Cross Entropy of Belief Function

Fan Liu[a], Yangxue Li[b], Yong Deng[b,*]

[a] *Yingcai Honors of shcool&School of Mathematical Sciences , University of Electronic Science and Technology of China, Chengdu, 610054, China*

[b] *Institute of Fundamental and Frontier Science, University of Electronic Science and Technology of China, Chengdu, 610054, China*

Abstract

Dempster-Shafer evidence theory as an extension of Probability has wide applications in many fields. Recently, A new entropy called Deng entropy was proposed in evidence theory. There were a lot of discussions and applications about Deng entropy. However, there is no discussion on how to apply Deng entropy to measure the correlation between the two evidences. In this article, we first review and analyze some of the work related to mutual information. Then we propose the extension of Deng Entropy: joint Deng entropy, Conditional Deng entropy and cross Deng entropy. In addition, we prove the relevant properties of this entropy. Finally, we also proposed a method to obtain joint evidence.

Keywords: Dempster Shafer evidence theory , Deng entropy, Cross entropy , Conditional entropy , Basic probability assignment, Belief function.

---

1. Introduction

Dempster-shafer evidence theory [1, 2] was proposed by Dempster [1] and developed by Shafer [2]. Evidence theory as a framework of uncertain reasoning is closely related to probability theory. It can be considered as a generalization of probability, assigning belief to power set of the propositions rather than single elements. This theory allows for the combination of evidence from different sources and draws a certain degree of conclusion, taking

*Institute of Fundamental and Frontier Science, University of Electronic Science and Technology of China, Chengdu, 610054, China

Email address: dengentropy@uestc.edu.cn, prof.deng@hotmail.com (Yong Deng)

into account all available evidence. So there are a lot of applications about it, and so on [3, 4, 5, 6, 7, 8].

Shannon entropy [9] can be used as an uncertainty measure in probability theory. But it cannot be directly applied in evidence theory. In order to solve this limitation, some scholars have proposed some solutions. Pal et al. [10] proposed an axioms capable of measuring the total uncertainty and derived a new expression. This measure is closely related to basic probability assignment(BPA) and satisfies additivity. Yager [11] introduces a new entropy combined with BPA and belief function and analyzes the property of this entropy. Klir [12] later reviewed some of the uncertainties in measuring uncertainty and pointed out the shortcomings of these methods. Therefore, the measure of discord is proposed. This method not only satisfies the intuitive reason, but also satisfies some ideal mathematical properties. George *et al*. [13] believe that the uncertainty of evidence theory is mainly due to the randomness and lack of accuracy of evidence. They built a new measurement uncertainty based on the distance of proposition. And so on [14, 15]. Subsequently, some discussions about Deng entropy and its application were also raised.

Recently, a new entropy [16] called Deng entropy has been proposed as a measure of uncertainty. It is closely related to the cardinality of BPA and BPA. When the belief is assigned to single elements, Deng entropy degenerates into Shannon entropy [9]. Abellán [17] compares Deng entropy with other entropies and analyzes some its properties. Xiao [18] based on Deng entropy proposed an improved version of the combination method to deal with conflicting evidence. And so on [19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30].

Although there are many applications and extensions about Deng Entropy, there is no relevant entropy to measure the BPA relationship. Therefore, based on Deng entropy we proposed cross and conditional entropy of BPA.

Our contribution can be summarized as follows.

- Review and analyze some related mutual information work.
- Propose joint evidence, conditional Deng entropy, joint Deng entropy, cross Deng entropy and prove some relations of these entropies.
- Propose a method for approximate calculation of joint evidence .
- Propose a method for approximate calculation of joint evidence.

- Verify proposed method though some examples.

The limitations and future work are as follows.

- How to obtain joint evidence is not clear.
- The proposed method for obtaining approximate joint evidence is through Dempster's rule of combination. The result is inconsistent with the intuition, when the evidence contradicts each other.
- In the future we will present some better approximation methods and demonstrate the effectiveness of the proposed method through a large number of practical examples.

The structure of this paper is as follows. Section 2 briefly introduces some basic knowledge. Some related work is introduced in section 3. Section 4 proposes our method and prove and discuss some properties. Section 5 gives some examples. Finally, conclusion is given.

2. Basic Knowledge

In this section, some basic knowledge are briefly introduced.

Definition 1 Basic probability assignment(BPA) is defined as follows [1, 2].

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

Note that $m(\phi) = 0$ and $\sum_{A \in 2^{\Theta}} m(A) = 1$. where Θ is the frame of discernment and 2^{Θ} is the power set of Θ.

Example 1
if $\Theta = \{a, b\}$, then $2^{\Theta} = \{\phi, \{a\}, \{b\}, \Theta\}$.

Definition 2 Deng entropy is defined as follows [16].

$$E_d = \sum_i m(F_i) log \frac{m(F_i)}{2^{|F_i|} - 1} \tag{2}$$

where $|F_i|$ is the cardinality of F_i.

Definition 3 Suppose two BPAs indited by m_1 and m_2, the Dempster's rule of combination is defined as follows [1, 2].

$$m_{1,2}(A) = (m_1 \oplus m_2)(A) = \frac{1}{1 - K} \sum_{B \cap C = A \neq \phi} m_1(B) m_2(C) \tag{3}$$

where $K = \sum_{B \cap C = \phi} m(B) m(C)$.

3. Related Work

In this section, some related work is reviewed.

3.1. Mutual Information in Probability

Mutual information is a measure that a random variable tells another random variable. It can be considered as a uncertainty reduction of another random variable where a random variable is known. High mutual information represents more uncertainty reduction, low mutual information represents less uncertainty reduction. In special cases, when mutual information is zero, there is no change in uncertainty.

Definition 3 Given two discrete random variables X and Y, their joint probability distribution is $P_{XY}(x, y) \triangleq P(x, y)$. Mutual information $I(X; Y)$ is defined as follows [31].

$$I(X, Y) = \sum_{x,y} P(x, y) \log \frac{P(x, y)}{P(x)P(y)} \tag{4}$$

where $P(x)$ and $P(y)$ are marginals, which indicates $P(x) = \sum_y P(x, y)$ and $P(y) = \sum_x P(x, y)$.

The definition of Shannon entropy and conditional entropy [9] is given in order to further understand the meaning of mutual information.

Definition 4 The shannon entropy is defined as follows [9].

$$H(X) = \sum_{x} P(x) \log P(x) \tag{5}$$

Definition 5 The conditional entropy is defined as follows [9].

$$H(Y|X) = \sum_{x,y} P(x, y) \log P(y|x) \tag{6}$$

where $P_{Y|X}(y|x) \triangleq P(y|x)$ is the conditional distribution.

Shannon Entropy is a measure of uncertainty [9, 31]. It can be proved that the *Eq.* 4 is equivalent to the following *Eq.* 7 [31].

$$I(X; Y) = H(X) - H(X|Y) \tag{7}$$

When given the random variable Y, the uncertainty of the random variable X is reduced [31].

3.2. Fusion of Dependent Evidence Based on Mutual Information

Recently, Su *et al*. [32] proposed a work about the fusion of dependent evidence. In the actual information fusion process, many evidences are not independent of each other [32]. In order to improve the availability of evidence, it is necessary to measure the degree of independence of evidence [32].

The author uses mutual information to judge the degree of mutual independence of evidence. However, this mutual information is not "real mutual information" among evidences. The author's method is mainly divided into 5 steps [32].

$STEP1$ Calculate mutual information $I(S_x, S_y)$ [32]. where S_x and S_y are the information source [32].

$STEP2$ Calculate the generalized correlation coefficient $R(S_x, S_y)$. where $R = \frac{I(S_x,S_y)}{\sqrt{H(S_x,S_y)}}$ [32].

$STEP3$ Build a coefficient matrix [32].

$STEP4$ Sum the coefficients of the same row [32].

$STEP5$ Calculate the weighting coefficients of each information source [32].

The author verifies his method through some numerical experiments and finds that the accuracy of classification can be improved when considering the degree of correlation between individual evidences [32].

4. Mutual Information in Evidence Theory

So when BPAs are not independent of each other, what is their mutual information? The mutual information of BPA should be an uncertaintity measure of another BPA when one BPA is known. When one evidence is known, the uncertainty of the other evidence should be reduced. So based on the above assumptions, a joint evidence distribution is defined.

Definition 6 The joint evidence is defined as follows.

$$m(F_i, F_j)^{1,2} \quad m(\phi, F_k)^{1,2} = 0 || m(F_t, \phi)^{1,2} = 0 \tag{8}$$

where $F_i, F_j, F_k, F_t \in 2^\Theta$. Note that the $m(F_i, F_j)^{1,2}$ should satify the following Eq. 9 and 10.

$$m_1(F_i) = \sum_j m(F_i, F_j)^{1,2} || m_2(F_j) = \sum_i m(F_i, F_j)^{1,2} \tag{9}$$

$$\sum_{i,j} m(F_i, F_j)^{1,2} = 1 \tag{10}$$

The interpretation of joint evidence is that simultaneously assigning the beliefs to the power set of Θ through witness. It should be noted that the joint evidence is different from the BPA after fusing BPAs by dempster's rule of combination, because the joint evidence is multidimensional. Note that the joint BPA is easy to expand into N dimensions .

A simple example is presented to help understand the definition of joint BPA.

Example 2

Given a frame of discernment of $\Theta = \{a, b\}$, there are two witnesses that assign the belief to the power set of Θ. The joint BPA can be seen in *Table* 1 .

$m(F_i, F_j)$	a	b	Θ	m_2
a	0.4	0.2	0.2	0.8
b	0.1	0	0	0.1
Θ	0.1	0	0	0.1
m_1	0.6	0.2	0.2	1

Table 1: An example of joint BPA

Definition 7 Once the joint distribution of evidence is determined, the definition of corss Deng entropy(mutual entropy) is as follows.

$$I(m_1; m_2) = \sum_{i,j} m(F_i, F_j)^{1,2} \log \frac{m(F_i, F_j)^{1,2}}{m_1(F_i)m_2(F_j)} \tag{11}$$

Definition 8 The conditional Deng entropy is defined as follows.

$$E_d(m_1|m_2) = -\sum_{i,j} m(F_i, F_j)^{1,2} \log \frac{m(F_i, F_j)^{1,2}}{m_2(F_j)(2^{|F_i|} - 1)} \tag{12}$$

Definition 9 The joint Deng entropy is defined as follows.

$$E_d(m_1, m_2) = -\sum_{i,j} m(F_i, F_j)^{1,2} \log \frac{m(F_i, F_j)^{1,2}}{(2^{|F_i|} - 1)(2^{|F_j|} - 1)} \tag{13}$$

The following is proof of the relationship between these entropies.

Theorem 1. $I(m_1;m_2) = E_d(m_1) - E_d(m_1|m_2)$

Proof 1. $I(m_1;m_2) = \sum_{i,j} m(F_i,F_j)^{1,2} \log \frac{m(F_i,F_j)^{1,2}}{m_1(F_i)m_2(F_j)}$
$= -\sum_{i,j} m(F_i,F_j)^{1,2} \log \frac{m_1(F_i)}{2^{|F_i|}-1} - (-\sum_{i,j} m(F_i,F_j)^{1,2} \log \frac{m(F_i,F_j)^{1,2}}{m_2(F_j)(2^{|F_i|}-1)})$
$= -\sum_{i} m_1(F_i) \log \frac{m_1(F_i)}{2^{|F_i|}-1} - (-\sum_{i,j} m(F_i,F_j)^{1,2} \log \frac{m(F_i,F_j)^{1,2}}{m_2(F_j)(2^{|F_i|}-1)})$
$= E_d(m_1) - E_d(m_1|m_2)$

Theorem 1 explains why the uncertainty of BPA m_1 becomes smaller when given the BPA m_2. When BPA m_1 and BPA $m2$ are not independent of each other and if one BPA is given, the uncertainty of the other BPA will become smaller. The amount of uncertainty reduction is equal to the conditional Deng entropy(*Eq.* 12) of the two BPAs.

Theorem 2. $I(m_1;m_2) = I(m_2;m_1)$

Proof 2. $I(m_1;m_2) = \sum_{i,j} m(F_i,F_j)^{1,2} \log \frac{m(F_i,F_j)^{1,2}}{m_1(F_i)m_2(F_j)}$
$= -\sum_{i,j} m(F_i,F_j)^{1,2} \log \frac{m_2(F_j)}{2^{|F_j|}-1} - (-\sum_{i,j} m(F_i,F_j)^{1,2} \log \frac{m(F_i,F_j)^{1,2}}{m_1(F_i)(2^{|F_j|}-1)})$
$= -\sum_{j} m_2(F_j) \log \frac{m_2(F_j)}{2^{|F_j|}-1} - (-\sum_{i,j} m(F_i,F_j)^{1,2} \log \frac{m(F_i,F_j)^{1,2}}{m_1(F_i)(2^{|F_j|}-1)})$
$= E_d(m_2) - E_d(m_2|m_1)$
$= I(m_2;m_1)$

Theorem 2 shows that Cross Deng entropy(*Eq.* 11) is commutative. Given two BPAs, their mutual information does not need to be considered in order. This property is close to the actual situation, Because the order of evidence is often unclear when collecting evidence from evidence sources.

Theorem 3. $E_d(m_1,m_2) = E_d(m_2|m_1) + E_d(m_1)$

Proof 3. $E_d(m_1,m_2) = -\sum_{i,j} m(F_i,F_j)^{1,2} \log \frac{m(F_i,F_j)^{1,2}}{(2^{|F_i|}-1)(2^{|F_j|}-1)}$
$= -\sum_{i,j} m(F_i,F_j)^{1,2} \log \frac{m(F_i,F_j)^{1,2} m_1(F_i)}{(2^{|F_i|}-1)(2^{|F_j|}-1)m_1(F_i)}$
$= -\sum_{i} m_1(F_i) \log \frac{m_1(F_i)}{2^{|F_i|}-1} - (-\sum_{i,j} m(F_i,F_j)^{1,2} \log \frac{m(F_i,F_j)^{1,2}}{m_1(F_i)(2^{|F_j|}-1)})$
$= E_d(m_2|m_1) + E_d(m_1)$

Theorem 3 shows that the joint Deng entropy of BPAs is equal to the Deng entropy of one of the BPAs plus the conditional Deng entropy of another BPA.

Theorem 4. $I(m_1; m_1) = E_d(m_1)$

Proof 4. $I(m_1; m_1) = E_d(m_1) - E_d(m_1|m_1) = E_d(m_1)$

Theorem 4 shows that the Deng entropy of of BPA is equal to its self-information.

However, in most cases the joint evidence is not clear. Generally only know each single piece of evidence, do not know the joint evidence. Because joint evidence is multidimensional and a single piece of evidence is one-dimensional. The fact is that high-dimensional evidence cannot be inferred from low-dimensional evidence, But low-dimensional evidence can be derived from high-dimensional evidence.

To address this limitation, we propose an approximate method. The approach of approximate method is very simple. There are two main steps:

STEP 1

the BPA is fused by Dempster's rule of combination through *Eq.* 3.

STEP 2

Get joint evidence by multiplying the corresponding row element with the column element. The specific process can be seen in *Table* 2.

Table 2: *STEP* 2

$m(F_i, F_j)$	F_1	F_2	F_j	$F_{2^{\|\Theta\|}-1}$	$m_{1,2}$
F_1	$m_{1,2}(F_1)m_{1,2}(F_1)$	$m_{1,2}(F_1)m_{1,2}(F_2)$	$m_{1,2}(F_1)m_{1,2}(F_j)$	$m_{1,2}(F_1)m_{1,2}(F_{2^{\|\Theta\|}-1})$	$m_{1,2}(F_1)$
F_2	$m_{1,2}(F_2)m_{1,2}(F_1)$	$m_{1,2}(F_2)m_{1,2}(F_2)$	$m_{1,2}(F_2)m_{1,2}(F_j)$	$m_{1,2}(F_2)m_{1,2}(F_{2^{\|\Theta\|}-1})$	$m_{1,2}(F_2)$
F_i	$m_{1,2}(F_i)m_{1,2}(F_1)$	$m_{1,2}(F_i)m_{1,2}(F_2)$	$m_{1,2}(F_i)m_{1,2}(F_j)$	$m_{1,2}(F_i)m_{1,2}(F_j)$	$m_{1,2}(F_j)$
$F_{2^{\|\Theta\|}-1}$	$m_{1,2}(F_{2^{\|\Theta\|}-1})m_{1,2}(F_1)$	$m_{1,2}(F_{2^{\|\Theta\|}-1})m_{1,2}(F_2)$	$m_{1,2}(F_{2^{\|\Theta\|}-1})m_{1,2}(F_j)$	$m_{1,2}(F_{2^{\|\Theta\|}-1})m_{1,2}(F_{2^{\|\Theta\|}-1})$	$m_{1,2}(F_{2^{\|\Theta\|}-1})$
$m_{1,2}(F_i)$	$m_{1,2}(F_1)$	$m_{1,2}(F_1)$	$m_{1,2}(F_j)$	$m_{1,2}(F_{2^{\|\Theta\|}-1})$	1

Although this method can obtain joint evidence, the way it is obtained is through Dempster's rule of combination. There is a lot of debate about Dempster's rule of combination. One obvious argument is that when the evidence is contradictory, it will produce an opposite conclusion. A discussion of Dempsterrule of combination can be found at [33].

5. Example

Given a frame of discernment of $\Theta = \{a, b\}$, there are two witnesses that assign the belief to the power set of Θ. The joint BPA can be seen in $Table$ 1 .

Through $Eq.$ 11

$I(m_1, m_2) =$

$= 0.4 \times \log \frac{0.4}{0.2\times0.8} + 0.2 \times \log \frac{0.2}{0.2\times0.8} + 0.2 \times \log \frac{0.2}{0.2\times0.8} + 0.1 \times \log \frac{0.1}{0.6\times0.1} + 0.1 \times \log \frac{0.1}{0.6\times0.1}$

$= 0.1710$

Calculate $I(m_1, m_2)$ by approximate method.

First merge the two BPAs through the Dempster rule of combination. The result can be seen in $Table$ 3.

$m(F_i, F_j)$	a	b	Θ	m
a	0.4793	0.0355	0.1775	0.6923
b	0.0355	0.0026	0.0131	0.0513
Θ	0.1775	0.0131	0.0657	0.2563
m	0.6923	0.0513	0.2563	1

Table 3: result of approximation method

Through the Through $Eq.$ 11. the new $I(m_1, m_2) = 0.2887$which can see that the mutual information of the two BPAs is still very low.

6. Conclusion

The "main story" of this article is about how to measure the relevance of two pieces of evidence. First, we review the work on mutual information and analyze the role of these mutual information. Then through the previous discussion to define cross Deng entropy and propose an approximation method to calculate joint evidence. Finally, verify our method with some examples.

References

[1] A. P. Dempster, Upper and lower probabilities induced by a multivalued mapping, The annals of mathematical statistics (1967) 325–339.

[2] G. Shafer, et al., A mathematical theory of evidence, volume 1, Princeton university press Princeton, 1976.

[3] J. Wang, K. Qiao, Z. Zhang, An improvement for combination rule in evidence theory, Future Generation Computer Systems 91 (2019) 1–9.

[4] W. Liu, X. Chen, B. Chen, J. Wang, L. Chen, An efficient algorithm for influence maximization based on propagation path analysis, in: Advances in Computer Science and Ubiquitous Computing, Springer, 2017, pp. 836–845.

[5] A. L. Kuzemsky, Temporal evolution, directionality of time and irreversibility, RIVISTA DEL NUOVO CIMENTO 41 (2018) 513–574.

[6] J. An, M. Hu, L. Fu, J. Zhan, A novel fuzzy approach for combining uncertain conflict evidences in the dempster-shafer theory, IEEE Access 7 (2019) 7481–7501.

[7] M. D. Mambe, T. N'Takpe, N. G. Anoh, S. Oumtanaga, A New Uncertainty Measure in Belief Entropy Framework, INTERNATIONAL JOURNAL OF ADVANCED COMPUTER SCIENCE AND APPLICATIONS 9 (2018) 600–606.

[8] M. D. Mambe, S. Oumtanaga, G. N. Anoh, A belief entropy-based approach for conflict resolution in iot applications, in: 2018 1st International Conference on Smart Cities and Communities (SCCIC), IEEE, pp. 1–5.

[9] C. E. Shannon, A mathematical theory of communication, Bell system technical journal 27 (1948) 379–423.

[10] N. R. Pal, J. C. Bezdek, R. Hemasinha, Uncertainty measures for evidential reasoning ii: A new measure of total uncertainty, International Journal of Approximate Reasoning 8 (1993) 1–16.

[11] R. R. Yager, Entropy and specificity in a mathematical theory of evidence, International Journal of General System 9 (1983) 249–260.

[12] G. J. Klir, A. Ramer, Uncertainty in the dempster-shafer theory: a critical re-examination, International Journal of General System 18 (1990) 155–166.

[13] T. George, N. R. Pal, Quantification of conflict in dempster-shafer framework: a new approach, International Journal Of General System 24 (1996) 407–423.

[14] Y. Yang, D. Han, A new distance-based total uncertainty measure in the theory of belief functions, Knowledge-Based Systems 94 (2016) 114–123.

[15] Y. Song, X. Wang, L. Lei, S. Yue, Uncertainty measure for interval-valued belief structures, Measurement 80 (2016) 241–250.

[16] Y. Deng, Deng entropy, Chaos, Solitons & Fractals 91 (2016) 549–553.

[17] J. Abellán, Analyzing properties of deng entropy in the theory of evidence, Chaos, Solitons & Fractals 95 (2017) 195–199.

[18] F. Xiao, An improved method for combining conflicting evidences based on the similarity measure and belief function entropy, International Journal of Fuzzy Systems 20 (2018) 1256–1266.

[19] Y. Li, Y. Deng, Generalized ordered propositions fusion based on belief entropy., International Journal of Computers, Communications & Control 13 (2018).

[20] L. Pan, Y. Deng, A new belief entropy to measure uncertainty of basic probability assignments based on belief function and plausibility function, Entropy 20 (2018) 842.

[21] K. Yuan, F. Xiao, L. Fei, B. Kang, Y. Deng, Modeling sensor reliability in fault diagnosis based on evidence theory, Sensors 16 (2016) 113.

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

[23] F. Xiao, An improved method for combining conflicting evidences based on the similarity measure and belief function entropy, International Journal of Fuzzy Systems 20 (2018) 1256–1266.

[24] J. Abellán, C. J. Mantas, J. G. Castellano, A random forest approach using imprecise probabilities, Knowledge-Based Systems 134 (2017) 72–84.

[25] J. Abellan, J. G. Castellano, Improving the Naive Bayes Classifier via a Quick Variable Selection Method Using Maximum of Entropy, ENTROPY 19 (2017).

[26] J. Abellan, E. Bosse, Drawbacks of Uncertainty Measures Based on the Pignistic Transformation, IEEE TRANSACTIONS ON SYSTEMS MAN CYBERNETICS-SYSTEMS 48 (2018) 382–388.

[27] R. Jirousek, P. P. Shenoy, A new definition of entropy of belief functions in the Dempster-Shafer theory, INTERNATIONAL JOURNAL OF APPROXIMATE REASONING 92 (2018) 49–65.

[28] A. D. Jaunzemis, M. J. Holzinger, M. W. Chan, P. P. Shenoy, Evidence gathering for hypothesis resolution using judicial evidential reasoning, Information Fusion 49 (2019) 26–45.

[29] J. Vandoni, E. Aldea, S. Le Hégarat-Mascle, Evidential query-by-committee active learning for pedestrian detection in high-density crowds, International Journal of Approximate Reasoning 104 (2019) 166–184.

[30] W. Zhu, H. Yang, Y. Jin, B. Liu, A method for recognizing fatigue driving based on dempster-shafer theory and fuzzy neural network, Mathematical Problems in Engineering 2017 (2017).

[31] W. Weaver, C. E. Shannon, The mathematical theory of communication. 1949, Urbana, Illinois: University of Illinois Press (1963).

[32] X. Su, L. Li, F. Shi, H. Qian, Research on the fusion of dependent evidence based on mutual information, IEEE Access 6 (2018) 71839–71845.

[33] J. Dezert, A. Tchamova, F. Dambreville, On the mathematical theory of evidence and dempster's rule of combination (2011).