Conditional Deng Entropy, Joint Deng Entropy and Generalized Mutual Information

Haoyang Zheng[a], Yong Deng[a,b,*]

[a] School of Computer and Information Science, Southwest University, Chongqing 400715, China
[b] School of Engineering, Vanderbilt University, TN 37235, USA

Abstract

Shannon entropy, conditional entropy, joint entropy and mutual information, can estimate the chaotic level of information. However, these methods could only handle certain situations. Based on Deng entropy, this paper introduces multiple new entropy to estimate entropy under multiple interactive uncertain information: conditional Deng entropy is used to calculate entropy under conditional basic belief assignment; joint Deng entropy could calculate entropy by applying joint basic belief assignment distribution; generalized mutual information is applied to estimate the uncertainty of information under knowing another information. Numerical examples are used for illustrating the function of new entropy in the end.

Keywords: Uncertainty measure, Dempster-Shafer evidence theory, Information entropy, Deng entropy.

---

1. Introduction

How to process marvelous information is a big issue in big data era. Many papers, based on their domain knowledge, proposed methods to process information. Evidence

*Corresponding author at: School of Computer and Information Science, Southwest University, Chongqing 400715, China. Tel.: +86 23 68254555; fax: +86 23 68254555.

Email address: zhenghaoyangswu@163.com, ydeng@swu.edu.cn, professordeng@163.com (Yong Deng)

theory [1] is one of the effective ways to obtain accurate information through combining uncertain information. After that, numerous methods to process uncertain information was developed, such as Yager's method [2], Murphy's method [3], Deng's method [4] [5], fuzzy number [6] [7], D numbers [8], ect.

C.E.Shannon [9] proposed a useful way Shannon entropy, to describe the chaotic level of information. Shannon entropy comes from the concept "entropy" [10], proposed by Papoulis in 1860s. Papoulis applied entropy in thermodynamics, and predetermined that entropy equals the ratio of the variation of heat to absolute temperature. Then C.E.Shannon extended entropy into information transmission. Based on Shannon entropy, many papers specialized in the entropy of multiple interrelated elements. Conditional entropy [10] [11] was proposed to estimate element's average uncertainty after knowing another element. Joint entropy [12] could use for estimating the uncertainty of the simultaneous occurrence of over two elements. Mutual information [13] [14] is the amount of information from a random variable, which includes other random variables. There are strong connections among conditional entropy, joint entropy and mutual information and could also expressed by some formulas.

Traditional information entropy could only be used in processing certain and exhaustive information(closed world assumption). To expand the application scope, Deng proposed Deng entropy [15].Deng entropy could handle more uncertain information, such as the information represented by basic belief assignment (BBA) under the framework of Dempster-Shafer evidence theory (D-S evidence theory). This work follows Deng entropy, combining the characteristics of multiple uncertain information and information entropy, proposes conditional Deng entropy, joint Deng entropy and generalized mutual information. These new information entropy inherit the characteristics of their corresponding traditional entropy. Also, when uncertain information is expressed by probability distribution, new entropy would degenerate to traditional entropy, and get the same results.

This paper will be organized as follows: Section 2 introduces some concepts and equations in D-S evidence theory. In Section 3, some proposed entropy is briefly introduced. Next, the conditional Deng entropy, joint Deng entropy and generalized mutual information is represented. Then in Section 5, the effect of these entropy is shown by some simple experiments. Finally, this paper is concluded.

2. D-S Evidence Theory

2.1. The frame of discernment

The frame of discernment (FD) is proposed to describe the whole circumstances in the event.

Definition 1. Θ is used to describe a set of mutually exclusive and collectively exhaustive elements E_i, which is indicated by

$$\Theta = \{E_1, E_2, \cdots, E_i, \cdots, E_N\} \tag{1}$$

Set Θ is called FD. The power set of Θ is denoted by 2^Θ, and

$$2^\Theta = \{\emptyset, \{E_1\}, \cdots, \{E_N\}, \{E_1, E_2\}, \cdots, \{E_1, E_2, \cdots, E_i\}, \cdots, \Theta\} \tag{2}$$

where $\emptyset$ is an empty set. For every single set A in 2^Θ, A is called an elementary proposition.

2.2. The basic mass assignment

Definition 2. A mass function m is a mapping from 2^Θ to a probability interval $[0, 1]$, formally defined by:

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

which satisfies the following conditions:

$$m(\emptyset) = 0 \qquad \sum_{A \in 2^\Theta} m(A) = 1 \quad 0 \leq m(A) \leq 1 \quad A \in 2^\Theta \tag{4}$$

If m(∅) = 0, it tells the frame of discernment is set in closed world assumption(CWA). On the contrary, the open world assumption(OWA), proposed by Smets [16], distributed the belief to ∅ : m(∅) $\neq$ 0.

2.3. Combining rule of evidence

Definition 3. If $m(A) > 0$, A is a focal element, and the set of some focal elements is named a body of evidence (BOE). When multiplying BOEs is available, the Dempster's combination rule can be used to obtain the combined evidence:

$$m(A) = \frac{\sum\limits_{B \bigcap C = A} m_1(B) m_2(C)}{1 - K} \tag{5}$$

where $K = \sum\limits_{B \bigcap C = \emptyset} m_1(B) m_2(C)$ is a normalized constant, which is called conflict. The combination rule establishes if and only if when $m(\emptyset) \neq 1$.

3. Information Entropy

Entropy, which is firstly applied in thermodynamics by Clausius, is an effective tool to describe the level of chaos. The bigger the entropy is, the greater the uncertainty and randomness are. Lots of studies are developed to apply entropy to different fields later.

3.1. Shannon Entropy

Similar to entropy in thermodynamics, the miscellaneous degree of information could be described by information entropy. In 1948, C.E.Shannon proposed Shannon entropy, which firstly makes the information can be qualified.

Definition 4. The entropy of a discrete random variable x_i is defined as:

$$H(x_i) = \sum_{i=1}^{N} P(x_i) \log_b \frac{1}{P(x_i)} = -\sum_{i=1}^{N} P(x_i) \log_b P(x_i) \tag{6}$$

Besides, we regulate $0 \log \frac{1}{0} = 0$. In Eq.6, b is the base of logarithms. We regulate the base number equals 2, and then we get:

$$H(X) = -\sum_{i=1}^{N} P(x_i) \log P(x_i) \tag{7}$$

In Eq.6, if we set b equals to natural number e, then the unit of entropy is nat; in Eq.7, the unit of entropy is bit.

$$1\ nat \approx 1.44\ bit$$

3.2. Joint Entropy

Joint entropy is a natural extension of entropy by means of joint probability distribution.

For two discrete random variables X and Y, the probability of the intersection of the events X=x and Y=y can be written as a function $p(x, y) = P(X = x, Y = y)$ and the corresponding probability (X, Y) is **the joint probability distribution** of X and Y.

Joint probability distributions satisfy:

$$p(x, y) \geq 0, \sum_x \sum_y p(x, y) = 1$$

where the sum is greater than any other possible values in the variables.

Definition 5. For every self information in the joint probability distribution, the entropy could be calculated as:

$$H(x, y) = -P(x, y) \log P(x, y) \tag{8}$$

Combined with the mathematical expectation of the self information, joint entropy is a measure of the uncertainty under the two variables XY.

$$H(X, Y) = -\sum_X \sum_Y P(X, Y) \log P(X, Y) \tag{9}$$

Here are some properties of joint entropy:

1. Nonnegativity:

$$H(X, Y) \geq 0$$

2. The entropy of the joint entropy is not less than that of the subsystem:

$$H(X, Y) \geq H(X);\ H(X, Y) \geq H(Y)$$

3. Subadditivity:

$$H(X,Y) \leq H(X) + H(Y)$$

3.3. Conditional Entropy

Conditional entropy is an extension of entropy based on conditional probability distribution.

The probability of an event will be effected by other events. The conditional probability of the event B, while A is known, is denoted by $P(B|A)$. It is defined by:

$$P(B|A) = \frac{P(A \bigcap B)}{P(A)} \ if P(A) > 0.$$

Definition 6. If we get the value of another random variable Y is y, then the posterior distribution of X is equal to $P(X|Y = y)$. Thus the conditional entropy of variable X under this condition is:

$$H(X|Y = y) = -\sum_{X} P(X|Y = y) \log P(X|Y = y) \tag{10}$$

While entropy $H(X)$ =is used for measuring the uncertainty of random variable X, conditional entropy $H(X|Y = y)$ measures the uncertainty after knowing Y=y. With the change of y, $H(X|Y = y)$ will change simultaneously.

Since knowing the probability distribution of Y, the expected value of the entropy of X after observing Y could be measured as:

$$\begin{aligned} H(X|Y) &= \sum_{j=1}^{N} P(Y = y_j) H(X|Y = y_j) \\ &= -\sum_{j=1}^{N} P(Y = y_j) \sum_{X} P(X|Y = y_j) \log P(X|Y = y_j) \\ &= -\sum_{X}\sum_{Y} P(Y) P(X|Y) \log P(X|Y) \\ &= -\sum_{X}\sum_{Y} P(X,Y) \log P(X|Y) \end{aligned} \tag{11}$$

Conditional entropy also follows similar rules for joint entropy. Besides, in Eq.12, which is called "Chain rule of entropy", shows the relationship among Shannon entropy, conditional entropy and joint entropy:

$$H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y) \tag{12}$$

3.4. Mutual Information

The mutual information of two random variables is a measure of the mutual dependence between the two variables. It expressed by the amount of information, represented by one random variable through the proportion of the other random variable.

Before observing Y, the uncertainty of X is $H(X)$. After knowing Y, the uncertainty of X will decrease to $H(X|Y)$.

Definition 7. The difference between $H(X)$ and $H(X|Y)$ is regarded as mutual information, and is defined as:

$$I(X;Y) = H(X) - H(X|Y) \tag{13}$$

Which means a measure of the amount of X contained in Y. Mutual information also follow several rules:

$$\begin{aligned} I(X;Y) &= \sum_{X,Y} P(X,Y) \log \frac{P(X,Y)}{P(X)P(Y)} \\ I(X;Y) &= I(Y;X) \\ I(X;Y) + H(X,Y) &= H(X) + H(Y) \end{aligned} \tag{14}$$

Fig.1 shows the relationship among these entropy.

3.5. Deng Entropy

Deng entropy is the generalization of Shannon entropy, which was first proposed by Deng [15]. It proposed an efficient way to measure uncertain information, which can not only under the situation when the uncertainty is represented by a probability distribution, but also the situation when the uncertainty is represented by BBA. Therefore,

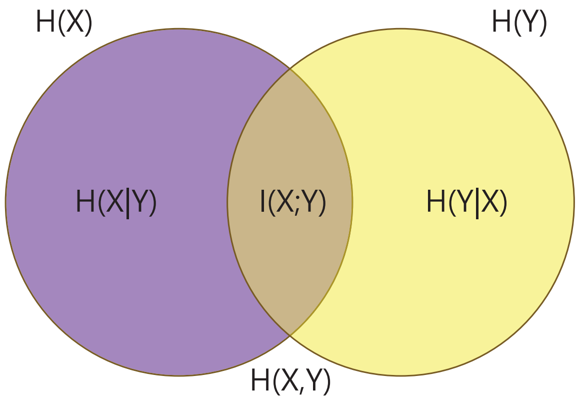

Figure 1: Relationship among multiple entropy

Deng entropy could be widely applied in D-S evidence theory. When the uncertainty is expressed as probability distribution, Deng entropy definitely degenerates to Shannon entropy. The related concepts are given below.

Definition 8. Let A_i be a proposition of BBA m; the cardinality of the set A_i is denoted by $|A_i|$. Deng entropy E_d of set A_i is defined as:

$$H(X) = -\sum_{i=1}^{n} m\left(x_i\right) \log \frac{m\left(x_i\right)}{2^{|x_i|}-1} \tag{15}$$

When the belief value is only assigned to a single element, Deng entropy can definitely degenerate to Shannon entropy, namely:

$$H(X) = -\sum_{i=1}^{n} m\left(x_i\right) \log \frac{m\left(x_i\right)}{2^{|x_i|}-1} = -\sum_{i=1}^{n} m(x_i) \log m(x_i) \tag{16}$$

4. Information Entropy under Open World

Traditional entropy can measure the amount of information under non-reliable and certain sources. However, the issue that it can not measure the amount of information under open world limits the development in this field.

4.1. *Two-dimensional Deng Entropy*

When the sum of joint BBA distribution of A equals to 1:

$$\sum_{X} m(X,Y) = m(X)$$

Then the two-dimensional Deng entropy degenerates into Deng entropy. However, it is impossible to enumerate frame of discernment. Therefore, $\sum_{X} m(X,Y) \neq m(X)$. Thus, we regulate:

$$H(\theta) = -\sum_{x \in \theta} m(x,Y) \log \frac{m(x)}{2^{|x|}-1} = 0 \tag{17}$$

Where θ is the complement of the known set.

When the discrete random variable changes from one to two, Deng entropy's formula has changed simultaneously:

$$H(X) = -\sum_{X} m(X,Y) \log \frac{m(X)}{2^{|X|}-1} \tag{18}$$

4.2. *Conditional Deng Entropy*

Combining conditional probability and Deng entropy, we deduce conditional Deng-self entropy under conditional BBA distribution:

$$H(Y|X=x) = -\sum_{Y} m(Y|X=x) \log \frac{m(Y|X=x)}{2^{|Y|}-1} \tag{19}$$

It should be noted that the denominator of real number in logarithm is $2^{|Y|}-1$ because we need for the entropy of Y in the case of knowing X.

According to Eq.11, conditional Deng entropy is:

$$H(Y|X) = \sum_{i=1}^{N} m(X=x_i) H(Y|X=x_i) = -\sum_{X,Y} m(X,Y) \log \frac{m(Y|X)}{2^{|Y|}-1} \tag{20}$$

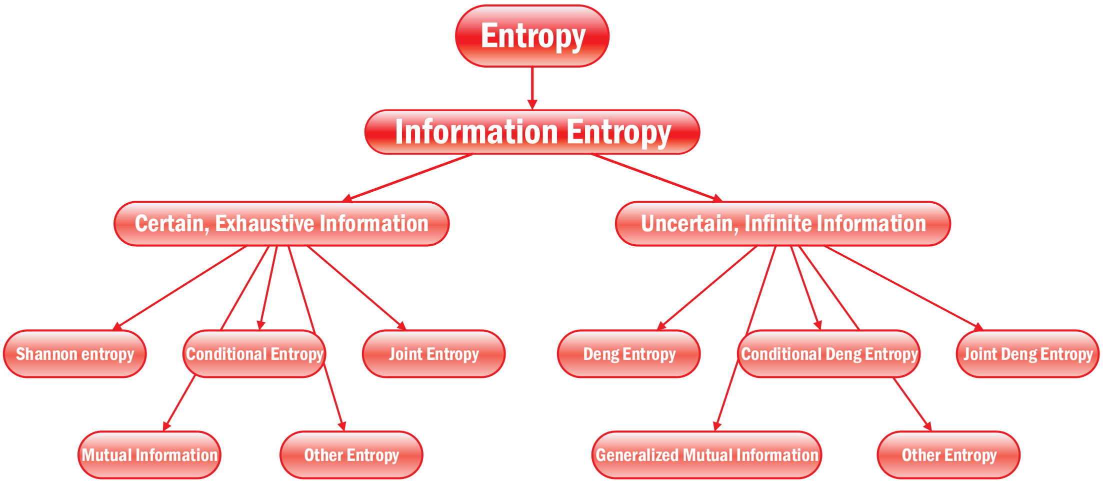

Figure 2: Information Entropy Classification

4.3. *Joint Deng Entropy*

Also, by means of the joint probability distribution of entropy, the joint entropy under two discrete random variables can be expressed as:

$$H(X,Y) = -\sum_{X,Y} m(X,Y) \log \frac{m(X,Y)}{(2^{|X|}-1)(2^{|Y|}-1)} \tag{21}$$

4.4. *Generalized Mutual Information*

The generalized mutual information under the evidence theory:

$$\begin{aligned} I(X;Y) = I(Y;X) = H(X) - H(X|Y) = H(Y) - H(Y|X) \\ = \sum_{X,Y} P(X,Y) \log \frac{P(X|Y)}{P(X)} = \sum X,Y P(X,Y) \log \frac{P(Y|X)}{P(Y)} \\ = \sum_{X,Y} P(X,Y) \log \frac{P(X,Y)}{P(X)P(Y)} \end{aligned} \tag{22}$$

Similar to Deng entropy, if the basic belief assignment does not contain uncertain element, conditional Deng entropy, joint Deng entropy and generalized mutual information can degenerate to conditional entropy, joint entropy and mutual information. Besides, some properties suitable for traditional entropy are also applied in multiple new entropy.

Table 1: Mass Function of Y

y_1	y_2	y_3	y_1,y_2	y_1,y_3	total
10%	15%	20%	25%	20%	90%

Table 2: Entropy

y_1	y_2	y_3	y_1,y_2	y_1,y_3	total
0.050	0.075	0.100	0.125	0.100	0.450

Until now, we can carry out the classification of the information entropy, as is shown in Fig.2

5. Numerical Examples

5.1. Abstract Instance

Example 5.1. *Given two frames of discernment $X = (x_1, x_2, x_3)$, $Y = (y_1, y_2, y_3)$.*

Suppose $m(x_1) = 25\%$. Mass function of Y are shown in Tab.1. According to Eq.17, the entropy of H(X,Y) is shown in Tab.2.

However, if we apply Eq.16, we found that $H(x_1) = -\sum_i m(x_1) \log \frac{m(x_1)}{2^{|x_1|}-1}$=0.500. This is because the frame of discernment of Y is not exhaustive. For the unknown sets, we can not measure the entropy of them exactly. Thus, according to Eq.17, the entropy $H(x_1)$ equals 0.450.

Example 5.2. *Suppose two frames of discernment $X = (x_1, x_2, x_3)$, $Y = (y_1, y_2, y_3)$. Mass function of X,Y is shown in Tab.3.*

Then the two-dimensional Deng entropy is shown in Tab.4: Since $H(x_1) = 0.450$, $H(x_2) = 0.474$. The total Deng entropy of X is $H(X) = X(x_1) + X(x_2) = 0.924$.

Table 3: Mass Function of X,Y

	BBA	y_1	y_2	y_3	y_1, y_2	y_1, y_3	total
x_1	25%	10%	15%	20%	25%	20%	90%
x_2	55%	10%	15%	20%	25%	30%	100%

Table 4: Two-dimensional Deng entropy

	y_1	y_2	y_3	y_1, y_2	y_1, y_3	total
x_1	0.050	0.075%	0.100	0.125	0.100	0.450
x_2	0.047	0.071%	0.095	0.119	0.142	0.474

5.2. Application

Power transformer is an important equipment of the power system, which is important to establish an effective diagnosis system to ensure the safe and reliable operation of the transformer. However, in the transformer fault diagnosis, the information people get is often incomplete and uncertain, even some redundant information. Therefore, how to remove the redundant information became the focus of one of the research hotpot.

Suppose in one of the power transformers, the collection of fault type is expressed as M. In set M, known possible errors are:

$\{m_1$=Multi points grounding or short circuit; m_2=Insulation aging; m_3=Insulation damage of turn; m_4=Suspended discharge$\}$

Fault number is regarded as D=$\{d_1, d_2, d_3, d_4\}$.

Example 5.3. *Suppose now we know the mass function of fault number D and the conditional BBA distribution, as is shown below:*

	d_1	d_2	d_3	d_4	d_1, d_2	d_1, d_3	d_1, d_2, d_4	Θ	*total*
m_1	34.52%	0.00%	7.23%	0.00%	0.00%	15.02%	0.00%	10.09%	66.86%
m_2	0.00%	26.54%	0.00%	0.00%	10.25%	0.00%	1.57%	9.46%	47.82%
m_3	17.85%	0.00%	15.43%	5.56%	10.43%	20.56%	0.00%	0.00%	69.83%
m_4	15.60%	0.00%	16.79%	0.00%	26.53%	0.00%	20.96%	15.98%	95.86%
m_1, m_3	8.76%	0.00%	16.79%	0.00%	26.53%	0.00%	20.96%	15.98%	52.02%
m_1, m_4	0.00%	2.85%	4.66	0.00%	0.00%	35.79%	0.00%	5.64%	48.94%
m_2, m_4	11.20%	15.84%	0.00%	0.00%	30.42%	3.44%	13.59%	0.00%	74.49%
Θ	12.07%	0.00%	0.00%	20.64%	14.65%	0.00%	0.00%	4.68%	52.04%
BBA	22.71%	5.28%	6.58%	10.25%	18.76%	10.25%	12.18%	13.99%	100.00%

Then we get the Joint BBA distribution:

	d_1	d_2	d_3	d_4	d_1, d_2	d_1, d_3	d_1, d_2, d_4	Θ	*total*
m_1	7.84%	0.00%	0.48%	0.00%	0.00%	1.54%	0.00%	1.41%	11.27%
m_2	0.00%	1.40%	0.00%	0.00%	1.92%	0.00%	0.19%	1.32%	4.84%
m_3	4.05%	0.00%	1.02%	0.57%	1.96%	2.11%	0.00%	0.00%	9.70%
m_4	3.54%	0.00%	1.10%	0.00%	4.98%	0.00%	2.55%	2.24%	14.41%
m_1, m_3	1.99%	0.00%	0.00%	0.59%	0.00%	1.55%	1.52%	1.39%	7.04%
m_1, m_4	0.00%	0.15%	0.31%	0.00%	0.00%	3.67%	0.00%	0.79%	4.91%
m_2, m_4	2.54%	0.84%	0.00%	0.00%	5.71%	0.35%	1.66%	0.00%	11.09%
Θ	2.74%	0.00%	0.00%	2.12%	2.75%	0.00%	0.00%	0.65%	8.26%
total	22.71%	2.39%	2.90%	3.27%	17.31%	9.22%	5.92%	7.81%	71.53%

Applying Eq.11, we attain the conditional Deng entropy of each fault number and rank them in ascending order:

m_2	m_1, m_4	m_1	m_3	m_4	m_1, m_3	m_2, m_4	Θ
0.146	*0.186*	*0.227*	*0.264*	*0.335*	*0.340*	*0.441*	*0.469*

In order to reduce information to facilitate decision making, we could choose set, whose entropy is small, to do subsequent calculation.

6. Conclusion

Information entropy is one of the effective ways to measure the amount of information and thus it is used for a wide range of areas, such as power system fault analysis [17], environmental risk assessment [18], financial decision-making processes [19], etc. However, how to describe the amount of information under uncertain information has been a big issue for a long time. Deng entropy proposed an effective method to measure the amount of information under uncertain and limitless. Thus, this paper inherits Deng entropy, combining conditional Deng entropy, joint Deng entropy and generalized mutual information. These new methods could measure the information content obtained from more than one discrete random variables. Though this work, we can measure the amount of information in different associative conditions. These methods are helpful to BBA's calculation of different situations in the network.

Acknowledgment

The authors greatly appreciate the reviews' suggestions and the editor's encouragement. The work is partially supported by National High Technology Research and Development Program of China (863 Program) (Grant No. 2013AA013801), National Natural Science Foundation of China (Grant Nos. 61174022,61573290,61503237), China State Key Laboratory of Virtual Reality Technology and Systems, Beihang University (Grant No.BUAA-VR-14KF-02).

Reference

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

[2] R. R. Yager, On the dempster-shafer framework and new combination rules, Information sciences 41 (2) (1987) 93–137.

[3] C. K. Murphy, Combining belief functions when evidence conflicts, Decision support systems 29 (1) (2000) 1–9.

[4] D. Yong, S. WenKang, Z. ZhenFu, L. Qi, Combining belief functions based on distance of evidence, Decision support systems 38 (3) (2004) 489–493.

[5] Y. Deng, Generalized evidence theory, Artificial intelligence.

[6] W. Jiang, Y. Luo, X.-Y. Qin, J. Zhan, An improved method to rank generalized fuzzy numbers with different left heights and right heights, Journal of Intelligent & Fuzzy Systems 28 (5) (2015) 2343–2355.

[7] W. Jiang, Y. Yang, Y. Luo, X. Qin, Determining basic probability assignment based on the improved similarity measures of generalized fuzzy numbers, International Journal of Computers Communications & Control 10 (3) (2015) 333–347.

[8] 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.

[9] C. E. Shannon, A note on the concept of entropy, Bell System Tech. J 27 (1948) 379–423.

[10] A. Papoulis, S. U. Pillai, Probability, random variables, and stochastic processes, Tata McGraw-Hill Education, 2002.

[11] A. Rosenberg, J. Hirschberg, V-measure: A conditional entropy-based external cluster evaluation measure., in: EMNLP-CoNLL, Vol. 7, 2007, pp. 410–420.

[12] J. Nuyts, The use of mutual information and joint entropy for anatomical priors in emission tomography, in: Nuclear Science Symposium Conference Record, 2007. NSS'07. IEEE, Vol. 6, IEEE, 2007, pp. 4149–4154.

[13] R. M. Fano, D. Hawkins, Transmission of information: A statistical theory of communications, American Journal of Physics 29 (11) (1961) 793–794.

[14] K. W. Church, P. Hanks, Word association norms, mutual information, and lexicography, Computational linguistics 16 (1) (1990) 22–29.

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

[16] A. Ayoun, P. Smets, Data association in multi-target detection using the transferable belief model, International Journal of Intelligent Systems 16 (10) (2001) 1167–1182.

[17] Z. He, Y. Cai, Q. Qian, A study of wavelet entropy theory and its application in power system, in: Intelligent Mechatronics and Automation, 2004. Proceedings. 2004 International Conference on, IEEE, 2004, pp. 847–851.

[18] H. Dehghan, H. Ghassemian, Measurement of uncertainty by the entropy: application to the classification of mss data, International journal of remote sensing 27 (18) (2006) 4005–4014.

[19] C. Tang, A. Leung, K. Lam, Entropy application to improve construction finance decisions, Journal of construction engineering and management 132 (10) (2006) 1099–1113.