# Deng entropy: a generalized Shannon entropy to measure uncertainty

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**

Shannnon entropy is an efficient tool to measure uncertain information. However, it cannot handle the more uncertain situation when the uncertainty is represented by basic probability assignment (BPA), instead of probability distribution, under the framework of Dempster Shafer evidence theory. To address this issue, a new entropy, named as Deng entropy, is proposed. The proposed Deng entropy is the generalization of Shannnon entropy. If uncertain information is represented by probability distribution, the uncertain degree measured by Deng entropy is the same as that of Shannnon's entropy. Some numerical examples are illustrated to shown the efficiency of Deng entropy.

*Keywords:* Uncertainty measure, Entropy, Deng entropy, Shannnon entropy, Dempster-Shafer evidence theory

*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. How to measure the uncertainty has attracted much attention [1, 2]. Various categorizations exist to accommodate different kinds of uncertainties. Numerous uncertainty theories has been developed, such as probability theory [3], fuzzy set theory [4], possibility theory [5], Dempster-Shafer evidence theory [6, 7], rough sets[8], DSmT[9, 10], generalized evidence theory [11] and D numbers[12, 13, 14, 15].

Since firstly proposed by Clausius in 1865 for thermodynamics [16],the study of uncertainty and entropy attracts great interests and various types of entropies are developed, such as information entropy [17], Tsallis entropy [18], nonadditive entropy [19, 20, 21]. Information entropy [17], derived from the Boltzmann-Gibbs (BG) entropy [22] in thermodynamics and statistical mechanics, has been an indicator to measures aleatoric uncertainty which is associated with the probability density function (PDF).

For the aleatoric uncertain information expressed by PDF, information entropy proposed by Shannon [17] is a good measure. However, with respect to other uncertain information including epistemic, irreducible, reducible and inferential uncertainty, classical information entropy is invalid. In this paper, one of these uncertainties, epistemic uncertainty, is taken into consideration. Dempster-Shafer theory evidence theory[6, 7] is mainly proposed to handle such uncertainty. In Dempster-Shafer evidence theory, the epistemic uncertainty simultaneously contains nonspecificity and discord [23] which are coexisting in a basic probability assignment function (BPA). Several uncertainty measures, such as AU [24, 25], AM [23], have been proposed to quantify such uncertainty in Dempster-Shafer theory. What's more, five axiomatic requirements have been further built in order to develop a justifiable measure. These five axiomatic requirements are range, probabilistic consistency, set consistency, additivity, subadditivity, respectively [26]. However, existing methods are not efficient to measure uncertain degree of BPA. To address this issue, a new entropy, named as Deng entropy, is proposed in this paper.

The paper is organized as follows. The preliminaries Dempster-Shafer evidence theory and entropy are briefly introduced in Section 2. Section 3 presents Deng entropy. Some numerical examples are illustrated in Section 4 to show the efficiency of Deng entropy. 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[6] and then developed by Shafer[7], has emerged from their works on statistical inference and uncertain reasoning. This theory is widely applied to uncertainty modeling [27, 28], decision making [29, 30, 31, 32, 33, 34, 35, 36, 37]and information fusion [38, 39] and uncertain information processing [40]. D-S theory mainly focus on the epistemic uncertainty, but it is also valid for aleatoric uncertainty. It has many merits by contrast probability theory. First, D-S theory can handle more uncertainty in real world. In contrast to the probability theory in which probability masses can be only assigned to singleton subsets, in D-S theory the belief can be assigned to both singletons and compound sets. Second, in D-S theory, prior distribution is not needed before the combination of information from individual information sources. Third, D-S theory allows one to specify a degree of ignorance in some situations instead of being forced to be assigned for probabilities. Some basic concepts in D-S theory are introduced.

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

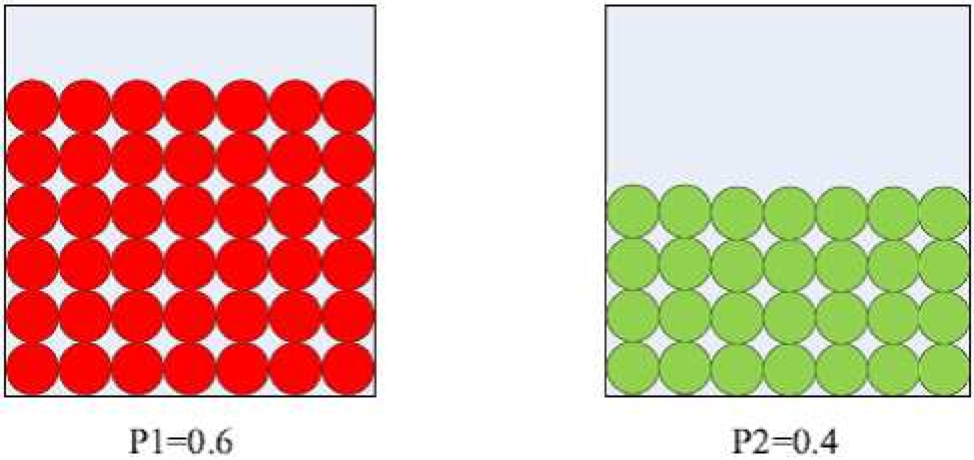


Figure 1: A game of picking ball which can be handled by probability theory

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 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.

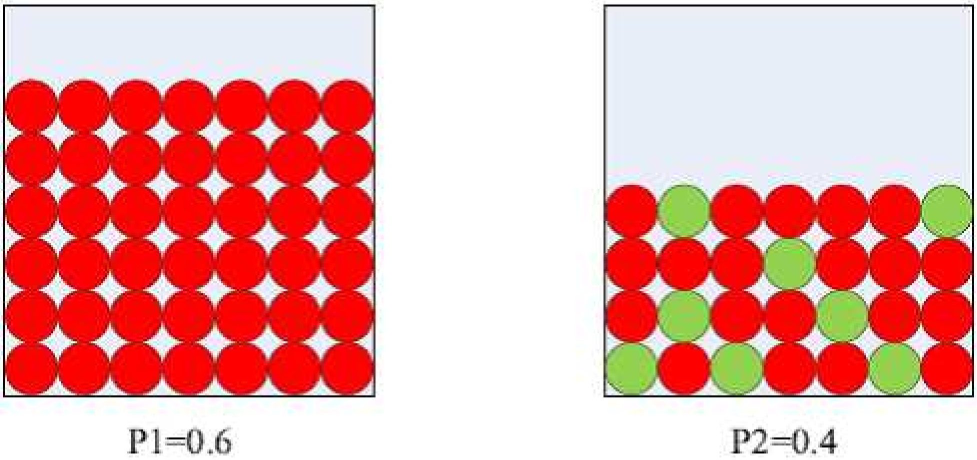


Figure 2: A game of picking ball where probability theory is unable but D-S theory is able to handle

*2.2. Existing entropy and open issue*

Entropy is associated with uncertainty, and it has been a measure of uncertainty and disorder. The concept of entropy is derived from physics [16]. In thermodynamics and statistical mechanics, the entropy often refers to Boltzmann-Gibbs entropy [22]. 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 $\sum_{i=1}^{W} p_i = 1$. 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 [17] 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. Numerical examples and discussions

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.*

$$H = 1 \times \log 1 = 0$$

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

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

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

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

Clearly, Example 1 and 2 have shown that the results of Shannon entropy and Deng entropy are identical when the belief is only assigned on single elements.

**Example 3.** *Given a frame of discernment $X = \{a, b, c\}$, for a mass function $m(a, b, c) = 1$,*

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

*For another mass function $m(a) = m(b) = m(c) = m(a, b) = m(a, c) = m(b, c) = m(a, b, c) = 1/7$,*

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

**Example 4.** *Given a frame of discernment $X = \{a_1, a_2, \cdots, a_N\}$, let us consider three special cases of mass functions as follows.*

- $m_1(F_i) = m_1(F_j)$ *and* $\sum_i m_1(F_i) = 1, \quad \forall F_i, F_j \subseteq X,\ F_i, F_j \neq \emptyset.$
- $m_2(X) = 1.$
- $m_3(a_1) = m_3(a_2) = \cdots = m_3(a_N) = 1/N.$

*Their associated Deng entropies change with $N$, as shown in Figure 3.*

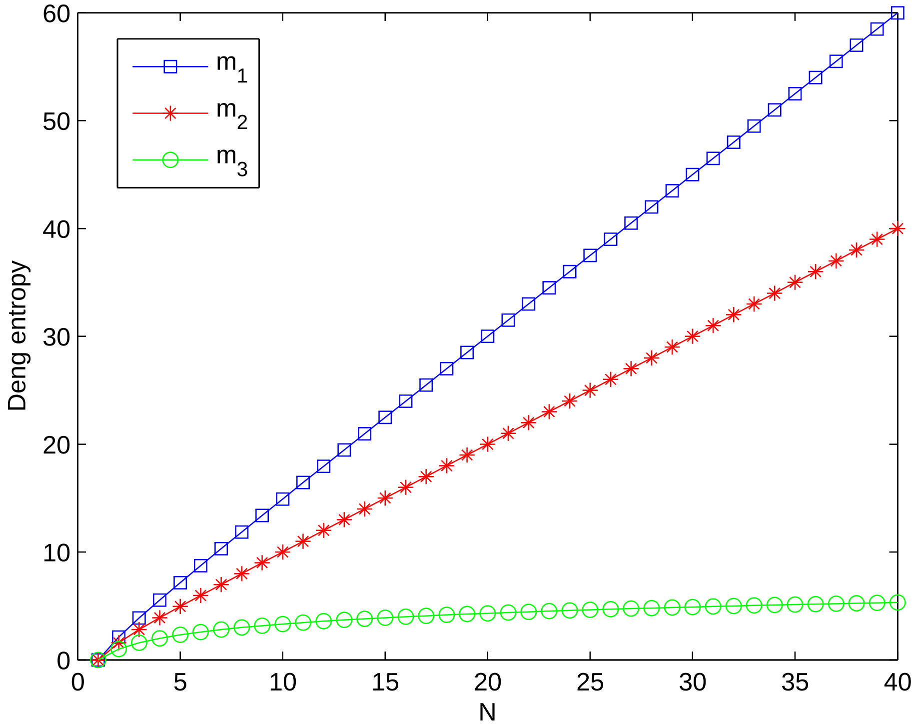


Figure 3: Deng entropy as a function of the size of frame of discernment in three types of mass functions

**Example 5.** *Given a frame of discernment X with 15 elements which are denoted as element 1, element 2, etc. A mass function is shown as follows.*

$$m(\{3,4,5\}) = 0.05, m(\{6\}) = 0.05, m(A) = 0.8, m(X) = 0.1$$

*Table 1 lists various Deng entropies when A changes, which is graphically shown in Figure 4. The results shows that the entropy of m increases monotonously with the rise of the size of subset A. It is rational that the entropy increases when the uncertainty involving a mass function increases.*

Table 1: Deng entropy when $A$ changes

| Cases | Deng entropy |
|---|---|
| $A = \{1\}$ | 1.8454 |
| $A = \{1, 2\}$ | 2.7242 |
| $A = \{1, 2, 3\}$ | 3.4021 |
| $A = \{1, \cdots, 4\}$ | 4.0118 |
| $A = \{1, \cdots, 5\}$ | 4.5925 |
| $A = \{1, \cdots, 6\}$ | 5.1599 |
| $A = \{1, \cdots, 7\}$ | 5.7207 |
| $A = \{1, \cdots, 8\}$ | 6.2784 |
| $A = \{1, \cdots, 9\}$ | 6.8345 |
| $A = \{1, \cdots, 10\}$ | 7.3898 |
| $A = \{1, \cdots, 11\}$ | 7.9447 |
| $A = \{1, \cdots, 12\}$ | 8.4994 |
| $A = \{1, \cdots, 13\}$ | 9.0540 |
| $A = \{1, \cdots, 14\}$ | 9.6086 |

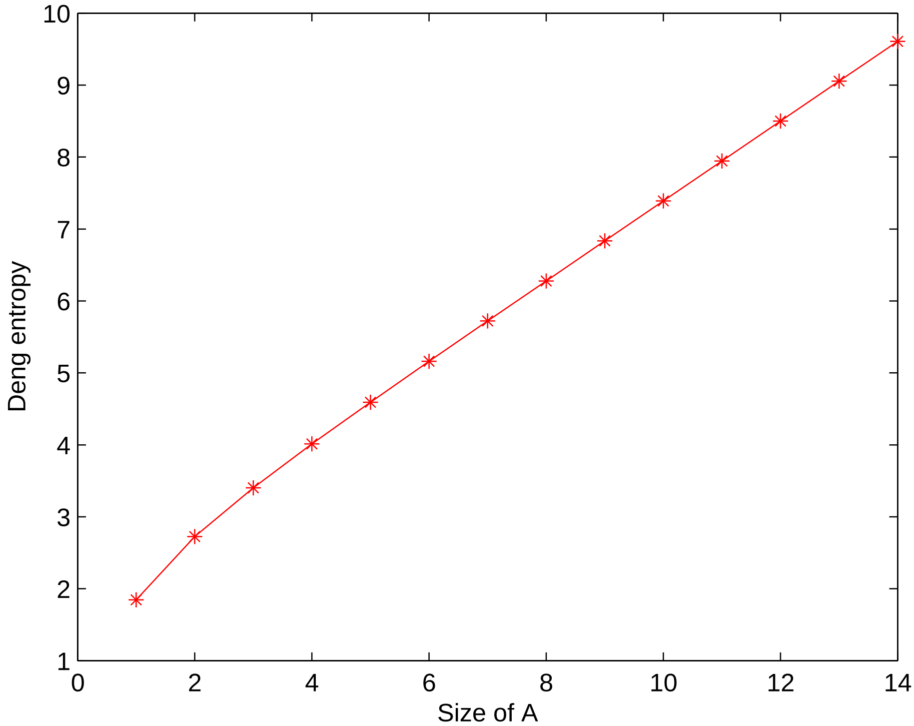


Figure 4: Deng entropy as a function of the size of $A$

## 5. Conclusion

Dempster Shafer evidence theory has widely used in many applications due to its advantages to handle the aleatoric and epistemic uncertainty. However, how to measure uncertain degree in evidence theory is still an open issue. The main contribution of this paper is that a new entropy, named as Deng entropy, is presented. Deng entropy is the generalization of Shannon entropy. In the case when the BPA is degenerated as probability, Deng entropy is degenerated as Shannon entropy. However, Numerical examples are illustrated to show the efficiency of Deng entropy. Some properties of Deng entropy are discussed. The new entropy provides a promising way to measure uncertain degree. One of the ongoing works is to present the cross entropy of BPA, which can be easily derived.


## 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] G. J. Klir, Uncertainty and information: foundations of generalized information theory, John Wiley & Sons, 2005.

[2] S. Sankararaman, S. Mahadevan, Model validation under epistemic uncertainty, Reliability Engineering & System Safety 96 (9) (2011) 1232–1241.

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

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

[5] 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.

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

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

[8] Z. Pawlak, Rough sets, International Journal of Computer & Information Sciences 11 (5) (1982) 341–356.

[9] J. Dezert, Foundations for a new theory of plausible and paradoxical reasoning, Information and Security 9 (2002) 13–57.

[10] F. Smarandache, J. Dezert, Advances and Applications of DSmT for Information Fusion (Collected works), second volume: Collected Works, Vol. 2, Infinite Study, 2006.

[11] Y. Deng, Generalized evidence theory, Applied Intelligence (2015) Second revision.

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

[13] X. Deng, Y. Hu, Y. Deng, S. Mahadevan, Supplier selection using AHP methodology extended by D numbers, Expert Systems with Applications 41 (1) (2014) 156–167.

[14] X. Deng, Y. Hu, Y. Deng, S. Mahadevan, Environmental impact assessment based on D numbers, Expert Systems with Applications 41 (2) (2014) 635–643.

[15] X. Deng, F. T. Chan, R. Sadiq, S. Mahadevan, Y. Deng, D-cfpr: D numbers extended consistent fuzzy preference relations, Knowledge-Based Systems 73 (1) (2015) 61–68.

[16] 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.

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

[18] C. Tsallis, Possible generalization of Boltzmann-Gibbs statistics, Journal of Statistical Physics 52 (1-2) (1988) 479–487.

[19] C. Tsallis, Nonadditive entropy: the concept and its use, The European Physical Journal A 40 (3) (2009) 257–266.

[20] C. Tsallis, Nonadditive entropy and nonextensive statistical mechanics-an overview after 20 years, Brazilian Journal of Physics 39 (2A) (2009) 337–356.

[21] C. Tsallis, U. Tirnakli, Nonadditive entropy and nonextensive statistical

mechanics–some central concepts and recent applications, in: Journal of Physics: Conference Series, Vol. 201, 2010, p. 012001.

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

[23] A.-L. Jousselme, C. Liu, D. Grenier, E. Bosse, Measuring ambiguity in the evidence theory, IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans 36 (5) (2006) 890–903.

[24] Y. Maeda, H. T. Nguyen, H. Ichihashi, Maximum entropy algorithms for uncertainty measures, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 1 (01) (1993) 69–93.

[25] D. Harmanec, G. J. Klir, Measuring total uncertainty in dempster-shafer theory: A novel approach, International journal of general system 22 (4) (1994) 405–419.

[26] G. J. Klir, M. J. Wierman, Uncertainty-based information: elements of generalized information theory, Springer, 1999.

[27] P. Smets, R. Kennes, The transferable belief model, Artificial Intelligence 66 (2) (1994) 191–234.

[28] T. L. Wickramarathne, K. Premaratne, M. N. Murthi, Toward efficient computation of the dempster–shafer belief theoretic conditionals, Cybernetics, IEEE Transactions on 43 (2) (2013) 712–724.

[29] L. V. Utkin, A new ranking procedure by incomplete pairwise comparisons using preference subsets, Intelligent Data Analysis 13 (2) (2009) 229–241.

[30] V.-N. Huynh, T. T. Nguyen, C. A. Le, Adaptively entropy-based weighting classifiers in combination using dempster–shafer theory for word sense disambiguation, Computer Speech & Language 24 (3) (2010) 461–473.

[31] Y. Zhang, X. Deng, D. Wei, Y. Deng, Assessment of E-Commerce security using AHP and evidential reasoning, Expert Systems with Applications 39 (3) (2012) 3611–3623.

[32] Y. Deng, F. Chan, A new fuzzy dempster MCDM method and its application in supplier selection, Expert Systems with Applications 38 (8) (2011) 9854–9861.

[33] Y. Deng, W. Jiang, R. Sadiq, Modeling contaminant intrusion in water distribution networks: A new similarity-based DST method, Expert Systems with Applications 38 (1) (2011) 571–578.

[34] Y. Deng, R. Sadiq, W. Jiang, S. Tesfamariam, Risk analysis in a linguistic environment: A fuzzy evidential reasoning-based approach, Expert Systems with Applications 38 (12) (2011) 15438–15446.

[35] B. Kang, Y. Deng, R. Sadiq, S. Mahadevan, Evidential cognitive maps, Knowledge-Based Systems 35 (2012) 77–86.

[36] R. R. Yager, N. Alajlan, Decision making with ordinal payoffs under dempster–shafer type uncertainty, International Journal of Intelligent Systems 28 (11) (2013) 1039–1053.

[37] J.-B. Yang, D.-L. Xu, Evidential reasoning rule for evidence combination, Artificial Intelligence 205 (2013) 1–29.

[38] F. Cuzzolin, A geometric approach to the theory of evidence, Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on 38 (4) (2008) 522–534.

[39] J. Schubert, Conflict management in Dempster-Shafer theory using the degree of falsity, International Journal of Approximate Reasoning 52 (3) (2011) 449–460.

[40] S. Rao, K. Annamdas, A comparative study of evidence theories in the modeling, analysis, and design of engineering systems, Journal of Mechanical Design 135 (6) (2013) 061006.