesian BBA, i.e., it only has singleton focal elements, it reaches the minimum value 0. When BBA $m(\cdot)$ is a vacuous BBA, i.e., $m(\Theta) = 1$, it reaches the maximum value $\log_2(|\Theta|)$. This definition was proved to have the uniqueness by Ramer [28] and it satisfies all the requirements of the non-specificity measure.

3.3. Measures for total uncertainty in belief functions theory

(1) Aggregated Uncertainty (AU) [16]

$$AU(m) = \max \left[- \sum_{\theta \in \Theta} p_\theta \log_2 p_\theta \right] \quad (10)$$

$$\text{s.t.} \begin{cases} p_\theta \in [0, 1], \forall \theta \in \Theta \\ \sum_{\theta \in \Theta} p_\theta = 1 \\ Bel(A) \leq \sum_{\theta \in A} p_\theta \leq 1 - Bel(\bar{A}), \forall A \subseteq \Theta \end{cases}$$

In fact for AU , given a BBA, the probability with the maximum Shannon entropy under the constraints established using the given BBA is selected and its corresponding Shannon entropy value is defined as the value for AU . Therefore, it is also called as “upper entropy” [29]. It is an aggregated total uncertainty (ATU) measure, which captures both non-specificity and discord. AU satisfies all the requirements [29] for uncertainty measure including probability consistency, set consistency, value range, monotonicity, sub-additivity and additivity for the joint BBA in Cartesian space.

(2) Ambiguity Measure (AM) [11]

$$AM(m) = - \sum_{\theta \in \Theta} BetP_m(\theta) \log_2 (BetP_m(\theta)) \quad (11)$$

where $BetP_m(\theta) = \sum_{B \subseteq \Theta} m(B)/|B|$ is the pignistic probability [30] of a BBA. In fact AM uses the Shannon entropy of the pignistic probability of a given BBA to represent the uncertainty.

3.4. Drawbacks of available total uncertainty measures for belief functions

The traditional total uncertainty measures have their own drawbacks. AM cannot satisfy the sub-additivity (for joint BBA in Cartesian space) which has been pointed out by Klir and Lewis [18]. Moreover, AM is criticized due to its logical non-monotonicity [29] as shown in Example 1.

3.4.1. Example 1

Suppose that the FOD is $\{\theta_1, \theta_2, \theta_3\}$. There are two BBAs as follows:

$$\begin{aligned} m_1(\{\theta_1, \theta_2\}) &= 1/3, & m_1(\{\theta_1, \theta_3\}) &= 1/2, & m_1(\{\theta_2, \theta_3\}) &= 1/6; \\ m_2(\{\theta_1, \theta_2, \theta_3\}) &= 1/3, & m_2(\{\theta_1, \theta_3\}) &= 1/2, & m_2(\{\theta_2, \theta_3\}) &= 1/6. \end{aligned}$$

Obviously, from the two BBAs, there exists

$$Bel_2(A) \leq Bel_1(A), \quad Pl_1(A) \leq Pl_2(A), \quad \forall A \subseteq \Theta.$$

That is to say, all the belief intervals of m_1 can be contained by those corresponding belief intervals of m_2 , which means that m_2 has a higher level of uncertainty. We can calculate their corresponding AM and AU as follows:

$$\begin{aligned} 1.5546 &= AM(m_1) \geq AM(m_2) = 1.5100; \\ AU(m_1) &= AU(m_2) = 1.5850 = \log_2 3. \end{aligned}$$

We can see that the results of AM are counter-intuitive, because AM violates the monotonicity, i.e., AM decreases the total quantity of uncertainty in this example where a clear increment of uncertainty of m_2 compared with that of m_1 . Although AU does not bring out a decrease of the quantity of uncertainty, it generates two equal values, which is also counter-intuitive, because m_2 should have higher level of uncertainty. We provide further analyses on this example. The belief function and plausibility function of m_1 and m_2 are

$$\begin{aligned} Bel_1(\{\theta_1\}) &= 0.0000, & Pl_1(\{\theta_1\}) &= 0.8333; \\ Bel_1(\{\theta_2\}) &= 0.0000, & Pl_1(\{\theta_2\}) &= 0.5000; \\ Bel_1(\{\theta_1, \theta_2\}) &= 0.3333, & Pl_1(\{\theta_1, \theta_2\}) &= 1.0000; \\ Bel_1(\{\theta_3\}) &= 0.0000, & Pl_1(\{\theta_3\}) &= 0.6667; \\ Bel_1(\{\theta_1, \theta_3\}) &= 0.5000, & Pl_1(\{\theta_1, \theta_3\}) &= 1.0000; \\ Bel_1(\{\theta_2, \theta_3\}) &= 0.1667, & Pl_1(\{\theta_2, \theta_3\}) &= 1.0000; \\ Bel_1(\Theta) &= 1.0000, & Pl_1(\Theta) &= 1.0000; \\ \\ Bel_2(\{\theta_1\}) &= 0.0000, & Pl_2(\{\theta_1\}) &= 0.8333; \\ Bel_2(\{\theta_2\}) &= 0.0000, & Pl_2(\{\theta_2\}) &= 0.5000; \\ Bel_2(\{\theta_1, \theta_2\}) &= 0.0000, & Pl_2(\{\theta_1, \theta_2\}) &= 1.0000; \\ Bel_2(\{\theta_3\}) &= 0.0000, & Pl_2(\{\theta_3\}) &= 0.6667; \\ Bel_2(\{\theta_1, \theta_3\}) &= 0.5000, & Pl_2(\{\theta_1, \theta_3\}) &= 1.0000; \\ Bel_2(\{\theta_2, \theta_3\}) &= 0.1667, & Pl_2(\{\theta_2, \theta_3\}) &= 1.0000; \\ Bel_2(\Theta) &= 1.0000, & Pl_2(\Theta) &= 1.0000. \end{aligned}$$

Since AU tries to find a probability maximizing the Shannon entropy, and the uniform p.m.f. (probability mass function) $P(\theta_1) = P(\theta_2) = P(\theta_3) = 1/3$, which has the maximum Shannon entropy, satisfies all the constraints (shown in Eq. (10)) established using Bel_1 , Pl_1 and those using Bel_2 and Pl_2 . Therefore, both AU_1 and AU_2 reach the maximum value $\log_2 3 = 1.5850$.

3.4.2. Example 2

Suppose that the FOD $\Theta = \{\theta_1, \theta_2\}$. A BBA over Θ is $m(\{\theta_1\}) = a$, $m(\{\theta_2\}) = b$, $m(\{\Theta\}) = 1 - a - b$, where $a, b \in [0, 0.5]$. Here, we calculate AU , AM values corresponding to different values of a and b . The change of AU and AM values with the change of a and b are illustrated in Fig. 1.

As shown in Fig. 1, AM reaches its maximum when $a = b$, because when $a = b$ (no matter a or b 's value is large or small), the pignistic probability is uniformly distributed. This is counter-intuitive, because AM cannot discern the effect of total set's mass assignment values to the uncertainty degree. For example, $m_1(\{\theta_1\}) = m_1(\{\theta_2\}) = 0.5$ and $m_2(\{\theta_1\}) = m_2(\{\theta_2\}) = 0.25$, $m_2(\Theta) = 0.5$, there exists $AM(m_1) = AM(m_2)$. Obviously, it is irrational to say that m_1 and m_2 have the same degree of total uncertainty. AU never changes with the change of a and b . The value of AU is always $\log_2 2 = 1$. The reason is analyzed as follows. The constraints for calculating AU in this example are

$$\begin{aligned} Bel(\{\theta_1\}) &= a \leq P(\theta_1) \leq 1 - b = Pl(\{\theta_1\}); \\ Bel(\{\theta_2\}) &= b \leq P(\theta_2) \leq 1 - a = Pl(\{\theta_2\}); \\ Bel(\Theta) &= 1 - a - b \leq P(\theta_1) \leq 1 = Pl(\Theta); \end{aligned}$$

Since AU tries to find a p.m.f. with the maximum Shannon entropy, and the uniformly distributed $P(\theta_1) = P(\theta_2) = 0.5$ always satisfies the constraints above (because $a, b \in [0, 0.5]$), no matter how a and b change, $P(\theta_1) = P(\theta_2) = 0.5$ is always picked up when calculating AU and thus, AU always equals to $\log_2 2 = 1$. However, intuitively, the degree of uncertainty should change with the change of a and b . So, AU cannot well describe the degree of total uncertainty here.

As we can see above, both AM and AU borrow the uncertainty measure (Shannon entropy) in the probability theory framework and they both have drawbacks. However, as aforementioned, the belief functions theory is not a successful generalization of the probability theory. There exists inconsistency between the two frameworks. We do not prefer such a switch between different frameworks, which might cause problems in representing the uncertainty in belief functions. Therefore, in our work, we design the total uncertainty measure directly in the framework of belief functions theory as introduced in the next section.

4. Belief interval distance-based total uncertainty measure

4.1. Belief interval and uncertainty

The belief interval [1] $[Bel(A), Pl(A)]$ of a given focal elements A in a BBA m can represent its uncertainty degree as analyzed below.

Case I: The most uncertain case is $Bel(A) = 0$ and $Pl(A) = 1$, i.e., the belief interval is $[0, 1]$.

Case II: The clearest case is $Bel(A) = 1$ and $Pl(A) = 1$ (A is assure to occur) or $Bel(A) = 0$ and $Pl(A) = 0$ (A is assure to never occur).

Case III: When $Bel(A) = a$, $Pl(A) = b$, $\forall a, b \in (0, 1)$, the belief interval is $[a, b]$. For A , the degree of imprecision can be represented by $b - a$, the probability of whether A occurs or not (randomness) is large than a and less than b .

Case IV: When $Bel(A) = Pl(A) = a$, $\forall a \in (0, 1)$, the belief interval is $[a, a]$. It means that A has no imprecision and whether A occurs or not cannot be clearly determined (it is with the randomness, i.e., the probability a).

Therefore, the information related to the uncertainty carried by a belief interval both include the randomness part and the imprecision part (non-specificity). Thus, given a BBA, we can fully utilize the information of the belief intervals to measure the total uncertainty of the BBA. Then, how to use the information of belief intervals?

Given the belief interval of A , i.e., $[Bel(A), Pl(A)]$, if the belief interval is farther from the most uncertain case $[0, 1]$, then A has smaller uncertainty; if the belief interval is nearer to the most uncertain case $[0, 1]$, then A has larger uncertainty. Then, how to describe the distance between belief intervals? Here we use the distance between interval numbers [31,32] as follows. Given two interval numbers $[a_1, b_1]$ and $[a_2, b_2]$, a strict distance is defined as

$$\begin{aligned} d^l([a_1, b_1], [a_2, b_2]) \\ = \sqrt{\left[\frac{a_1 + b_1}{2} - \frac{a_2 + b_2}{2}\right]^2 + \frac{1}{3}\left[\frac{b_1 - a_1}{2} - \frac{b_2 - a_2}{2}\right]^2} \quad (12) \end{aligned}$$

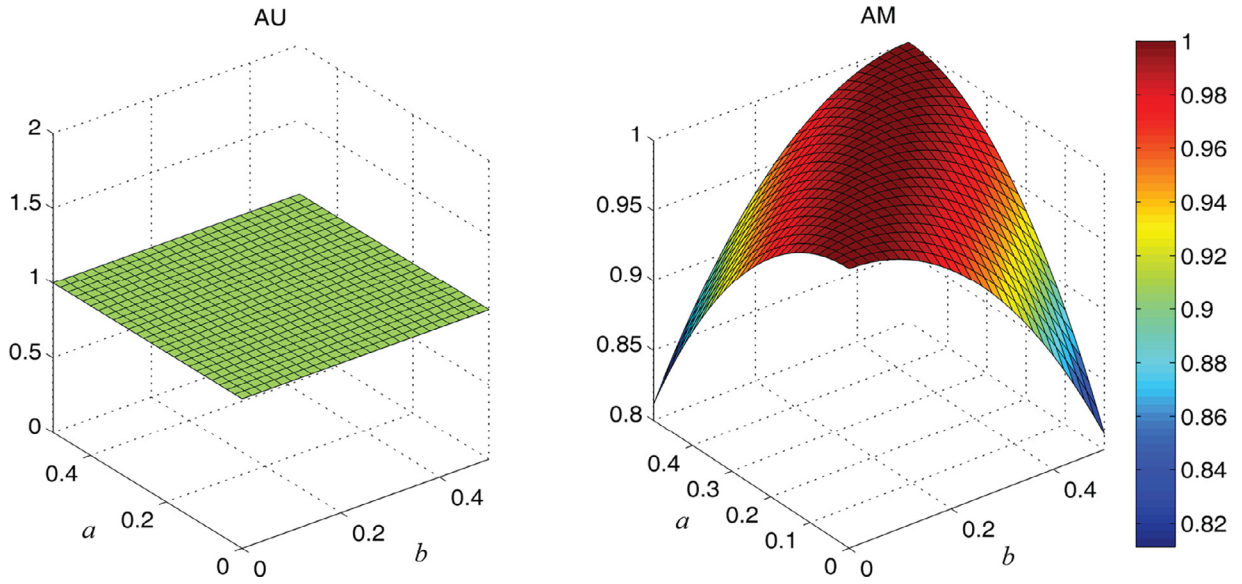


Fig. 1. The change of total uncertainty measures in Example 2.

The belief interval $[Bel(A), Pl(A)]$ can be considered as an interval number. If we replace $[a_1, b_1]$ by $[Bel(A), Pl(A)]$ and $[a_2, b_2]$ by $[0, 1]$, then, the distance obtained can be used to represent how far it is from A to the most uncertain case. As shown in Eq. (12), in d^l , the proportion of $[(Bel(\{\theta_i\}) + Pl(\{\theta_i\}))/2 - 1]^2$ and $[(Pl(\{\theta_i\}) - Bel(\{\theta_i\}))/2 - 1]^2$ is $1/3$. Generally speaking, an interval number is a kind of fuzzy data [31]. To define the distance between two fuzzy data, the value of such a proportion depends on the shape [31] (could be the symmetric triangular, normal, parabolic, square root fuzzy data, and interval number) of the fuzzy data model. Here, we use the belief intervals (interval numbers), therefore, there is a rational and natural assumption that the interval is with a uniform distribution. Under such an assumption, the proportion should be $1/3$ and d^l corresponds to Mallows' distance between two distributions, hence a strict distance metric for interval numbers. By using such a strict distance metric for interval numbers, we can further define a total uncertainty measure for a BOE as follows.

4.2. New total uncertainty measure based on belief intervals

Suppose that m is a BBA over the FOD $\Theta = \{\theta_1, \dots, \theta_n\}$. First, calculate the belief interval $[Bel(\{\theta_i\}), Pl(\{\theta_i\})]$, $i = 1, \dots, n$ for each singleton θ_i . Then, calculate the distance between each $[Bel(\{\theta_i\}), Pl(\{\theta_i\})]$ and $[0, 1]$, which represents the degree of departure from the most uncertain case for each singleton. Since larger distance (larger departure from the most uncertain case) means smaller total uncertainty, one minus the normalized and averaged belief interval distances (for all singletons) is used as the total uncertainty measure for the BBA m as shown in Eq. (13)

$$TU^l(m) = 1 - \frac{1}{n} \cdot \sqrt{3} \cdot \sum_{i=1}^n d^l([Bel(\{\theta_i\}), Pl(\{\theta_i\})], [0, 1]) \quad (13)$$

where $\sqrt{3}$ is the normalization factor¹.

The newly proposed TU^l has no drawbacks as those of AU and AM pointed out in Examples 1 and 2. See the Examples 3 and 4 in Section 5 for details.

Here, we provide an illustrative example to show how TU^l is calculated.

Suppose that FOD is $\Theta = \{\theta_1, \theta_2, \theta_3\}$. A BBA over Θ is $m(\{\theta_1\}) = 0.3$, $m(\{\theta_2, \theta_3\}) = 0.5$, $m(\Theta) = 0.2$.

First calculate the belief function and plausibility function for singletons

$$\begin{aligned} Bel(\{\theta_1\}) &= m(\{\theta_1\}) = 0.3, \\ Bel(\{\theta_2\}) &= m(\{\theta_2\}) = 0.0, \\ Bel(\{\theta_3\}) &= m(\{\theta_3\}) = 0.0, \\ Pl(\{\theta_1\}) &= m(\{\theta_1\}) + m(\Theta) = 0.5, \\ Pl(\{\theta_2\}) &= m(\{\theta_2, \theta_3\}) + m(\Theta) = 0.7, \\ Pl(\{\theta_3\}) &= m(\{\theta_2, \theta_3\}) + m(\Theta) = 0.7. \end{aligned}$$

Then, calculate the distance between each belief interval of singleton and the interval $[0, 1]$

$$\begin{aligned} d^l([Bel(\{\theta_1\}), Pl(\{\theta_1\})], [0, 1]) &= d^l([0.3, 0.5], [0, 1]) \\ &= \sqrt{\left[\frac{0.3+0.5}{2} - \frac{0+1}{2}\right]^2 + \frac{1}{3}\left[\frac{0.5-0.3}{2} - \frac{1-0}{2}\right]^2} = 0.2517; \\ d^l([Bel(\{\theta_2\}), Pl(\{\theta_2\})], [0, 1]) &= d^l([0, 0.7], [0, 1]) \\ &= \sqrt{\left[\frac{0+0.7}{2} - \frac{0+1}{2}\right]^2 + \frac{1}{3}\left[\frac{0.7-0}{2} - \frac{1-0}{2}\right]^2} = 0.1732; \\ d^l([Bel(\{\theta_3\}), Pl(\{\theta_3\})], [0, 1]) &= d^l([0, 0.7], [0, 1]) \\ &= \sqrt{\left[\frac{0+0.7}{2} - \frac{0+1}{2}\right]^2 + \frac{1}{3}\left[\frac{0.7-0}{2} - \frac{1-0}{2}\right]^2} = 0.1732; \end{aligned}$$

then $TU^l = 1 - \sqrt{3}/3 \cdot (0.2517 + 0.1732 + 0.1732) = 0.6547$.

4.3. Desired properties of the new total uncertainty measure

4.3.1. Range

The range of TU^l is $[0, 1]$, which is desired for practical use. For a vacuous BBA $m(\Theta) = 1$, all the belief intervals for singletons are $[0, 1]$. Therefore, by using Eq. (13), its degree of total uncertainty is 1, which is the maximum. For a categorical BBA $m(\theta_i) = 1$, the belief interval for θ_i is $[1, 1]$ and those of other singletons are $[0, 0]$. Therefore, by using Eq. (13), its degree of total uncertainty is 0, which is the minimum.

Proof of the range: For a vacuous BBA, when using Eq. (13), $TU^l = 1$. On the other hand, if $TU^l = 1$, i.e., all the belief intervals for singletons are $[0, 1]$, the corresponding BBA must be a vacuous one. So the maximum value is unique. If a BBA is not a vacuous one, then

¹ The distance $d^l([Bel(\{\theta_i\}), Pl(\{\theta_i\})], [0, 1])$ reaches its maximum value when the belief interval is $[0, 0]$ or $[1, 1]$. Therefore, the normalization factor is $1/d^l([0, 0], [0, 1]) = 1/d^l([1, 1], [0, 1]) = \sqrt{3}$.

some singletons' belief intervals can be strictly contained by $[0, 1]$. Therefore, according to Eq. (13), the corresponding TU^I will be smaller than 1. Therefore, the unique maximum value of TU^I is 1.

For a categorical BBA $m(\theta_i) = 1$, when using Eq. (13), $TU^I = 0$. On the other hand, if $TU^I = 0$, the only possible belief interval for singletons are either $[0, 0]$ or $[1, 1]$. For a BBA, one and only one belief interval for singleton could be $[1, 1]$. So, the corresponding BBA should be $m(\{\theta_i\}) = 1, m(\{\theta_j\}) = 0, \forall j \neq i$, which represents the clearest case. If a BBA is not $m(\{\theta_i\}) = 1, m(\{\theta_j\}) = 0, \forall j \neq i$, some singletons' belief intervals are neither $[0, 0]$ nor $[1, 1]$. Therefore, according to Eq. (13), the corresponding TU^I will be larger than 0. So, the minimum value of TU^I is 0.

End of Proof

4.3.2. Monotonicity

Given an uncertainty measure UN and two BBAs m_1 and m_2 , if

$$\forall A \in \mathcal{P}(\Theta) : Pl_1(A) \leq Pl_2(A), Bel_1(A) \geq Bel_2(A) \quad \text{or}$$

$$\forall A \in \mathcal{P}(\Theta) : [Bel_1(A), Pl_1(A)] \subseteq [Bel_2(A), Pl_2(A)]$$

there exists $UN(m_1) \leq UN(m_2)$, then it is said that UN satisfies the monotonicity [29]. The physical meaning of the monotonicity is that an uncertainty measure in the theory of belief functions must not decrease the total quantity of uncertainty in situations where there is a clear decrease in information (increment of uncertainty).

Proof of the monotonicity: For two BBAs defined over the same FOD, if it holds that $\forall A \in \mathcal{P}(\Theta) : [Bel_1(A), Pl_1(A)] \subseteq [Bel_2(A), Pl_2(A)]$, then, there exists $\forall \theta_i \in \Theta : [Bel_1(\{\theta_i\}), Pl_1(\{\theta_i\})] \subseteq [Bel_2(\{\theta_i\}), Pl_2(\{\theta_i\})]$. According to the property of the distance of interval numbers in Eq. (12), there exists $d^I([Bel_1(\{\theta_i\}), Pl_1(\{\theta_i\})], [0, 1]) \geq d^I([Bel_2(\{\theta_i\}), Pl_2(\{\theta_i\})], [0, 1])$.

Then, according to Eq. (13), $TU^I(m_1) \leq TU^I(m_2)$, i.e., the monotonicity holds.

End of Proof

4.4. On other properties

It is declared in [11,29] that AU and AM satisfy the probability consistency and set consistency.

Probability consistency: When m is a Bayesian BBA (all focal elements are singletons), AU reduces to Shannon entropy in probability theory

$$AU(m) = AM(m) = - \sum_{\theta \in \Theta} m(\{\theta\}) \log_2(m(\{\theta\}))$$

Set consistency: When m focuses on a single set (i.e., $m(A) = 1, \forall A \subseteq \Theta$), AU reduces to a Hartley measure in classical theory

$$AU(m) = AM(m) = \log_2(|A|)$$

It should be noted that it is meaningless to apply these two requirements or properties to our new measure. First, our new TU^I is designed directly in the framework of belief functions theory, but not a generalization of the measures in probability framework or those in classical set theory. Second, the theory of belief functions is not a successful generalization of the probability theory. It cannot degenerate back to the probability framework. For example, a Bayesian BBA is not a probability but still a special BBA, which cannot satisfy the additivity for mutually exclusive events and cannot satisfy the unity for the total set in Kolmogorov probability axioms [33]. Therefore, the validity and strictness of such a consistency is doubtful.

Furthermore, AU also satisfies the sub-additivity and additivity, which defined for the joint BBA in Cartesian space. In our work, we do not concern the joint BBA in Cartesian space, therefore, we will not analyze the sub-additivity and additivity in this paper.

We will further verify our new total uncertainty measure by examples and simulations in the next section.

5. Experiments and simulations

5.1. Example 3

This Example 3 is a revisiting of Example 1. Using Eq. (13), we can obtain that $TU^I(m_1) = 0.6667$ and $TU^I(m_2) = 0.7778$, i.e., m_2 has a higher level of uncertainty, which is intuitive. It can be proved that TU^I satisfy the monotonicity. The proof can be found at Section 4.3.2.

5.2. Example 4

This Example 4 is a revisiting of Example 2. Our new total uncertainty measure TU^I , non-specificity measure NS , dissonance measure $Diss$ are also calculated. The results are shown in Fig. 2.

AU never changes as analyzed in Example 2, which is counter-intuitive. AM values never changes, so far as $a = b$ holds, which is also counter-intuitive as aforementioned in Example 2. As we can see in Fig. 2, TU^I brings out rational results. It reaches the maximum when $a = b = 0$, i.e., $m(\Theta) = 1$. When $a + b$ is fixed, the maximum values are reached if $a = b$ holds. For example, suppose that $a + b = 0.3$. The TU^I values are shown in Fig. 3. This makes sense, because $m(\Theta) = 1 - (a + b)$ is fixed, the non-specificity part is fixed. When $a = b$, the conflict part reaches its maximum value.

Note that NS reaches 0 (minimum), when $a = b = 0.5$ ($m(\Theta) = 0$); it reaches 1 (maximum), when $a = b = 0$ ($m(\Theta) = 1$). $Diss$ reaches 1 (maximum), when $a = b = 0.5$; it reaches 0 (minimum), when $a = 0$ or $b = 0$ (no conflict).

5.3. Example 5

Suppose that the FOD is $\Theta = \{\theta_1, \theta_2, \theta_3\}$ and a BBA over Θ is $m(A) = 1, \forall A = \Theta$. We make changes to the BBA step by step. In each step, $m(\Theta)$ has a decrease of $\Delta = 0.05$ and each singleton mass $m(\{\theta_i\}), i = 1, 2, 3$ has an increase of $\Delta/3$. Finally, $m(\Theta)$ reaches 0 and $m(\{\theta_i\}) = 1/3$. We calculate total uncertainty measures including TU^I , AM , AU , non-specificity NS , and dissonance ($Diss$) at each step. The changes of uncertainty values are shown in Fig. 4.

As we can see in Fig. 4, with the change of the BBA in each step, the non-specificity changes from the maximum to zero. This is intuitive, because the mass assignments are transferred from the focal elements with larger cardinality to those with smaller cardinality. The measure for conflict, i.e., $Diss$ here, becomes larger and reaches the maximum in the final. This is also intuitive, because the BBA changes from a vacuous BBA (no conflict) to a Bayesian BBA (with uniform distribution on all singletons) in the final. The traditional total uncertainty measures AU and AM never change and they are always at the maximum value. This is counter-intuitive because the total uncertainty degree of a vacuous BBA should be the largest, and thus, it should be larger than that of a Bayesian BBA (even if, it is a Bayesian BBA with uniform distribution on all singletons). The reasons for the irrational results of AU and AM are analyzed as follows. They are both defined based on some probabilistic transformation from BBAs. In this example, for the probabilistic transformation used in AM and AU , the results are always a uniformly distributed probability mass function (p.m.f.): $P(\theta_i) = 1/3, i = 1, 2, 3$, therefore, AM and AU will never change here. Our new measure TU^I provides rational results. It becomes smaller and smaller when the BBA changes from a vacuous one to a Bayesian one as shown in Fig. 2. Therefore, according to our opinion, it is inappropriate to define total uncertainty measure in the theory of belief functions by indirectly using the uncertainty measure in probability framework, i.e., Shannon entropy. It should be better not to switch the framework but to directly design in the framework of belief functions theory. This is also the motivation of our work in this paper.

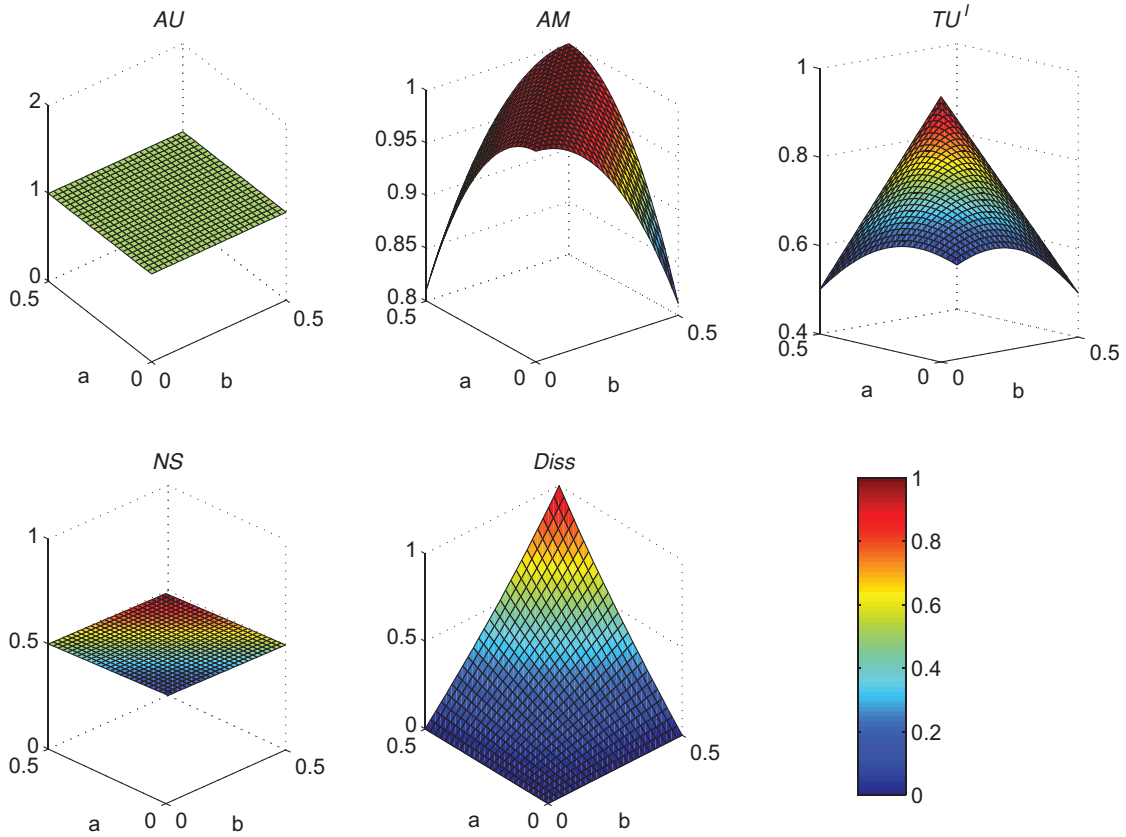


Fig. 2. The change of total uncertainty measures in Example 4.

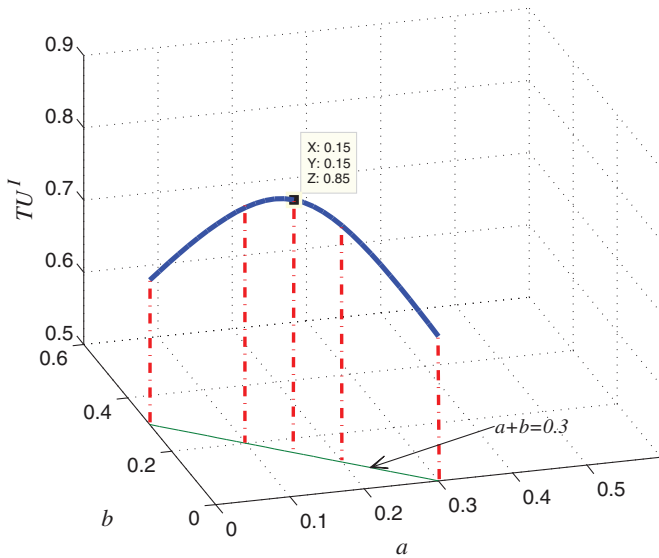


Fig. 3. TU^I for those with fixed $a + b$.

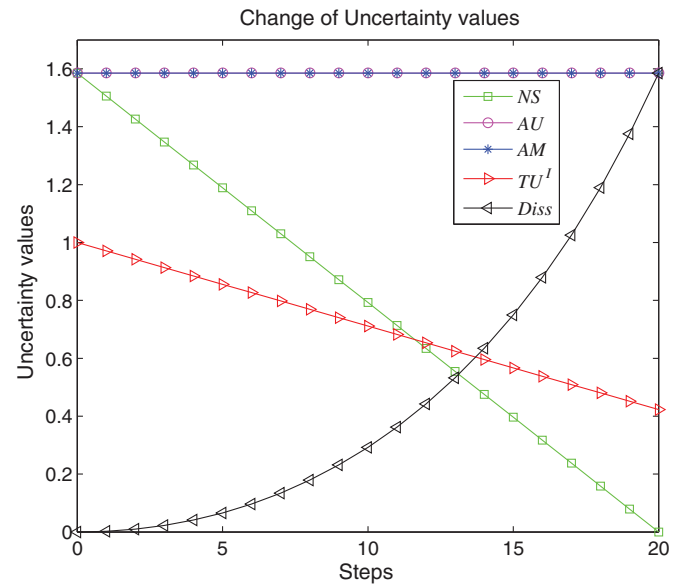


Fig. 4. The change of total uncertainty measures in Example 5.

5.4. Example 6

The FOD is $\Theta = \{\theta_1, \theta_2, \theta_3\}$. A BBA over Θ is $m(A) = 1, \forall A = \Theta$. We make changes to the BBA step by step. In each step, $m(\Theta)$ has a decrease of $\Delta = 0.05$ and mass assignment of one singleton $m(\{\theta_1\})$ has an increase of $\Delta = 0.05$. In the final step, $m(\Theta) = 0$ and $m(\{\theta_1\}) = 1$. We calculate total uncertainty measures including TU^I , AM, AU, non-specificity measure NS, and dissonance (Diss) of the BBA at each step. The changes of these uncertainty values are shown in Fig. 5.

The BBA changes from a vacuous one to a categorical one. As shown in Fig. 5, non-specificity NS becomes smaller and smaller, which is intuitive. The conflict part Diss is always 0, which is also intuitive, because at most two focal elements are available for the BBA in each step (including $\{\theta_1\}$ and Θ), therefore, there is no conflict. AU, AM, TU^I all decrease intuitively. However, AU is not sensitive to the BBA's change at the beginning stage (nearly unchanged). This has already been criticized by Jousselme et al. [11].

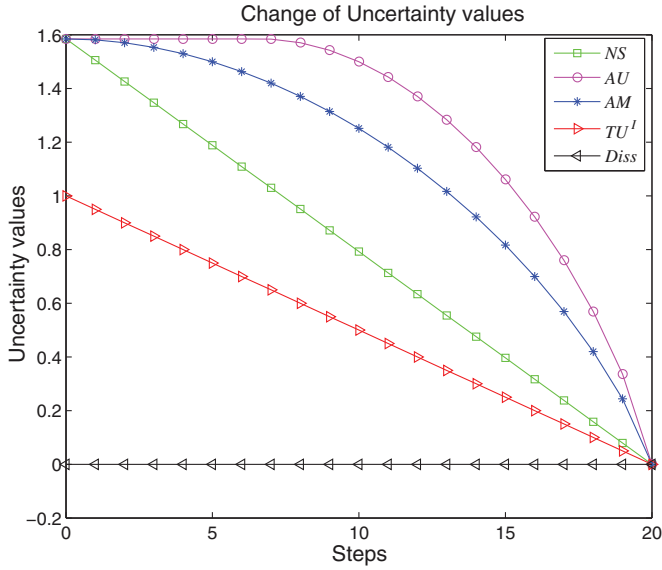


Fig. 5. The change of total uncertainty measures in Example 6.

5.5. Example 7

Examples 7–9 are used for reference from [11]. Suppose that the FOD is $\Theta = \{\theta_1, \dots, \theta_8\}$. A BBA over Θ is $m(A) = 1, \forall A = \Theta$. We make changes to the BBA step by step. In each step, $m(\Theta)$ has a decrease of $\Delta = 0.05$, and the mass assignment for one focal element B with cardinality $s < 8$ has an increase of $\Delta = 0.05$. In the final step, $m(B) = 1$. Here, we can set $s = 2, 4, 6$, respectively. Given an s value, we repeat the whole procedure of BBA change. Under the different s , the changes of values for uncertainty measures including TU^I , AM , AU , NS , and $Diss$ are shown in Fig. 6.

As shown in Fig. 6, when the mass assignments are transferred to a smaller size focal element, the non-specificity NS intuitively decreases. The total uncertainty AU , AM , TU^I all decrease. If s has a smaller value, the total uncertainty and the non-specificity will decrease faster, because the mass assignments are transferred to a smaller size focal element. Comparatively, AU is not sensitive to the change of the BBA. The conflict part ($Diss$) is always 0. The reason is analyzed as follows. At the beginning, the BBA is a vacuous one; in the middle steps, the BBA has two focal elements including the total sets; in the final step, it has only one focal element. Therefore, there is no conflict.

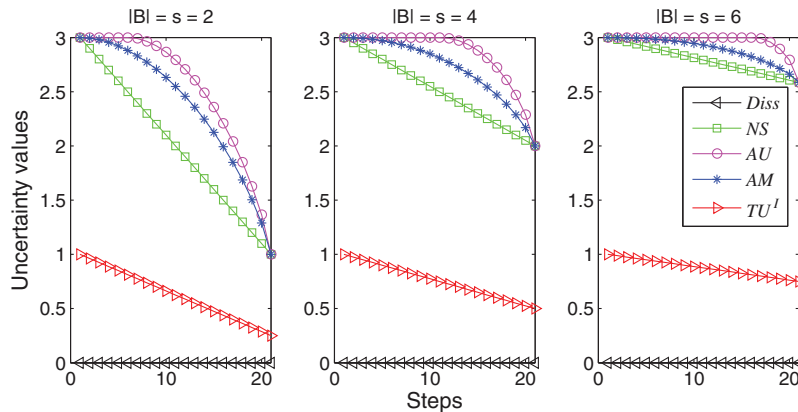


Fig. 6. The change of total uncertainty measures in Example 7.

Table 1

Algorithm 1: Random generation of BBA.

Input: Θ : FOD; N_{\max} : Maximum number of focal elements;
Output: m : a BBA
Generate the power set of Θ : $\mathcal{P}(\Theta)$;
Generate a random permutation of $\mathcal{P}(\Theta) \rightarrow \mathcal{R}(\Theta)$;
Generate an integer between 1 and $N_{\max} \rightarrow k$;
FOR Each First k elements of $\mathcal{R}(\Theta)$ **do**
Generate a value within $[0, 1] \rightarrow mv(i), i = 1, \dots, k$;
END
Normalize the vector $mv = [mv(1), \dots, mv(2)] \rightarrow mv'$
 $m(A_j) = mv'(j), j = 1, \dots, k$.

5.6. Example 8

Suppose that the FOD is $\Theta = \{\theta_1, \dots, \theta_{10}\}$. A BBA over Θ is $m(\Theta) = 1 - a$, and $m(A) = a, |A| = s \leq 10$. Given an a , and the initial $|A| = 1$, we make changes to the BBA step by step. $|A|$ increases by 1 in each step. In the final, the BBA becomes a vacuous one. We set $a = 0.3, 0.5, 0.8$, respectively. Given an a value, we repeat the whole procedure of BBA change. Under the different a , the changes of values for uncertainty measures including TU^I , AM , AU , NS , and $Diss$ are shown in Fig. 7.

As shown in Fig. 7, non-specificity NS , total uncertainty measures including AU , AM , and our TU^I all intuitively increase with the change of BBA at each step. If a has a larger value, they will increase faster, because relatively more mass assignments are transferred to a larger size focal element. Comparatively, AU is not sensitive to the BBA's change. The conflict part ($Diss$) is zero and never changes. The reason is analyzed as follows. Before the final step, the BBA has two focal elements including the total sets; in the final step, it has only one focal element. Therefore, there is no conflict.

5.7. Example 9

Suppose that the size of an FOD is $|\Theta| = 5$. Randomly generate 10 BBAs with $k, 1 \leq k \leq 2^5 - 1$ focal elements using the algorithm [11] in Table 1.

In this simulation, the number of focal elements is set to 15. Those generated BBAs are then combined one by one using Dempster's rule of combination in Eq. (4). At each instant t of the evidence combination, the values of NS , AU , AM , $Diss$, and TU^I are calculated. The whole procedure is repeated for 100 times and the averaging values of these different uncertainty measures at different t are shown in Fig. 8. It is shown that all the uncertainty measures, except for $Diss$, decrease with the increase of the combination steps. This makes sense, because it is intuitive that the total uncertainty decreases in the information fusion procedure such as the evidence combination. The

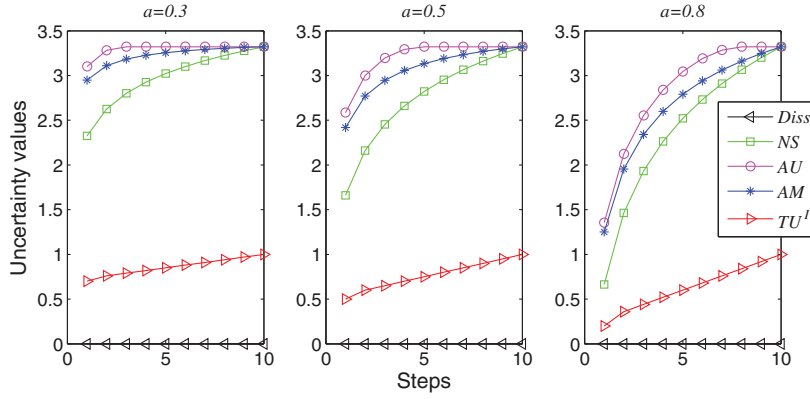


Fig. 7. The change of total uncertainty measures in Example 8.

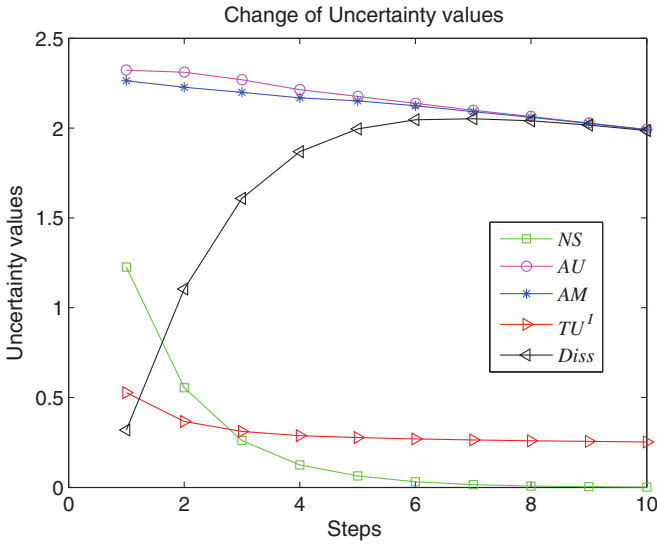


Fig. 8. The change of total uncertainty measures in Example 8.

non-specificity measure drops faster and more significantly than the total uncertainty measures. This is because that in the evidence combination based on Dempster's rule, the focal elements are split into focal elements with smaller cardinality. Also, since the focal elements are split into smaller size focal elements, the conflict in the BBA will significantly increase at the beginning steps. At the final steps, since there is less further split of focal elements and the consensus effect in combination becomes more dominant, the conflict in the BBA decreases slightly as shown in Fig. 8.

6. Application of the new total uncertainty measure

Here we use our new total uncertainty measure in feature evaluation for pattern classification to further show its rationality.

Three classes of samples are artificially generated. There are 200 samples in each class. Each sample has four dimensions. In each class, each dimension of samples is Gaussian distributed with different mean and standard deviation (Std for short) as shown in Fig. 9 and Table 2.

As shown in Fig. 9 and Table 2, since the three Gaussian probability density functions (PDFs) in feature 2 are quite well separated, the class discriminability of feature 2 is the best; the class discriminability of feature 4 is the worst (totally overlapped), and that of the feature 3 is also very bad. The class discriminability of feature 1 is in the middle. This can also be verified by using the discrimination

Table 2

Gaussian distribution parameters of the samples.

Features		Class 1	Class 2	Class 3
Feature 1	Mean	0	4	5.5
	Std	1	1.2	1.2
Feature 2	Mean	0	5	8
	Std	1	1.2	1.2
Feature 3	Mean	4.8	5	5.2
	Std	1.2	1.2	1.2
Feature 4	Mean	5	5	5
	Std	1.2	1.2	1.2

criterion [34] as follows:

$$J = \frac{\text{tr}(S_w)}{\text{tr}(S_b)} \quad (14)$$

where tr denotes the trace of a matrix. Suppose that there are C classes and each class C_i has N_i samples. The degree of inner-class cohesion S_w and the degree of inter-class separability S_b are as follows [34]:

$$\begin{cases} S_w = \sum_{i=1}^C P(C_i) E \left[\left(X - \frac{1}{N_i} \sum_{X \in C_i} X \right) \left(X - \frac{1}{N_i} \sum_{X \in C_i} X \right)^T \right] \\ S_b = \sum_{i=1}^C P(C_i) \left(\frac{1}{N_i} \sum_{X \in C_i} X - M \right) \left(\frac{1}{N_i} \sum_{X \in C_i} X - M \right)^T \end{cases} \quad (15)$$

where X is a feature (vector) of a sample and

$$M = \frac{1}{C} \sum_{i=1}^C \left(\frac{1}{N_i} \sum_{X \in C_i} X \right) \quad (16)$$

is the mean of all the classes' centroids. If J of some feature (or set of features) in Eq. (14) is smaller, then such a feature (or set of features) is more crisp and discriminable.

For the four dimensions of the artificially generated samples, there exists

$$J(1) = 0.2666, J(2) = 0.1336, J(3) = 488.9156, J(4) = 981.3525$$

It means that in terms of discriminability, feature 2 is the best; feature 4 is the worst; feature 3 is also very bad, and feature 1 is in the middle, which is accordant to the judgment based on the Gaussian parameters.

Now, we use different total uncertainty measures (including AM , AU , and our new TU^I) for the feature evaluation. Intuitively, the feature with smaller total uncertainty measure should be better (higher discriminability).

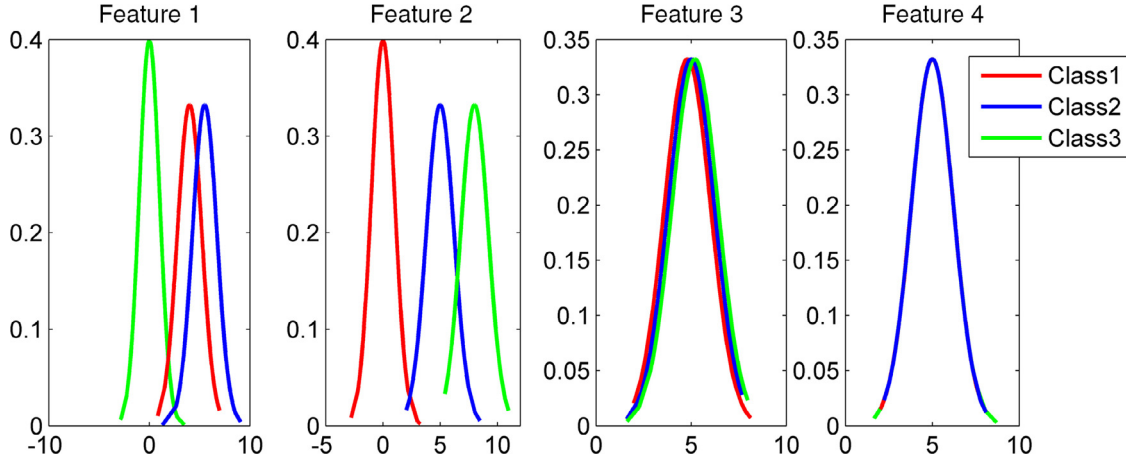


Fig. 9. Probability density functions (PDFs) of different features of the samples in three classes.

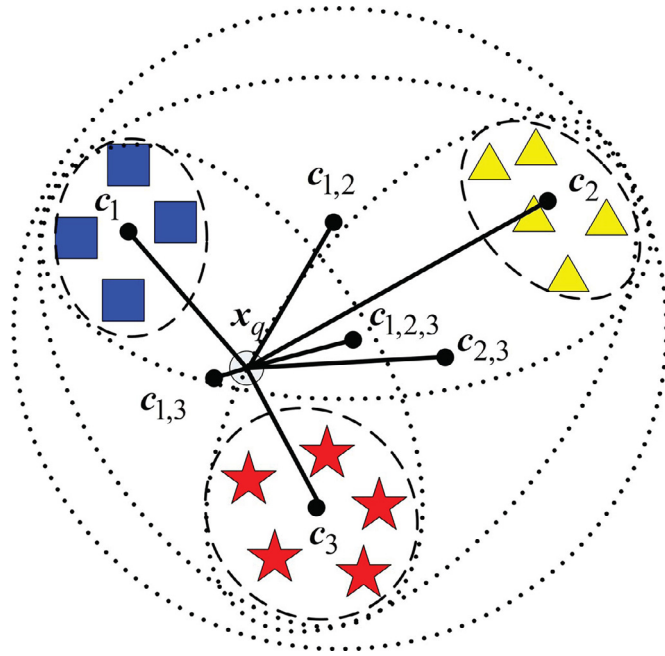


Fig. 10. Illustration of BBA generation.

First, we generate BBA from each sample x_q in the three-class artificial samples on different feature $i \in \{1, 2, 3\}$ according to [35]

$$m_{x_q}^i(A_j) = \frac{|A_j|^{-\alpha/(\beta-1)} d_{qj}^{-2/(\beta-1)}}{\sum_{A_k \neq \emptyset} |A_k|^{-\alpha/(\beta-1)} d_{qk}^{-2/(\beta-1)} + \delta^{-2/(\beta-1)}} \quad (17)$$

where parameters $\alpha = 1$, $\beta = 2$ as suggested in [34]. Here we give an illustrative example as in Fig. 10.

In Fig. 10, three different colors represent three different classes. c_1 denotes the centroid of samples in class 1; $c_{1,2}$ denotes the centroid of samples in class 1 and class 2; $c_{1,2,3}$ denotes the centroid of samples in classes 1, 2, and 3. Calculate the distance d between x_q and those centroids of single classes and compound classes. Then according to Eq. (17), the BBA can be generated.

Second, we calculate AU, AM, and our proposed TU^l for all the generated BBAs (corresponding to each sample on different features). Then, calculate the average values of AU, AM and TU^l for different features. The results are shown in Table 3.

Table 3

Average total uncertainty measures.

Measures (Ave)	Feature 1	Feature 2	Feature 3	Feature 4
AU	1.4022	1.1968	1.5850	1.5850
AM	1.1855	0.9860	1.5844	1.5847
TU^l	0.4496	0.3787	0.6218	0.6221

We consider the average total uncertainty measures as scores for different features. The feature with smaller score is better (with high discriminability)

As shown in Table 3 the feature evaluation based on the average TU^l values is accordant to the results obtained using discrimination criterion. That is, feature 2 is the best; feature 4 is the worst; feature 3 is also very bad, and feature 1 is in the middle.

AMcan also provide the correct feature evaluation result. Note that feature 3 is more ambiguous than feature 4. The difference of average TU^l for features 3 and 4 is 0.0003; The difference of average AM for features 3 and 4 is also 0.0003. However, the value range of TU^l is $[0, 1]$, while the value range of AM is $[0, \log_2 3]$. Therefore, our proposed TU^l is more sensitive when measuring the ambiguity in different features, which agrees with the analysis in above examples.

The average AU values for both features 3 and 4 are equal. They are both the maximum value $\log_2 3 = 1.5850$. The reason is analyzed below.

According to the BBA generation approach used, because in features 3 and 4, all the centroids (7 centroids, 3 singleton classes, and 4 compound classes) are very close, all the distances between x_q and those centroids are almost equal when generating BBAs for x_q on features 3 and 4. Therefore, each focal element of the BBA generated for x_q on features 3 and 4 has the mass value approximating to $1/7$. According to all the BBAs generated, for feature 3, $m(A) = (1/7) \pm 0.010$; for feature 4, $m(A) = (1/7) \pm 0.007$, $A \subseteq \{c_1, c_2, c_3\}$. According to Eq. (10), since AU is calculated based on the maximization of Shannon entropy given the constraints established using belief functions as

$$\begin{aligned} \underbrace{\frac{1}{7}}_{Bel(\{c_1\})} &\approx m(\{c_1\}) \leq P(c_1) \\ &\leq \underbrace{m(\{c_1\}) + m(\{c_1, c_2\}) + m(\{c_1, c_3\}) + m(\{c_1, c_2, c_3\})}_{Pl(\{c_1\})} \approx \frac{4}{7} \end{aligned}$$

$$\begin{aligned}
\frac{1}{7} &\approx \underbrace{m(\{c_2\})}_{\text{Bel}(\{c_2\})} \leq P(c_2) \\
&\leq \underbrace{m(\{c_2\}) + m(\{c_1, c_2\}) + m(\{c_2, c_3\}) + m(\{c_1, c_2, c_3\})}_{\text{Pl}(\{c_2\})} \approx \frac{4}{7} \\
\frac{1}{7} &\approx \underbrace{m(\{c_3\})}_{\text{Bel}(\{c_3\})} \leq P(c_3) \\
&\leq \underbrace{m(\{c_3\}) + m(\{c_1, c_3\}) + m(\{c_2, c_3\}) + m(\{c_1, c_2, c_3\})}_{\text{Pl}(\{c_3\})} \approx \frac{4}{7}
\end{aligned}$$

$P(c_1) = P(c_2) = P(c_3) = 1/3$, which corresponds to the maximum Shannon entropy, always satisfies the constraints, so, the AU value is fixed to $\log_2 3 = 1.5850$. Based on AU , we cannot judge which one is better for features 3 and 4.

In summary, our proposed TU^I can be well used in feature evaluation for pattern classification, and it performs better than AM and AU .

7. Conclusions

In this paper, a new total uncertainty measure based on the distance of belief intervals is proposed in the framework of belief functions theory. There is no switch between frameworks in the new measure. It is based on a new perspective that for a body of evidence, the farther it is from the most uncertain case, the less uncertainty it carries. It has been experimentally shown that our new measure can rationally represent the total uncertainty in BOEs. The new measure also has some desired properties and behaves more rational than traditional measures like AU and AM in some cases as shown in our experiments and simulations. Furthermore, our new measure can provide more rational results in practical applications such as the feature evaluation.

Till now, the theory of belief functions is still not so solid and needs further refining. The available works mainly focus on the uncertainty modeling and reasoning in belief functions theory and there have emerged many valuable results. However, the performance evaluation in belief functions theory is far from mature, which has become its bottle neck for the further development. In future work, we will focus on performance evaluation in the theory of belief functions and try to use our new total uncertainty measure to implement the theoretical evaluation and the application-based evaluation in the theory of belief functions.

Acknowledgments

This work was supported by the Grant for State Key Program for Basic Research of China (973) (No. 2013CB329405), National Natural Science Foundation (Nos. 61203222, 61573275), Foundation for Innovative Research Groups of the National Natural Science Foundation of China (No. 61221063), Science & Technology Project of Shaanxi Province (No. 2013KJXX-46), Specialized Research Fund for the Doctoral Program of Higher Education (No. 20120201120036), and Fundamental Research Funds for the Central Universities (No. xjj2014122).

References

- [1] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press Princeton, 1976.
- [2] G. Lin, J. Liang, Y. Qian, An information fusion approach by combining multigranulation rough sets and evidence theory, *Inf. Sci.* 314 (2015) 184–199.
- [3] G. Dong, G. Kuang, Target recognition via information aggregation through Dempster–Shafer's evidence theory, *IEEE Geosci. Remote Sens. Lett.* 12 (6) (2015) 1247–1251.
- [4] Z.-g. Liu, Q. Pan, J. Dezert, Evidential classifier for imprecise data based on belief functions, *Knowl. Based Syst.* 52 (2013) 246–257.
- [5] Z.-g. Liu, Q. Pan, J. Dezert, G. Mercier, Credal c-means clustering method based on belief functions, *Knowl. Based Syst.* 74 (2015) 119–132.
- [6] O. Basir, X. Yuan, Engine fault diagnosis based on multi-sensor information fusion using Dempster–Shafer evidence theory, *Inf. Fusion* 8 (4) (2007) 379–386.
- [7] D. Han, J. Dezert, J.-M. Tacnet, C. Han, A fuzzy-cautious OWA approach with evidential reasoning, in: *Proceedings of the 15th International Conference on Information Fusion (FUSION)*, ISIF, Singapore, 2012, pp. 278–285.
- [8] R.R. Yager, N. Alajlan, Dempster–Shafer belief structures for decision making under uncertainty, *Knowl. Based Syst.* 80 (2015) 58–66.
- [9] G.J. Klir, B. Parviz, A note on the measure of discord, in: *Proceedings of the Eighth International Conference on Uncertainty in Artificial Intelligence*, Morgan Kaufmann Publishers Inc., 1992, pp. 138–141.
- [10] D. Dubois, H. Prade, A note on measures of specificity for fuzzy sets, *Int. J. Gen. Syst.* 10 (4) (1985) 279–283.
- [11] A.-L. Jousselme, C. Liu, D. Grenier, É. Bossé, Measuring ambiguity in the evidence theory, *IEEE Trans. Syst. Man Cybern. Part A Syst. Hum.* 36 (5) (2006) 890–903.
- [12] U. Höhle, Entropy with respect to plausibility measures, in: *Proceedings of the 12th IEEE International Symposium on Multiple-Valued Logic*, 1982, pp. 167–169.
- [13] R.V. Hartley, Transmission of information, *Bell Syst. Tech. J.* 7 (3) (1928) 535–563.
- [14] R.R. Yager, Entropy and specificity in a mathematical theory of evidence, *Int. J. Gen. Syst.* 9 (4) (1983) 249–260.
- [15] R. Körner, W. Näther, On the specificity of evidences, *Fuzzy Sets Syst.* 71 (2) (1995) 183–196.
- [16] D. Harmanec, G.J. Klir, Measuring total uncertainty in Dempster–Shafer theory: a novel approach, *Int. J. Gen. Syst.* 22 (4) (1994) 405–419.
- [17] C.E. Shannon, The mathematical theory of communication, *Bell Syst. Tech. J.* 27 (1949) 379–423, 623–656.
- [18] G. Klir, H.W. Lewis, III, Remarks on “measuring ambiguity in the evidence theory”, *IEEE Trans. Syst. Man Cybern. Part A Syst. Hum.* 38 (4) (2008) 995–999.
- [19] J. Pearl, Reasoning with belief functions: an analysis of compatibility, *Int. J. Approx. Reason.* 4 (5) (1990) 363–389.
- [20] J. Dezert, A. Tchamova, On the validity of Dempster's fusion rule and its interpretation as a generalization of Bayesian fusion rule, *Int. J. Intell. Syst.* 29 (3) (2014) 223–252.
- [21] F. Smarandache, J. Dezert, *Advances and Applications of DSMT for Information Fusion—Collected Works*, 3, American Research Press, 2009.
- [22] J.F. Lemmer, Confidence factors, empiricism and the Dempster–Shafer theory of evidence, in: *Proceedings of the First Conference on Uncertainty in Artificial Intelligence (UAI-85)*, 1985, pp. 160–176.
- [23] L.A. Zadeh, A simple view of the Dempster–Shafer theory of evidence and its implication for the rule of combination, *AI Mag.* 7 (2) (1986) 85–90.
- [24] P. Wang, A defect in Dempster–Shafer theory, in: *Proceedings of the Tenth International Conference on Uncertainty in Artificial Intelligence*, Morgan Kaufmann Publishers Inc., 1994, pp. 560–566.
- [25] A. Gelman, The boxer, the wrestler, and the coin flip: a paradox of Bayesian inference, robust Bayes, and belief functions, *Am. Stat.* 60 (2) (2005) 146–150.
- [26] G.J. Klir, A. Ramer, Uncertainty in the Dempster–Shafer theory: a critical re-examination, *Int. J. Gen. Syst.* 18 (2) (1990) 155–166.
- [27] D. Dubois, H. Prade, Properties of measures of information in evidence and possibility theories, *Fuzzy Sets Syst.* 100 (1999) 35–49.
- [28] A. Ramer, Uniqueness of information measure in the theory of evidence, *Fuzzy Sets Syst.* 24 (2) (1987) 183–196.
- [29] J. Abellán, A. Masegosa, Requirements for total uncertainty measures in Dempster–Shafer theory of evidence, *Int. J. Gen. Syst.* 37 (6) (2008) 733–747.
- [30] P. Smets, R. Kennes, The transferable belief model, *Artif. Intell.* 66 (2) (1994) 191–234.
- [31] A. Irpino, R. Verde, Dynamic clustering of interval data using a Wasserstein-based distance, *Pattern Recognit. Lett.* 29 (11) (2008) 1648–1658.
- [32] L. Tran, L. Duckstein, Comparison of fuzzy numbers using a fuzzy distance measure, *Fuzzy* 130 (3) (2002) 331–341.
- [33] A. Papoulis, S.U. Pillai, *Probability, Random Variables, and Stochastic Processes*, fourth, Tata McGraw-Hill Education, 2002.
- [34] R.O. Duda, P.E. Hart, D.G. Stork, *Pattern Classification*, John Wiley & Sons, 2012.
- [35] M.-H. Masson, T. Denoeux, ECM: An evidential version of the fuzzy c-means algorithm, *Pattern Recognit.* 41 (4) (2008) 1384–1397.