The maximum Deng entropy

Bingyi Kang[a], Yong Deng[a,b,c,*]

[a]*School of Computer and Information Science, Southwest University, Chongqing, 400715, China*
[b]*School of Electronics and Information, Northwestern Polytechnical University, Xi'an, 710072, China*
[c]*Department of Civil & Environmental Engineering, School of Engineering, Vanderbilt University, Nashville, TN, 37235, USA*

Abstract

Dempster Shafer evidence theory has widely used in many applications due to its advantages to handle uncertainty. Deng entropy, has been proposed to measure the uncertainty degree of basic probability assignment in evidence theory. It is the generalization of Shannon entropy since that the BPA is degenerated as probability, Deng entropy is identical to Shannon entropy. However, the maximal value of Deng entropy has not been disscussed until now. In this paper, the condition of the maximum of Deng entropy has been disscussed and proofed, which is usefull for the application of Deng entropy.
Keywords: Uncertainty measure, Entropy, Deng entropy, Shannnon entropy, Dempster-Shafer evidence theory

*Corresponding author: School of Computer and Information Science, Southwest University, Chongqing, 400715, China.
Email address: ydeng@swu.edu.cn,prof.deng@hotmail.com (Yong Deng)

1. Introduction

How to measure the uncertainty has attracted much attention [1, 2]. A lot of 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].

Since firstly proposed by Clausius in 1865 for thermodynamics [14], various types of entropies are presented, such as information entropy [15], Tsallis entropy [16], nonadditive entropy [17, 18, 19]. Information entropy [15], derived from the Boltzmann-Gibbs (BG) entropy [20] in thermodynamics and statistical mechanics, has been an indicator to measures uncertainty which is associated with the probability density function (PDF).

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 [21] which are coexisting in a basic probability assignment function (BPA). Several uncertainty measures, such as AU [22, 23], AM [21], 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 [24]. Existing methods are not efficient to measure uncertain degree of BPA. To address this issue, a new entropy, named as Deng entropy [25], is proposed to measure the uncertainty of basic probability assignment for the evidence theory. In this paper, a discussin of the maximal value of Deng entropy has been discussed and proofed, which is useful for the real application of Deng entropy.

The paper is organized as follows. The preliminaries Dempster-Shafer evidence theory and Deng entropy are briefly introduced in Section 2. Section 3 proofed and discussed the maximal value of Deng entropy. Finally, this paper is concluded in Section 4.

2. Preliminaries

In this section, some preliminaries are briefly introduced.

2.1. Dempster-Shafer evidence theory

Dempster-Shafer theory (short for D-S theory) is presented by Dempster and Shafer [6, 7]. This theory is widely applied to uncertainty modeling [26, 27], decision making [28, 29, 30, 31, 32, 33, 34], information fusion [35, 36] and uncertain information processing [37]. D-S theory has many advantages to handle uncertain information. 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 $K < 1$.

2.2. Deng entropy [25]

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

$$E_d = -\sum_i m(F_i) \log \frac{m(F_i)}{2^{|F_i|} - 1} \tag{7}$$

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,

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

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

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

For mass function $m(a) = m(b) = m(c) = 1/19, m(a, b) = m(a, c) = m(b, c) = 3/19, m(a, b, c) = 7/19$*,*

$$\begin{aligned} E_d = & -\tfrac{1}{19} \times \log \tfrac{1/19}{2^1-1} - \tfrac{1}{19} \times \log \tfrac{1/19}{2^1-1} - \tfrac{1}{19} \times \log \tfrac{1/19}{2^1-1} \\ & -\tfrac{3}{19} \times \log \tfrac{3/19}{2^2-1} - \tfrac{3}{19} \times \log \tfrac{3/19}{2^2-1} - \tfrac{3}{19} \times \log \tfrac{3/19}{2^2-1} - \tfrac{7}{19} \times \log \tfrac{7/19}{2^3-1} \\ = & \; 4.2479 \end{aligned}$$

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* $\sum_i m_1(F_i) = 1, \quad \forall F_i, F_j \subseteq X, \ F_i, F_j \neq \emptyset.$
- $m_2(X) = 1.$

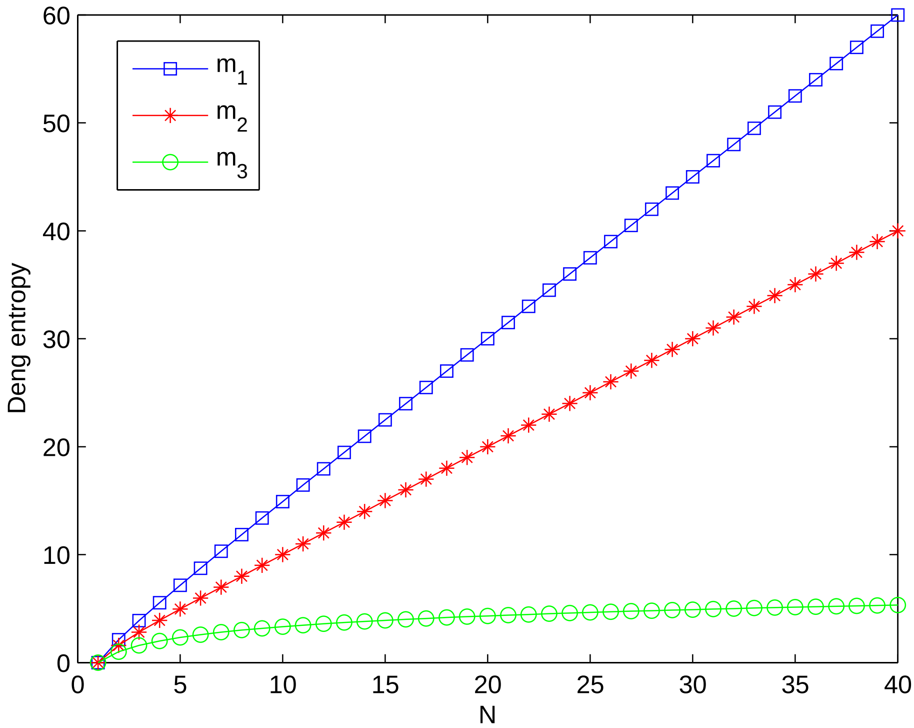

Figure 1: Deng entropy as a function of the size of frame of discernment in three types of mass functions

- $m_3(a_1) = m_3(a_2) = \cdots = m_3(a_N) = 1/N$.

Their associated Deng entropies change with N, as shown in Figure 1.

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 2. 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

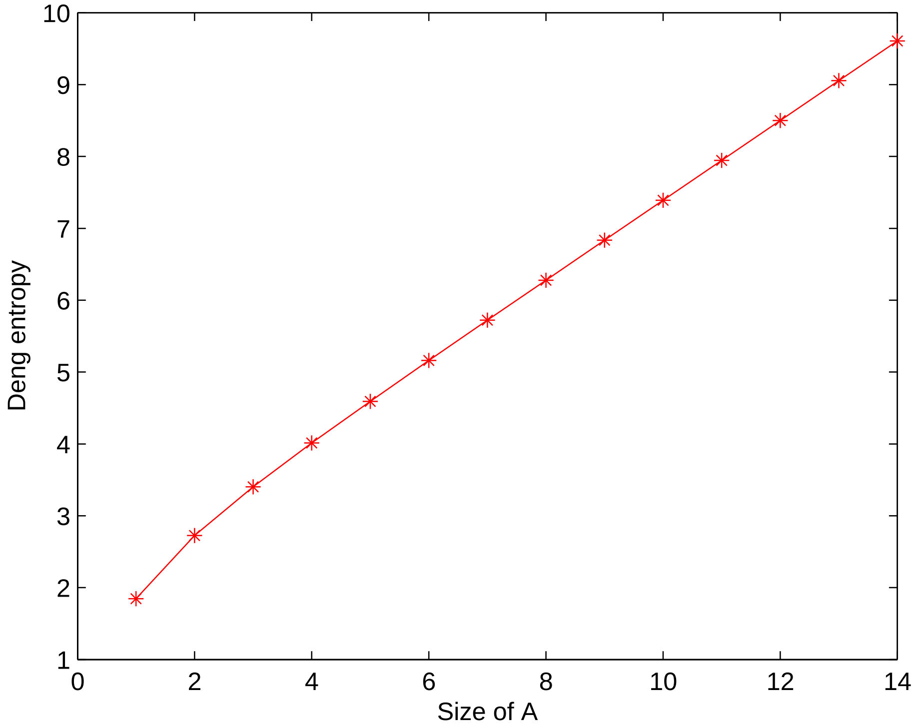

Figure 2: Deng entropy as a function of the size of A

3. Discussion of the maximum Deng entropy

Theorem 3.1 (The maximum Deng entropy)**.** *The maximum Deng entropy:* $E_d = -\sum\limits_i m(F_i) \log \frac{m(F_i)}{2^{|F_i|}-1}$ *if and only if* $m(F_i) = \frac{2^{|F_i|}-1}{\sum\limits_i 2^{|F_i|}-1}$

Proof.

$$D = -\sum_i m(F_i) \log \frac{m(F_i)}{2^{|F_i|}-1} \tag{8}$$

$$\sum_i m(F_i) = 1 \tag{9}$$

$$D_0 = -\sum_i m(F_i) \log \frac{m(F_i)}{2^{|F_i|}-1} + \lambda \left(\sum_i m(F_i) - 1 \right) \tag{10}$$

$$\frac{\partial D_0}{\partial m(F_i)} = -\log \frac{m(F_i)}{2^{|F_i|}-1} - m(F_i) \frac{1}{\frac{m(F_i)}{2^{|F_i|}-1} \ln a} \cdot \frac{1}{2^{|F_i|}-1} + \lambda = 0 \tag{11}$$

$$-\log \frac{m(F_i)}{2^{|F_i|}-1} - \frac{1}{\ln a} + \lambda = 0 \tag{12}$$

$$\frac{m(F_1)}{2^{|F_1|}-1} = \frac{m(F_2)}{2^{|F_2|}-1} = \cdots = \frac{m(F_n)}{2^{|F_n|}-1} \tag{13}$$

$$\frac{m(F_1)}{2^{|F_1|}-1} = \frac{m(F_2)}{2^{|F_2|}-1} = \cdots = \frac{m(F_n)}{2^{|F_n|}-1} = k \tag{14}$$

$$m(F_i) = k\left(2^{|F_i|}-1\right) \tag{15}$$

$$k = \frac{1}{\sum\limits_i 2^{|F_n|}-1} \tag{16}$$

$$m(F_i) = \frac{2^{|F_i|}-1}{\sum\limits_i 2^{|F_i|}-1} \tag{17}$$

□

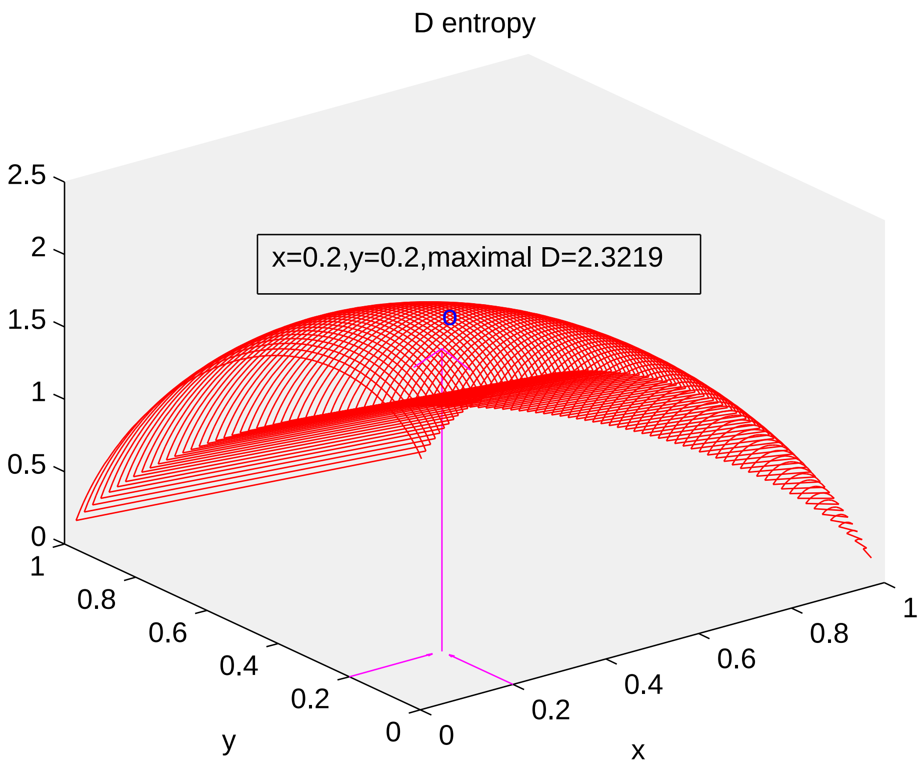

Figure 3: Deng entropy when frame of discernment is $\{a, b\}$

Let $X = \{a, b\}$, $\mathrm{m}(a) = x, m(b) = y, m(a, b) = 1 - x - y$, which is under the condition of $0 \leq x \leq 1, 0 \leq y \leq 1, 0 \leq 1 - x - y \leq 1$, then relation between Deng entropy and (x, y) can be denoted as the Figure 3.

4. Conclusion

Deng entropy which is the generalization of Shannon entropy is used to measure uncertain degree in evidence theory. In the case when the BPA is degenerated as probability, Deng entropy is degenerated as Shannon entropy.

In this paper, the condition of the maximal value of Deng entropy has been discussed and proofed, which is helpful for the application of Deng entropy.

Acknowledgments

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) doi 10.1007/s10489–015–0661–2.

[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] 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.

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

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

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

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

[19] 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.

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

[21] 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.

[22] 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.

[23] 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.

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

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

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

[27] 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.

[28] L. V. Utkin, A new ranking procedure by incomplete pairwise compar-

isons using preference subsets, Intelligent Data Analysis 13 (2) (2009) 229–241.

[29] 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.

[30] 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.

[31] 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.

[32] 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.

[33] 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.

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

[35] 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.

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

[37] 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.