The information volume of uncertain information: (5) Divergence measure

Lipeng Pan[a], Yong Deng[a,b,c,*]

[a] *Institute of Fundamental and Frontier Science, University of Electronic Science and Technology of China, Chengdu, 610054, China*
[b] *School of Eduction, Shannxi Normal University, Xi'an, 710062, China*
[c] *Department of Biostatics, Medical Center, Vanderbilt University, Nashville, TN. 37235, USA*

Abstract

Dempster-Shafer Evidence theory is an extension of probability theory, which can describe uncertain information more reasonably. Divergence measure is always an important concept in probability theory. Therefore, how to propose a reasonable divergence measurement has always been a research hot spot in evidence theory. Recently, Deng proposed the concept of information volume based on Deng entropy. It is interesting to note that compared with the uncertainty measure of Deng entropy, information volume of Deng entropy contains more information. Obviously, it might be more reasonable to use information volume of Deng entropy to represent uncertainty information. Based on this, in the paper, we combined the characteristics of non-specific measurement of Deng entropy, and propose a new divergence measure. The new divergence measurement not only satisfies the axiom of distance measurement, but also has some advantages that cannot be ignored. In addition, when the basic probability assignment(BPA) degenerates into probability distribution, the measured result of the new divergence measure is the same as that of the traditional Jensen-Shannon divergence. If the mass function is assigned in probability distribution, the proposed divergence is degenerated as Kullback-Leibler divergence. Finally, some numerical examples are illustrated to show the efficiency of the proposed

*Corresponding author
Email address: dengentropy@uestc.edu.cn; prof.deng@hotmail.com (Yong Deng)

divergence measure of information volume.

Keywords: Mass function, probability distribution, information volume, Jensen-Shannon divergence, Kullback-Leibler divergence, divergence measure of mass function

1. Introduction

In engineering practice and scientific research, a lot of uncertain information often appears. To deal with this uncertain information, researchers have proposed a series of theories, such as probability theory[1], Dempster-Shafer Evidence theory[2, 3], fuzzy sets[4], intuitive fuzzy sets[5], and Pythagorean fuzzy sets[6] and so on[7, 8]. It's worth noting that since the evidence theory satisfies the weaker condition than the probability theory, it can describe the uncertain information more reasonably. It has been concerned by researchers and applied to target recognition[9, 10], decision-making[11, 12] and so on[13, 14].

In Dempster-Shafer Evidence theory, there are many concepts reflected from probability theory, such as information entropy, cross entropy, conditional entropy, Jensen-Shannon divergence. Among them, the researchers always paid attention to the divergence of two basic probability assignment functions, because Jensen-Shannon divergence could be used to measure the difference between two basic probability assignment(BPAs). Recently, Deng proposed a concept of information quantity based on Deng entropy[15]. It can be written like this, given a BPA, the corresponding information volume is larger than its uncertainty measured by Deng entropy. The so called Deng distribution is defined as the BPA condition of the maximum Deng entropy. The information volume of Deng distribution is called the maximum information volume, which is lager than the maximum Deng entropy. In addition, both the total uncertainty case and the Deng distribution have the same information volume. Obviously, information volume would be a more reasonable description of uncertain information.

Therefore, in this paper, we consider not only the information volume of BPA, but also the non-specific characteristics of Deng entropy, and propose a new divergence measure. The new divergence measure has some advantages, such as satisfying the axiom of distance measurement, high resolution and compatibility with traditional Jensen-Shannon divergence. In addition, some numerical examples are used to explain these advantages.

The rest of this paper is organized as follows. In section **2**, some preliminaries are briefly reviewed. In section **3**, based on information volume, the new divergence measure is proposed. In section **4**, numerical examples are expounded to illustrated the proposed method. In section **5**, we have a brief conclusion.

2. Preliminaries

Several preliminaries are briefly introduced in this section, including basic probability assignment, information volume based on Deng entropy and Jensen-Shannon divergence.

2.1. Dempster-Shafer evidence theory

D-S evidence theory assigns probabilities to the power set of events [2, 3], so as to better grasp the unknown and uncertainty of the things, it offers a useful fusion tool for uncertain information. Some preliminaries in D-S theory are introduced as follows. For additional details about D-S evidence theory, refer to [2, 3].

Definition 1. *(Frame of discernment)*
For a mutually exclusive set Θ composed of A_i, it is defined as

$$\Theta = \{A_1, A_2, \cdots, A_n\} \tag{1}$$

The power set of Θ is indicated by 2^Θ :

$$2^\Theta = \{\phi, \{A_1\}, \{A_2\}, \cdots, \{A_1, A_2\}, \cdots, \Theta\} \tag{2}$$

Definition 2. *(Basic Probability Assignment)*
For a frame of discernment $\Theta = \{A_1, A_2, \cdots, A_n\}$, a mapping of m from 0 to 1

is defined as :

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

which satifies

$$m(\phi) = 0 \tag{4}$$

$$\sum_{A \subseteq \Theta} m(A) = 1 \tag{5}$$

where $m(A)$ represents the degree of evidence supports A.

Definition 3. *(Belief Function and Plausibility Function)*
For any $A \subseteq \Theta$, the belief function Bel :$2^{\Theta} \to [0, 1]$ is defined as

$$Bel(A) = \sum_{B \subseteq A} m(B) \tag{6}$$

The plausibility function Pl : $2^{\Theta} \to [0, 1]$ is defined as

$$Pl(A) = 1 - Bel(\bar{A}) = \sum_{B \cap A \neq \phi} m(B) \tag{7}$$

Obviously, $Pl(A) \geq Bel(A)$, the $Bel(A)$ and $Pl(A)$ are lower limit function and upper limit function of A. When A is a single set in power set 2^{Θ}, the equal sign is established.

2.2. Information volume of mass function

Recently, Deng poeposed the concept of information volume[15, 16, 17]. firstly, He define Deng distribution as the BPA condition of the maximum Deng entropy. Then, based on Deng entropy, the information volume of mass function is proposed.

The maximum Deng entropy and the BPA condition of it have been analyzed in [18]. However, the terminology, the BPA condition of the maximum Deng entropy, is not convenient for discussing. As a result, Deng distribution was defined as follows:

Definition 4. *(Deng distribution)[15]*

$$m_D(A) = \frac{(2^{|A|} - 1)}{\sum_{A \in 2^\Theta} (2^{|A|} - 1)} \tag{8}$$

which is the BPA distribution of the maximum Deng entropy. Namely, if and only if under this conditions, Deng entropy can reach its maximum value.

Definition 5. *(The information volume of mass function)[15]*
Let the frame of discernment be $\Theta = \{\theta_1, \theta_2, \theta_3, \cdots, \theta_N\}$. Use index i to denote the times of this loop, and use $m(A_i)$ to denote different mass function of different loops. Based on Deng entropy, the information volume of mass function can be calculated by following steps:

step 1: *Input mass function $m(A_0)$.*

step 2: *Continuously separate the mass function of the element whose cardinal is larger than 1 until convergence. Concretely, repeat the loop from step 2-1 to step 2-3 until Deng entropy is convergent.*

step 2-1: *Focus on the element whose cardinal is larger than 1, namely, $|A_i| > 1$. And then, separate its mass function based on the proportion of Deng distribution:*

$$m_D(A_i) = \frac{(2^{|A_i|} - 1)}{\sum_{A_i \in 2^\Theta} (2^{|A_i|} - 1)} \tag{9}$$

step 2-2: *Based on Deng entropy, calculate the uncertainty of all the mass functions except for those who have been divided. The result is denoted as $H_i(m)$.*

step 2-3: *Calculate $\Delta_i = H_i(m) - H_{i-1}(m)$. When Δ_i satisfies following condition, jump out of this loop.*

$$\Delta_i = H_i(m) - H_{i-1}(m) < \varepsilon \tag{10}$$

where ε is the allowable error.

step 3: *Output $H_{IV-mass}(m) = H_i(m)$, which is the information volume of the mass function.*

2.3. The JensenShannon divergence between two probability distributions

Lin introduced an information-theoretical based divergence measure between two or more probability distributions, called as JensenShannon (JS) divergence[19]. Unlike others divergence measures, the main property of JS divergence is that, it does not require the condition of absolute continuity for the probability distributions involved. JS divergence defines a true metric in the space of probability distributions actually it is the square of a metric [84]. The main concepts are defined as below.

Definition 6. *(The JensenShannon divergence between two probability distributions)*
Let us consider a discrete random variable X, and let $P_1 = \{p_{11}, p_{12}, \cdots, p_{1M}\}$ and $P_2 = \{p_{21}, p_{22}, \cdots, p_{2M}\}$ be two probability distributions for X. The Jensen-Shannon divergence between the probability distributions P_1 and P_2 is denoted as:

$$JS\left(P_1, P_2\right) = \frac{1}{2}\left[S\left(P_1, \frac{P_1 + P_2}{2}\right) + S\left(P_2, \frac{P_1 + P_2}{2}\right)\right] \tag{11}$$

where $S\left(P_1, P_2\right) = \sum_i p_{1i} log \frac{p_{1i}}{p_{2i}} \left(i = 1, 2, cdots, M\right)$ is the KullbackLeibler divergence and $\sum_i p_{ji} = 1 \left(i = 1, 2, \cdots, M; j = 1, 2\right)$

There are some properties for the JS divergence:

(1) $JS\left(P_1, P_2\right)$ is symmetric and always well defined;

(2) $JS\left(P_1, P_2\right)$ is bounded, $0 JS\left(P_1, P_2\right) 1$;

(3) its square root, $\sqrt{JS\left(P_1, P_2\right)}$ verifies the triangle inequality.

3. divergence measure of information volume

In DempsterShafer evidence theory, how to measure the discrepancy and conflict among evidences is still an open issue that is critical for the fusion of evidences. Obviously, DempsterShafer evidence theory is a generalization of probability theory. By integrating the DempsterShafer evidence theory with above mentioned JensenShannon divergence, based on the information volume and non-specific characteristics of deng entropy, a novel divergence measure named divergence measure of information volume(IV-JS) which is designed for the mass function is defined as below.

Definition 7. *(The IV-JS divergence between two BBAs)*
Let A_i be a hypothesis of the mass function m, and let m_1 and m_2 betwo BBAs on the same frame of discernment X, containing N mutually exclusive and exhaustive hypotheses. $\hat{m}_1$ and $\hat{m}_2$ are the information volume of m_1 and m_2 in the case of the limit value δ, respectively. The IV-JS divergence between the two BBAs m_1 and m_2 is denoted as:

$$IV - JS\left(m_1, m_2\right) = \frac{1}{2}\left[S\left(\hat{m}_1, \frac{\hat{m}_1 + \hat{m}_2}{2}\right) + S\left(\hat{m}_2, \frac{\hat{m}_1 + \hat{m}_2}{2}\right)\right] \quad (12)$$

where $S\left(\hat{m}_1, \hat{m}_2\right) = \sum_i \frac{1}{2^{|A_i|-1}} \hat{m}_1\left(A_i\right) log_2 \frac{\hat{m}_1(A_i)}{\hat{m}_2(A_i)}$, and $\sum_i \hat{m}_j\left(A_i\right) = 1$.

It is obvious that the fraction value tends to infinity when the BPA is zero and the value of its logarithm also tends to infinity. The proposed method will fail in this case, so a very small number $1.0000e - 16$ is used to replace zero value when the above case occurs. It has been proven that this will not affect the calculation results [88]. The Belief JensenShannon divergence is similar with JensenShannon divergence in form, however, the IV-JS divergence utilizes the mass function by taking the place of probability distribution function. In such a situation that all of the belief functions hypothesis are assigned to the single elements, the BPA will turn into probability; the IV-JS divergence degenerates to JensenShannon divergence in this case.

The property can be inferred as below:

(1) $IV - JS\left(m_1, m_2\right)$ is symmetric and always well defined;

(2) $IV - JS\left(m_1, m_2\right)$ is bounded, $0 IV - JS\left(m_1, m_2\right) 1$;

(3) its square root, $\sqrt{IV - JS\left(m_1, m_2\right)}$ verifies the triangle inequality.

4. Some numerical examples

Example 1. Supposing that there are two BPAs m_1 and m_2 in the frame of discernment $X = \{AB, C\}$ which is complete, and the two BPAs are given as follows:

$m_1 : m_1(A) = 0.6, m_1(B) = 0.2, m_1(C) = 0.2;$

$m_2 : m_2(A) = 0.6, m_2(B) = 0.2, m_2(C) = 0.2;$

Next, the information volume of m_1 and m_2 is expressed as follows:

$\hat{m}_1 : \hat{m}_1(A) = 0.6, \hat{m}_1(B) = 0.2, \hat{m}_1(C) = 0.2;$

$\hat{m}_2 : \hat{m}_2(A) = 0.6, \hat{m}_2(B) = 0.2, \hat{m}_2(C) = 0.2;$

As shown in Example 1, it can be see that $\hat{m}_1$ has the same BPAs as $\hat{m}_2$, where $\hat{m}_1(A) = \hat{m}_2(A) = 0.6$, $\hat{m}_1(B) = \hat{m}_2(B) = 0.2$ and $\hat{m}_1(C) = \hat{m}_2(C) = 0.2$. Then, the specific calculation processes of $IV - JS(m_1, m_2)$ are listed as follows:

$$IV - JS(m_1, m_2) = \frac{1}{2}\frac{1}{2^{|1|}-1}0.6 \times log\left(\frac{2\times 0.6}{0.6+0.6}\right) + \frac{1}{2}\frac{1}{2^{|1|}-1}0.6 \times log\left(\frac{2\times 0.6}{0.6+0.6}\right) + \frac{1}{2}\frac{1}{2^{|1|}-1}0.2\times log\left(\frac{2\times 0.2}{0.2+0.2}\right) + \frac{1}{2}\frac{1}{2^{|1|}-1}0.2\times log\left(\frac{2\times 0.2}{0.2+0.2}\right) + \frac{1}{2}\frac{1}{2^{|1|}-1}0.2\times log\left(\frac{2\times 0.2}{0.2+0.2}\right) + \frac{1}{2}\frac{1}{2^{|1|}-1}0.2 \times log\left(\frac{2\times 0.2}{0.2+0.2}\right) = 0$$

This example verifies that when m_1 has the same BPAs as m_2, the $IV - JS$ between m_1 and m_2 is 0 which accords with an intuitionistic result.

Example 2. Supposing that there are two BPAs m_1 and m_2 in the frame of discernment $X = \{AB, C\}$ which is complete, and the two BPAs are given as follows:

$m_1 : m_1(A) = 0.6, m_1(B) = 0.2, m_1(C) = 0.2;$

$m_2 : m_2(A) = 0.7, m_2(B) = 0.2, m_2(C) = 0.1;$

The information volume of m_1 and m_2 is expressed as follows:

$\hat{m}_1 : \hat{m}_1(A) = 0.6, \hat{m}_1(B) = 0.2, \hat{m}_1(C) = 0.2;$

$\hat{m}_2 : \hat{m}_2(A) = 0.7, \hat{m}_2(B) = 0.2, \hat{m}_2(C) = 0.1;$

Next, $IV - JS(m_1, m_2) = 0.0150$, and $IV - JS(m_2, m_1) = 0.0150$

From the above results, it can be see that the $IV - JS$ between m_1 and m_2 is equal to the divergence measure between m_2 and m_1 . Consequently, the symmetric property of $IV - JS$ is verified in this example.

Example 3. Supposing that there are two BPAs m_1 , m_2 and and m_3 in the frame of discernment $X = \{AB\}$ which is complete, and the two BPAs are given as follows:

$m_1 : m_1(A) = 0.99, m_1(B) = 0.01;$

$m_2 : m_2(A) = 0.90, m_2(B) = 0.10;$

$m_3 : m_3(A) = 0.01, m_3(B) = 0.99.$

Next, the information volume of m_1 , m_2 and and m_3 is expressed as follows:

$\hat{m}_1 : \hat{m}_1(A) = 0.99, \hat{m}_1(B) = 0.01;$

$\hat{m}_2 : \hat{m}_2(A) = 0.90, \hat{m}_2(B) = 0.10;$

$\hat{m}_3 : \hat{m}_3(A) = 0.01, \hat{m}_3(B) = 0.99.$

After that, their corresponding square root values can be calculated as follows:

$\sqrt{IV - JS(m_1, m_2)} = 0.1799,$

$\sqrt{IV - JS(m_2, m_3)} = 0.8481,$

$\sqrt{IV - JS(m_1, m_3)} = 0.9588.$

It can be noticed that $\sqrt{IV - JS(m_1, m_2)} + \sqrt{IV - JS(m_2, m_3)} = 1.0280,$ so that $\sqrt{IV - JS(m_1, m_2)} + \sqrt{IV - JS(m_2, m_3)} \geq \sqrt{IV - JS(m_1, m_3)}$ which satisfies the triangle inequality property of $IV - JS$.

Example 4. Supposing that there are two BPAs m_1 and m_2 in the frame of discernment $X = \{AB, C\}$ which is complete, and the two BPAs are given as follows:

$m_1 : m_1(A) = 0.3, m_1(B) = 0.2, m_1(C) = 0.1, m_1(A, B) = 0.2, m_1(A, C) = 0.1, m_1(B, C) = 0.1, m_1(A, B, C) = 0.1;$

$m_2 : m_2(A) = 0.1, m_2(B) = 0.1, m_2(C) = 0.1, m_2(A, B) = 0.15, m_2(A, C) = 0.05, m_2(B, C) = 0.1, m_2(A, B, C) = 0.4.$

Through **section 2.2.**, we can get the information volume of m_1 and m_2. Finally, through $Eq.$12, the calculation results are as follows:

$IV - JS(m_1, m_2) = 0.1286.$

At the same time, Xiao also proposed BJS divergence[20], and the calculation results are as follows: $BJS(m_1, m_2) = 0.1286.$

Example 5. Supposing that there are two BPAs m_1 and m_2 in the frame of discernment $X = \{AB, C\}$ which is complete, and the two BPAs are given as follows:

$m_1 : m_1(A) = 0.1, m_1(B) = 0.1, m_1(C) = 0.1, m_1(A, B) = 0.1, m_1(A, C) = 0.1, m_1(B, C) = 0.1, m_1(A, B, C) = 0.4;$

$m_2 : m_2(A) = 0.1, m_2(B) = 0.1, m_2(C) = 0.1, m_2(A, B) = 0.15, m_2(A, C) = 0.05, m_2(B, C) = 0.1, m_2(A, B, C) = 0.4.$

Through **section 2.2.**, we can get the information volume of m_1 and m_2. Finally, through $Eq.$12, the calculation results are as follows:

$IV - JS(m_1, m_2) = 0.0098$

Xiao's[20] calculation results are as follows:$BJS(m_1, m_2) = 0.0098$.

Obviously, the two BPAs in Example 5 are more similar and less distant, which is intuitive. At the same time, the new method is more sensitive than Xiao's method because $0.1286 - 0.0098 > 0.1282 - 0.0098$.

5. Conclusion

In this paper, we propose a new divergence measure(IV-JS). Compared with existing methods, we fully consider the differences of focal elements from the perspective of information volume through the non-specific characteristics of Deng entropy. The new method satisfies the axiom of distance measure and is compatible with the traditional divergence measure. Some numerical examples show that the new method has relatively high resolution. Therefore, this method can be used to solve the problem of evidence conflict or classification in the future.

Acknowledgment

The work is partially supported by National Natural Science Foundation of China (Grant No. 61973332).

References

[1] M. Loeve, Probability theory ii, Graduate texts in mathematics 46 (1978) 15.

[2] A. P. Dempster, et al., Upper and lower probabilities generated by a random closed interval, The Annals of Mathematical Statistics 39 (3) (1968) 957–966.

[3] G. Shafer, A mathematical theory of evidence, Vol. 42, Princeton university press, 1976.

[4] L. A. Zadeh, Fuzzy sets, Information and control 8 (3) (1965) 338–353.

[5] K. T. Atanassov, Intuitionistic fuzzy sets, Vol. 20, Springer, 1986, pp. 87–96.

[6] R. R. Yager, Pythagorean fuzzy subsets, in: 2013 joint IFSA world congress and NAFIPS annual meeting (IFSA/NAFIPS), IEEE, 2013, pp. 57–61.

[7] Z. Pawlak, Rough sets, International journal of computer & information sciences 11 (5) (1982) 341–356.

[8] Y. Deng, D numbers: theory and applications, Journal of Information &Computational Science 9 (9) (2012) 2421–2428.

[9] F. Xiao, A new divergence measure for belief functions in D-S evidence theory for multisensor data fusion, Information Sciences (2019) DOI: 10.1016/j.ins.2019.11.022.

[10] W. Zhang, Y. Deng, Combining conflicting evidence using the DEMATEL method, Soft Computing 23 (17) (2019) 8207–8216.

[11] H. Zhang, Y. Deng, Weighted belief function of sensor data fusion in engine fault diagnosis, Soft computing () (2019) DOI: 10.1007/s00500–019–04063–7.

[12] H. Wang, X. Deng, Z. Zhang, W. Jiang, A new failure mode and effects analysis method based on Dempster–Shafer theory by integrating evidential network, IEEE Access 7 (2019) 79579–79591.

[13] B. Kang, P. Zhang, Z. Gao, G. Chhipi-Shrestha, K. Hewage, R. Sadiq, Environmental assessment under uncertainty using Dempster–Shafer theory and Z-numbers, Journal of Ambient Intelligence and Humanized Computing (2019) DOI: 10.1007/s12652–019–01228–y.

[14] X. Su, L. Li, H. Qian, M. Sankaran, Y. Deng, A new rule to combine dependent bodies of evidence, Soft Computing 23 (20) (2019) 9793–9799.

[15] Y. Deng, The information volume of uncertain informaion: (1) mass function, viXra preprint viXra (2020) 2006.0028.

[16] J. Deng, Y. Deng, The information volume of uncertain information: (2) fuzzy membership function, viXra preprint viXra (2020) 2006.0037.

[17] T. Wen, Y. Deng, The information volume of uncertain information: (3) information fractal dimension, viXra preprint viXra (2020) 2006.0035.

[18] B. Kang, Y. Deng, The maximum deng entropy, IEEE Access 7 (2019) 120758–120765.

[19] J. Lin, Divergence measures based on the shannon entropy, IEEE Transactions on Information theory 37 (1) (1991) 145–151.

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