A new information unit

Yong Deng[a,b,*]

[a] *School of Electronics and Information, Northwestern Polytechnical University, Xi'an, 710072, China*
[b] *Department of Civil & Environmental Engineering, School of Engineering, Vanderbilt University, Nashville, TN, 37235, USA*

Abstract

It is well known that "Bit" is the unit in information theory to measure information volume with Shannon entropy. However, one assumption to use bit as information unit is that each hypothesis is exclusive with each other. This assumption is also the basic assumption in probability theory which means that two events cannot happen synchronously. However, the assumption is violated such as the "Entangled state". A typical example is Schr?dinger's cat where a cat may be simultaneously both alive and dead. At this situation, bit is not suitable to measure the information volume. To address this issue, a new information unit, called as "*Deng*" and abbreviated as "**D**", is proposed based on Deng entropy. The proposed information unit may be used in entangle information processing and quantum information processing

Keywords: Entropy, Shannon entropy, Deng entropy, Dempster-Shafer evidence theory, Information volume, Information unit, Entangled state,

*Corresponding author: Yong Deng, School of Electronics and Information, Northwestern Polytechnical University, Xi'an, 710072, China.

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

1. Introduction

Uncertainty is ubiquitous in nature. Several uncertainty theories have been developed, such as probability theory [1], fuzzy set theory [2], possibility theory [3], Dempster-Shafer evidence theory [4, 5], generalized evidence theory [6] and D numbers[7].

Since firstly proposed by Clausius in 1865 for thermodynamics [8],the study of uncertainty and entropy attracts great interests and various types of entropies are developed, such as information entropy [9].

This paper is inspired by our previous work [10] and based on the proposed Deng entropy [11]. The main contribution of this work is that a new information unit is presented. Compared with existing unit ”Bit”, the new unit ”Deng” is more generalized. We argue that only on the situation that each hypothesis is exclusive with each other can we use the unit ”Bit”. The exclusive assumption is the base of probability theory which means that two events cannot happen synchronously. However, in quantum mechanics, this assumption is violated. A typical example is Schr?dinger’s cat where the cat may be simultaneously both alive and dead[12].

The paper is organized as follows. The preliminaries Dempster-Shafer evidence theory, Shannon entropy and Deng entropy are briefly introduced in Section 2. Section 3 presents the idea of information unit with Deng entropy. Some numerical examples are illustrated in Section 4. Finally, this paper is concluded in Section 5.

2. Preliminaries

In this section, some preliminaries are briefly introduced.

2.1. Dempster-Shafer evidence theory

Dempster-Shafer theory (short for D-S theory), also called belief function theory, as introduced by Dempster[4] and then developed by Shafer[5], has emerged from their works on statistical inference and uncertain reasoning.

Let X be a set of mutually exclusive and collectively exhaustive events, indicated by

$$X = \{\theta_1, \theta_2, \cdots, \theta_i, \cdots, \theta_{|X|}\} \tag{1}$$

where set X is called a frame of discernment. The power set of X is indicated by 2^X, namely

$$2^X = \{\emptyset, \{\theta_1\}, \cdots, \{\theta_{|X|}\}, \{\theta_1, \theta_2\}, \cdots, \{\theta_1, \theta_2, \cdots, \theta_i\}, \cdots, X\} \tag{2}$$

For a frame of discernment $X = \{\theta_1, \theta_2, \cdots, \theta_{|X|}\}$, a mass function is a mapping m from 2^X to $[0, 1]$, formally defined by:

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

which satisfies the following condition:

$$m(\emptyset) = 0 \quad and \quad \sum_{A \in 2^X} m(A) = 1 \tag{4}$$

In D-S theory, a mass function is also called a basic probability assignment (BPA). Assume there are two BPAs indicated by m_1 and m_2, the Dempster's rule of combination is used to combine them as follows:

$$m(A) = \begin{cases} \frac{1}{1-K} \sum\limits_{B \cap C = A} m_1(B) m_2(C) \ , & A \neq \emptyset; \\ 0 \ , & A = \emptyset. \end{cases} \tag{5}$$

with

$$K = \sum_{B \cap C = \emptyset} m_1(B) m_2(C) \tag{6}$$

Note that the Dempster's rule of combination is only applicable to such two BPAs which satisfy the condition $K < 1$.

D-S theory has more advantages in in handling uncertainty compared with classical probability theory. When information is adequate, probability theory is effective to handle that situation. However, when information is not adequate, probability theory is invalid to such uncertain situation. Here is an example.

As shown in Figure 1, assume there are two boxes. There are red balls the left box, and green balls in the right box. The number of balls in each box is unknown. Now, a person is assigned to pick a boll from these two boxes. We know that he chooses the the left box with a probability $P1 = 0.6$, and chooses the right box with a probability $P2 = 0.4$. Based on probability theory, it can be obtained that the probability of picking a red ball is 0.6, the probability of picking a green ball is 0.4, namely $p(R) = 0.6$, $p(G) = 0.4$.

Now, let us change the configuration, as shown in Figure 2. In the left box, there are still only red balls. But in the right box, there are not only red balls but also green balls. In accord with above, the exact number of balls in each box is still unknown, and the ratio of them are completely unknown. This

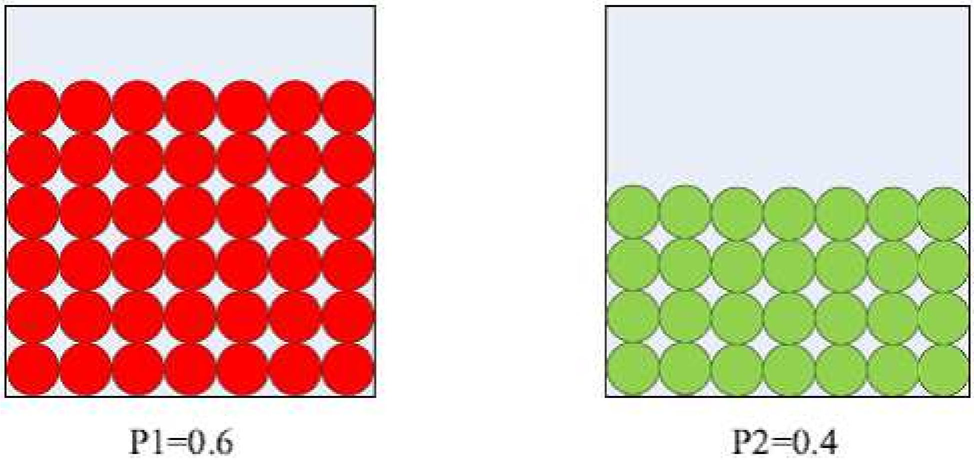

Figure 1: A game of picking ball which can be handled by probability theory

person also has 0.6 probability to choose the left box and 0.4 probability to choose the right box. The question is how possible that a red ball is picked. Obviously, in this case due to lack of adequate information, $p(R)$ and $p(G)$ cannot be obtained. Facing the situation of inadequate information, probability theory is incapable. However, if using D-S theory to analyze this problem, we can obtain a BPA that $m(R) = 0.6$ and $m(R, G) = 0.4$, which means the probability of red ball being picked is 0.6 and the probability of red ball or green ball being picked is 0.4. In the framework of D-S theory, the uncertainty has been expressed more effective. D-S theory has more ability to express uncertain information than probability theory.

2.2. Existing entropy

Entropy is associated with uncertainty, and it has been a measure of uncertainty and disorder. The concept of entropy is derived from physics [8]. In thermodynamics and statistical mechanics, the entropy often refers

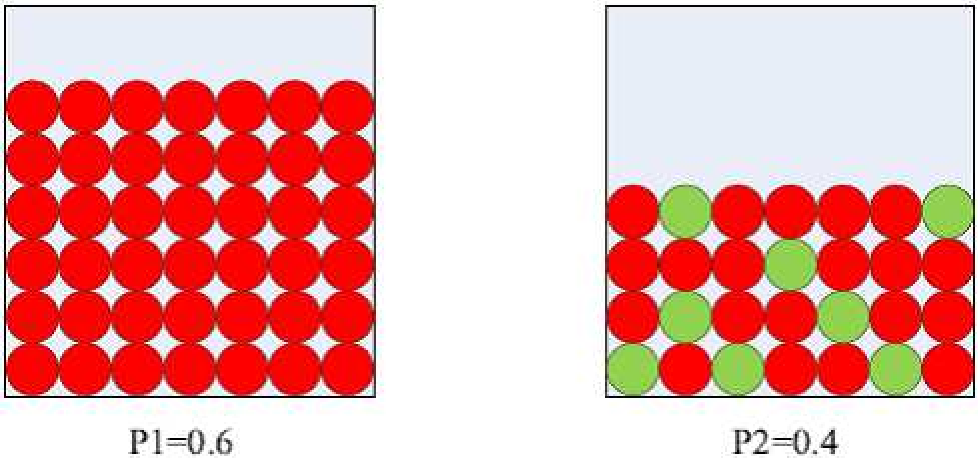

Figure 2: A game of picking ball where probability theory is unable but D-S theory is able to handle

to Boltzmann-Gibbs entropy [13]. According to Boltzmann's H theorem, the Boltzmann-Gibbs (BG) entropy of an isolated system S_{BG} is obtained in terms of the probabilities associated the distinct microscopic states available to the system given the macroscopic constraints, which has the following form

$$S_{BG} = -k \sum_{i=1}^{W} p_i \ln p_i \tag{7}$$

where k is the Boltzmann constant, W is the amount of distinct microscopic states available to the isolated system, p_i is the probability of microscopic state i satisfying $\sum\limits_{i=1}^{W} p_i = 1$. Equal probabilities, i.e. $\forall i, p_i = 1/W$, is a particular situation. In that situation, BG entropy has the following form

$$S_{BG} = k \ln W \tag{8}$$

In information theory, Shannon entropy [9] is often used to measure the information volume of a system or a process, and quantify the expected value of the information contained in a message. Information entropy, denoted as H, has a similar form with BS entropy

$$H = -\sum_{i=1}^{N} p_i \log_b p_i \tag{9}$$

where N is the amount of basic states in a state space, p_i is the probability of state i appears satisfying $\sum_{i=1}^{W} p_i = 1$, b is base of logarithm. When $b = 2$, the unit of information entropy is bit. If each state equally appears, the quantity of H has this form

$$H = \log_2 N \tag{10}$$

In information theory, quantities of H play a central role as measures of information, choice and uncertainty. For example, the Shannon entropy of the game shown in Figure 1 is $H = 0.6 \times \log_2 0.6 + 0.4 \times \log_2 0.4 = 0.9710$. But, it is worthy to notice that the uncertainty of this game shown in Figure 2 cannot be calculated by using the Shannon entropy.

According to mentioned above, no matter the BG entropy or the information entropy, the quantity of entropy is always associated with the amount of states in a system. Especially, for the case of equal probabilities, the entropy or the uncertainty of a system is a function of the quantity of states. Moreover, in that particular case, the entropy is the maximum.

3. Deng entropy

With the range of uncertainty mentioned above, Deng entropy can be presented as follows

$$E_d = -\sum_i m(F_i) \log \frac{m(F_i)}{2^{|F_i|} - 1} \tag{11}$$

where, F_i is a proposition in mass function m, and $|F_i|$ is the cardinality of F_i. As shown in the above definition, Deng entropy, formally, is similar with the classical Shannon entropy, but the belief for each proposition F_i is divided by a term $(2^{|F_i|} - 1)$ which represents the potential number of states in F_i (of course, the empty set is not included).

Specially, Deng entropy can definitely degenerate to the Shannon entropy if the belief is only assigned to single elements. Namely,

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

4. The new information unit:D

In the section, a lot of examples are given to show the effectiveness of Deng entropy.

Example 1. *Assume there is a mass function $m(a) = 1$, the associated Shannon entropy H and Deng entropy E_d are calculated as follows.*

$$H = 1 \times \log 1 = 0\boldsymbol{B}$$

$$E_d = 1 \times \log \frac{1}{2^1 - 1} = 0\boldsymbol{D}$$

*It means that in the first situation, the information volume is 0 Bit (**B**) while in the second situation, the information volume is 0 Deng (**D**)*

Example 2. *Given a frame of discernment* $X = \{a, b, c\}$*, for a mass function* $m(a) = m(b) = m(c) = 1/3$*, the information volume based on Shannon entropy* H *and Deng entropy* E_d *are as follows*

$$H = -\frac{1}{3} \times \log \frac{1}{3} - \frac{1}{3} \times \log \frac{1}{3} - \frac{1}{3} \times \log \frac{1}{3} = 1.5850\boldsymbol{B}$$

$$E_d = -\frac{1}{3} \times \log \frac{1/3}{2^1 - 1} - \frac{1}{3} \times \log \frac{1/3}{2^1 - 1} - \frac{1}{3} \times \log \frac{1/3}{2^1 - 1} = 1.5850\boldsymbol{D}$$

Clearly, Example 1 and 2 have shown that the results of Shannon entropy and Deng entropy are identical. At this situation, each hypothesis is exclusive with each other. The unit Deng (**D**) is degenerated as Bit (**B**).

Example 3. *Let's consider Figure 1 and Figure 2. The information volume can be calculated as follows*

$$-0.6 \times \log 0.6 - 0.4 \times \log 0.4 = 0.9710\boldsymbol{B}$$

$$-0.6 \times \log \frac{0.6}{2^1 - 1} - 0.4 \times \log \frac{0.4}{2^2 - 1} = 1.6049\boldsymbol{D}$$

Example 4. *Given a frame of discernment* $X = \{a, b, c\}$*, for a mass function* $m(a, b, c) = 1$*, the information volume is calculated as follows*

$$-1 \times \log \frac{1}{2^3 - 1} = 2.8074\boldsymbol{D}$$

For another mass function $m(a) = m(b) = m(c) = m(a, b) = m(a, c) = m(b, c) = m(a, b, c) = 1/7$*, the information volume is calculated as follows*

$$\begin{aligned} &-\tfrac{1}{7} \times \log \tfrac{1/7}{2^1-1} - \tfrac{1}{7} \times \log \tfrac{1/7}{2^1-1} - \tfrac{1}{7} \times \log \tfrac{1/7}{2^1-1} \\ &\quad -\tfrac{1}{7} \times \log \tfrac{1/7}{2^2-1} - \tfrac{1}{7} \times \log \tfrac{1/7}{2^2-1} - \tfrac{1}{7} \times \log \tfrac{1/7}{2^2-1} - \tfrac{1}{7} \times \log \tfrac{1/7}{2^3-1} \\ &= 3.8877\boldsymbol{D} \end{aligned}$$

The following example is used in [10].

Example 5. *In this section, we will show the irrationality of the requirement of range through an illustrative example. Let us consider an example as follows. Suppose there are 32 students participating in a course examination. After the examination finished, a student won the first place. In order to know who is the first one, we go to ask their course teacher. But the teacher doesn't want to directly tell us. Instead, she just answers "Yes" or "No" to our questions. The problem is how many times do we need ask at most in order to know who* ***IS*** *the first* ***ONE****?*

Assume the times is t*, it is easy to answer the problem through calculating the information volume by using information entropy*

$$t = \log_2 32 = 5 \tag{12}$$

For example, the first student's number, denoted as x*, is assumed to be No.2. The top 1 student No.2 can be found through five times asking at most. That means the information volume is 5* ***B****.*

Now, let's consider another situation. Assume we have been told that there are students tied for first. In this case, how many times do we need ask at most to know who ***ARE*** *the first* ***ONES****?*

In this case, obviously

$$t \geq \log_2 32 \tag{13}$$

According to Deng entropy, the information volume is as follows

$$\begin{aligned} E_d = & \tfrac{1}{2^{32}-1} \times \log_2(\tfrac{1}{2^{32}-1}/(2^1-1)) + \tfrac{1}{2^{32}-1} \times \log_2(\tfrac{1}{2^{32}-1}/(2^1-1)) + \cdots \\ & + \tfrac{1}{2^{32}-1} \times \log_2(\tfrac{1}{2^{32}-1}/(2^{32}-1)) \\ \approx & \, 48\boldsymbol{D} \end{aligned} \tag{14}$$

One thing should be pointed that, in [10], our conclusion is that we need 32 times to determine the top 1 student(s). However, it is not correct since that, according to the result above, we need 48 times to obtain the result. The difference of these two values is the information volume caused by Entangled state.

5. Conclusion

For the aleatoric uncertain information expressed by PDF, information entropy proposed by Shannon [9] is a good measure. However, with respect to other uncertain information including epistemic, irreducible, reducible and inferential uncertainty, classical information entropy is invalid. The "bit" can be used to measure information volume with Shannon entropy. When the assumption is violated such as the "Entangled state", the bit is not suitable to measure the information volume. To address this issue, a new information unit, called as "Deng" and abbreviated as "**D**", is proposed based on Deng entropy. Unit Deng is the generalization of unit bit, while the former is based on Deng entropy and the latter is based on Shannon entropy. The proposed information unit may be used in entangle information processing and quantum information processing.

Acknowledgments

The author greatly appreciates Professor Shan Zhong, the China academician of the Academy of Engineering, for his encouragement to do this research, and Professor Yugeng Xi in Shanghai Jiao Tong University for his support to this work. Professor Sankaran Mahadevan in Vanderbilt University discussed many relative topics about uncertain modelling. The author's Ph.D students in Shanghai Jiao Tong University, Peida Xu and Xiaoyan Su, discussed the uncertain measure. The author's Ph.D students in Southwest University, Xinyang Deng and Qing Wang, graduate students in Southwest

University, Qi Zhang and Meizhu Li, do some numerical experiments of this work. The author greatly appreciates the continuous supporting for the last ten years. This work is partially supported by National Natural Science Foundation of China, Grant Nos. 30400067, 60874105 and 61174022, Chongqing Natural Science Foundation for distinguished scientist, Grant No. CSCT, 2010BA2003, Program for New Century Excellent Talents in University, Grant No. NCET-08-0345, Shanghai Rising-Star Program Grant No.09QA1402900, the Chenxing Scholarship Youth Found of Shanghai Jiao Tong University, Grant No. T241460612, the open funding project of State Key Laboratory of Virtual Reality Technology and Systems, Beihang University (Grant No.BUAA-VR-14KF-02).

References

[1] W. Feller, An introduction to probability theory and its applications, Vol. 2, John Wiley & Sons, 2008.

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

[3] D. Dubois, H. M. Prade, H. Farreny, R. Martin-Clouaire, C. Testemale, Possibility theory: an approach to computerized processing of uncertainty, Vol. 2, Plenum press New York, 1988.

[4] A. P. Dempster, Upper and lower probabilities induced by a multivalued mapping, Annals of Mathematics and Statistics 38 (2) (1967) 325–339.

[5] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, Princeton, 1976.

[6] Y. Deng, Generalized evidence theory, Applied Intelligence (2015) Accepted.

[7] Y. Deng, D numbers: Theory and applications, International Journal of Computer & Information Sciences 9 (9) (2012) 2421–2428.

[8] R. Clausius, The mechanical theory of heat: with its applications to the steam-engine and to the physical properties of bodies, J. van Voorst, 1867.

[9] C. E. Shannon, A mathematical theory of communication, ACM SIGMOBILE Mobile Computing and Communications Review 5 (1) (2001) 3–55.

[10] X. Deng, Y. Deng, On the axiomatic requirement of range to measure uncertainty, Physica A: Statistical Mechanics and its Applications 406 (2014) 163–168.

[11] Y. Deng, Deng entropy: a generalized shannon entropy to measure uncertainty (2015) http://vixra.org/abs/1502.0222.

[12] Schr?dinger's cat http://en.wikipedia.org/wiki/Schr?dinger.

[13] J. L. Lebowitz, Boltzmann's entropy and time's arrow, Physics today 46 (1993) 32–32.