The information volume of uncertain information: (7) information quality

Dingbin Li[a,b], Yong Deng[a,c,d,*]

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

Abstract

Information quality is a concept that can be used to measure the information of probability distribution. Dempster-Shafer evidence theory can describe uncertain information more reasonably than probability theory. Therefore, it is a research hot spot to propose information quality applicable to evidence theory. Recently, Deng proposed the concept of information volume based on Deng entropy. It is worth noting that, compared with the Deng entropy, the information volume of the Deng entropy contains more information. Obviously, it may be more reasonable to use information volume of Deng entropy to represent uncertain information. Therefore, this article proposes a new information quality, which is based on the information volume of Deng entropy. In addition, when the basic probability (BPA) degenerates into a probability distribution, the proposed information quality is consistent with the information quality proposed by Ygare and Petry. Finally, several numerical examples illustrate the effectiveness of this new method.

Keywords: Information volume; Information quality; Deng entropy; Mass function; Deng distribution.

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

1. Introduction

In recent decades, people have developed a large number of theories to express and deal with uncertainties in uncertain environments, such as probability theory [1], fuzzy set theory [2], Dempster-Shafer evidence theory [3, 4], rough set [5], D number [6]. As a measure of the uncertainty of probability distribution, information quality [7] is widely used in decision-making [8], pattern recognition [9] and so on. Due to Dempster-Shafer evidence theory expresses uncertain information better than probability theory, so it is necessary to propose an information quality suitable for evidence theory.

Recently, Deng proposed a concept of information volume of mass function [10] based on Deng entropy [11]. First, by defining a new distribution, called Deng distribution, the BPA is reconstructed through the Deng distribution. When a BPA is given, the corresponding information is greater than the uncertainty of Deng Entropy. Obviously, the amount of information volume of mass function can better represent uncertain information. Therefore, this paper proposes an information quality based on the information volume. When BPA degenerates into a probability distribution, the information quality proposed in this paper will degenerate into information quality proposed by Ygare and Petry.

The rest of the article is organized as follows. In the second section, Dempster-Shafer evidence theory, Deng entropy, information volume of mass function and information quality are introduced. In the third section, a new information quality based on information volume that can measure the uncertainty of BPA is proposed. In the fourth section, some numerical examples are used to illustrate the effectiveness of the method. The fifth section is a brief conclusion.

2. Preliminary

This section will introduce some preliminary knowledge, including evidence theory, Deng entropy, Information volume of mass function and information quality.

2.1. Evidence theory

Dempster-Shafer evidence theory [3, 4] can be used to deal with uncertainty, and the ability to express uncertain information better than probability theory.

Definition 1. *Assuming the frame of discernment is Θ, and the set Θ have N elements. Then it can be expressed as:*

$$\Theta = \{H_1, \cdots, H_n, \cdots, H_N\} \tag{1}$$

The power of Θ is expressed as 2^Θ, contains all possible subsets of Θ, and 2^Θ is expressed as:

$$\begin{aligned} 2^\Theta &= \{A_1, A_2, \cdots, A_{2^N}\} \\ &= \{\phi, \{\theta_1\}, \{\theta_2\}, \cdots, \{\theta_N\}, \{\theta_1, \theta_2\}, \cdots, \{\theta_1, \theta_N\}, \cdots, \Theta\} \end{aligned} \tag{2}$$

Definition 2. *The basic probability assignment is defined as follows:*

$$m(\phi) = 0 \quad and \quad \sum_{A \subseteq \Theta} m(A) = 1 \tag{3}$$

The mass m(A) indicates the degree of support for evidence A.

2.2. Deng entropy

In information theory, entropy can be used to measure the uncertainty of the system. In recent years, Deng has proposed a new entropy to measure the uncertainty of evidence theory-Deng entropy [11].

Definition 3. *Deng entropy is defined as follows:*

$$H_{DE}(m) = -\sum_{A \in 2^\Theta} m(A) log_2(\frac{m(A)}{2^{|A|} - 1}) \tag{4}$$

where $|A|$ is the cardinal of a certain focal element A. Deng entropy can be rewritten as follows:

$$H_D = \sum_{A \in \Theta} m(A) log_2(2^{|A|} - 1) - \sum_{A \in \Theta} m(A) log_2 m(A) \tag{5}$$

$\sum_{A \in \Theta} m(A) log_2(2^{|A|} - 1)$ and $-\sum_{A \in \Theta} m(A) log_2 m(A)$ are measurements of non-specific and discord, respectively.

2.3. Infromation volume of mass function

Recently, Deng proposed the concept of information volume [12, 13]. First, he defined the BPA condition of maximum Deng entropy as Deng distribution. Then, the information volume of mass function based on Deng entropy is proposed.

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:

$$m_D(A_i) = \frac{(2^{|A_i|} - 1)}{\sum_{A \in 2^\Theta} (2^{|A_i|} - 1)} \tag{6}$$

For example, given a focal element $A_{i-1} = \{\theta_x, \theta_y\}$ and its mass function $m(A_{i-1})$, the separating proportion is $\frac{1}{5} : \frac{1}{5} : \frac{3}{5}$. The ith times of separation divides $m(A_{i-1})$ and yields following new mass function: $m(X_i), m(Y_i), m(Z_i)$, where $X_i = \{\theta_x\}$, $y_i = \{\theta_y\}$, $Z_i = \{\theta_x, \theta_y\}$. In addition, they satisfy these equations:

$$m(X_i) + m(Y_i) + m(Z_i) = m(A_{i-1}) \tag{7}$$

$$m(X_i) : m(Y_i) : m(Z_i) = \frac{1}{5} : \frac{1}{5} : \frac{3}{5} \tag{8}$$

step 2-2: Based on the 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 $\Delta = H_i(m) - H_{i-1}(m)$. When Δ_i satisfies following condition, jump out this loop.

$$\Delta = H_i(m) - H_{i-1}(m) < varepsilon \tag{9}$$

where Δ is the allowable error.

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

2.4. *Information quality*

As a measure of uncertainty of probability distribution, The information quality [7] has been applied to decision making [14] pattern classification and maximum fusion [15] etc [16].

Definition 4. *Given a probability distribution, the information quality is defined as follows [7]:*

$$IQ_{p_i} = \|p_i\|^2 = \sum_{j=1}^{n} \|p_{ij}\|^2 \tag{10}$$

Where

$$\|p_i\| = \sqrt{(p_i * p_i)} = (\sum_{j=1}^{n} \|p_{ij}\|^2)^{1/2} \tag{11}$$

The bigger the value of information quality, the more certain information provided by the probability distribution.

3. The information quality of information volume

In Dempster Shafer evidence theory, how to measure the information quality of evidence is still an open question. Obviously, Dempster Shafer evidence theory is a generalization of probability theory. Therefore, the above information quality can be generalized to evidence theory, propose a new information quality based on information volume, named information quality of information volume (IV-IQ) which is designed for the mass function.

Definition 5. *Given a basic probability assignment m, the frame of discernment is X. The information quality of information volume is defined as follows:*

$$IV - IQ(m) = \sum_{A_i \subseteq X} \left(\frac{\hat{m}(A_i)}{2^{|A_i|} - 1}\right)^2 \tag{12}$$

where $\hat{m}$ is the information volume of m in the case of the limit value Δ, and $|A_i|$ is the cardinal of a certain focal element A_i.

4. Numerical examples

In this section, the information quality of information volume is better explained by examples. In the following example, the allowable error is 0.001.

Example 1. *Assuming a framework of discernment $X = \{A, B, C\}$, the BPA is given as follows:*

$$m = [(A, \frac{1}{3}), (B, \frac{1}{3})(C, \frac{1}{3})]$$

Because there is no focal element whose cardinal is larger than 1, the step 2-1 can be skipped for all the times of the loop. the information volume of m is expressed as follows:

$$\hat{m} = [(A, \frac{1}{3}), (B, \frac{1}{3})(C, \frac{1}{3})]$$

$$IV - IQ = (\frac{1}{3})^2 + (\frac{1}{3})^2 + (\frac{1}{3})^2 = 0.3333$$

From Example 1, it can be seen that when the BPA degenerates into a probability distribution, the values of $IV - IQ(m)$ is identical to the information quality proposed by Yager and Petry.

Example 2. *Assuming a framework of discernment $X = \{A, B, C\}$, the BPA is given as follows:*

$$m = [(A, \frac{1}{7}), (B, \frac{1}{7})(C, \frac{1}{7})(AB, \frac{1}{7})(AC, \frac{1}{7})(BC, \frac{1}{7})(ABC, \frac{1}{7})]$$

Through *section 2.3.*, we can get the information volume of m and $H_{IV-mass}(m)$ Finally, through Equation 12, the calculation results are as follows:

$$IV - IQ = 0.0693,\ H_{IV-mass}(m) = 5.199486$$

Example 3. *Assuming a framework of discernment $X = \{A, B, C\}$, the BPA is given as follows:*

$$m = [(A, \frac{1}{19}), (B, \frac{1}{19})(C, \frac{1}{19})(AB, \frac{3}{19})(AC, \frac{3}{19})(BC, \frac{3}{19})(ABC, \frac{7}{19})]$$

Through *section 2.3.*, we can get the information volume of m and $H_{IV-mass}(m)$. Finally, through Equation 12, the calculation results are as follows:

$$IV - IQ = 0.0204,\ H_{IV-mass}(m) = 6.469009$$

It should be noted that the $IV - IQ$ in Example 2 is 0.0693 and the $IV - IQ$ in Example 3 is 0.0204, which means the mass function of Example 2 has more information than the value of Example 3. This is consistent with the result of comparing $H_{IV-mass}(m)$.

Example 4. *Assuming a framework of discernment $X = \{A, B, C\}$, the BPA is given as follows:*

$$m = [(ABC, 1)]$$

Through *section 2.3.*, we can get the information volume of m. Finally, through Equation 12, the calculation result is as follows:

$$IV - IQ = 0.0204$$

The example shows that Deng distribution and total uncertainty have the same information quality. This is consistent with the results of the information volume of mass function, and also with intuition.

5. Conclusion

In this paper, in order to propose an information quality applicable to evidence theory. This paper fully considers the non-specific characteristics of Deng entropy, and uses the BPA obtained by the information volume of mass function. This paper presents an information quality based on the information volume of mass function. This method is compatible with information quality proposed by Yager and Petry. Several numerical examples illustrate the effectiveness of this method.

Acknowledgment

The work is partially supported by National Natural Science Foundation of China (Grant Nos. 61973332).

References

[1] Loeve, Probability theory ii, Biometrika II (3) (1978) 378C382.

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

[3] A. P. Dempster, Upper and lower probabilities induced by a multi-valued mapping, Annals of Mathematical Statistics 38 (2) (1967) 325–339.

[4] G.-C. Rota, A mathematical theory of evidence: G. shafer, princeton university press, 1976, 297 pp., Advances in Mathematics 24 (3) (1977) 341.

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

[6] B. liu, Y. Deng, Risk Evaluation in Failure Mode and Effects Analysis Based on D Numbers Theory , International Journal of Computers Communications & Control 14 (5) (2019) 672–691.

[7] R. R. Yager, F. Petry, An intelligent quality-based approach to fusing multi-source probabilistic information, Information Fusion 31 (2016) 127–136.

[8] H. Seiti, R. Tagipour, A. Hafezalkotob, F. Asgari, Maintenance strategy selection with risky evaluations using RAHP, Journal of Multi-Criteria Decision Analysis 24 (5-6) (2017) 257–274.

[9] Z. Liu, Y. Liu, J. Dezert, F. Cuzzolin, Evidence combination based on credal belief redistribution for pattern classification, IEEE Transactions on Fuzzy Systems (2019) 1–1doi:10.1109/TFUZZ.2019.2911915.

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

[11] D. Yong, Deng entropy, Chaos Solitons and Fractals the Interdisciplinary Journal of Nonlinear Science and Nonequilibrium and Complex Phenomena 91 (2016) 549–553.

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

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

[14] H. Seiti, A. Hafezalkotob, R. Fattahi, Extending a pessimistic–optimistic fuzzy information axiom based approach considering acceptable risk: Application in the selection of maintenance strategy, Applied Soft Computing 67 (2018) 895–909.

[15] T. T. Nguyen, T. C. Phan, Q. V. H. Nguyen, K. Aberer, B. Stantic, Maximal fusion of facts on the web with credibility guarantee, Information Fusion 48 (2019) 55 – 66.

[16] R. R. Yager, Interval valued entropies for dempster–shafer structures, Knowledge-Based Systems 161 (2018) 390–397.