Information Management by Data Quantification with FuzzyEntropy and Similarity Measure

Chua Hong Siang[*], Sanghyuk Lee[**]

[*]School of Engineering, Computing, and Science Swinburne University of Technology
[**]Dept. of Electrical and Electronic Engineering, Xian Jiaotong-Liverpool University

Abstract   Data management with fuzzy entropy and similarity measure were discussed and verified by applying reliable data selection problem. Calculation of certainty or uncertainty for data, fuzzy entropy and similarity measure are designed and proved. Proposed fuzzy entropy and similarity are considered as dissimilarity measure and similarity measure, and the relation between two measures are explained through graphical illustration.Obtained measures are useful to the application of decision theory and mutual information analysis problem. Extension of data quantification results based on the proposed measures are applicable to the decision making and fuzzy game theory.

• **Key Words** : certainty; uncertainty; data management; fuzzy entropy; similarity measure

1. Introduction

Data quantification is the one of interesting research theme, in which data vagueness is represented by the real value. Studies on quantifying the degree of uncertainty has been debated between fuzzy set theory and probability [1], however coexistence seemed obvious due to two approaches are complementary rather than competitive. With the obtained research result can give the advantage for dealing with system management including reliable data selection, pattern recognition or even fuzzy game theoretic problem. Design of fuzzy entropy for calculation of uncertainty has been studied by numerous researchers [2-4]. Most of results were concentrated in the designing of fuzzy entropies [2,3], and some parts of them also showed the implicit results of fuzzy entropies [2]. Hence, to apply real data explicit fuzzy entropy has to be needed. In our previous results, fuzzy entropies based on the distance measure has been reported[5,6]. With those designed fuzzy entropies reliable data selection problem has been

solved [7]. Counter meaning of fuzzy entropy with respect to fixed data has been considered as the similarity measure and in our previous results [5]. Relation between fuzzy entropy and similarity measure has also studied [7]. In result [5], counter meaning of similarity measure was defined by dissimilarity measure, in which dissimilarity measure was derived through similarity and vice versa. Those relations give us the result that two measures can be obtained through counter measure designing. Obtained similarity measures were also designed with the distance measure, especially well-known Hamming distance measure. Hence, these data analysis make possible to manage the system optimization or design the efficient system management. Fuzzy entropy and similarity measure are introduced to describe the uncertainty and certainty of data, hence data analysis or quantification to the decision theory and fuzzy game theory has been followed. In next chapter, fuzzy entropy and similarity results are introduced and discussed. With application

example data quantification results from fuzzy entropy and similarity are verified. Applications to decision theory and fuzzy game theory are shown in Chapter 3. Finally, conclusions are followed in Chapter 4.

2. Fuzzy Entropy and Similarity Measure

Liu's definition of fuzzy entropy is illustrated in the Definition 2.1, which illustrates the four properties of fuzzy entropy definition [2].

Definition 2.1 For $\forall A \in F(X)$ and $\forall D \in P(X)$, fuzzy entropy has following four properties

(E1) $e(D) = 0, \forall D \in P(X)$

(E2) $e([1/2]_X) = max_{A \in F(X)} e(A)$

(E3) $e(A^*) \le e(A)$, for any sharpening A^* of A

(E4) $e(A) = e(A^C), \forall A \in F(X)$

where $[1/2]_X$ is the fuzzy set in which the value of the membership function is $1/2$. $F(X)$ is fuzzy set and $P(X)$ is ordinary set. Next, similarity measure between two sets is defined in Definition 2.2 [2]. On the contrary the properties of Definition 2.1 similarity measure shows that the degree of closeness between two sets containing fuzzy sets or ordinary sets.

Definition 2.2 For $\forall A, B \in F(X)$ and $\forall D \in P(X)$, similarity measure has following four properties

(S1) $s(A, B) = s(B, A)$, $\forall A, B \in F(X)$

(S2) $s(D, D^c) = 0$, $\forall D \in P(X)$

(S3) $s(C, C) = \max_{A, B \in F} s(A, B)$, $\forall C \in F(X)$

(S4) $\forall A, B, C \in F(X)$, if $A \subset B \subset C$, then $s(A, B) \ge s(A, C)$ and $s(B, C) \ge s(A, C)$,

$F(X)$ and $P(X)$ denote fuzzy set and ordinary set, respectively.

2.1 Illustrations of Fuzzy Entropy and Similarity measure

Entropy of fuzzy data set with respect to the ordinary set can be designed using distance measure. Our previous results are followed as follows [5,6]:

$$e(A, A_{near}) = d(A \cap A_{near}, [1]_X) + d(A \cup A_{near}, [0]_X) - 1$$
$$e(A, A_{near}) = 1 - d(A \cap A_{near}, [0]_X) - d(A \cup A_{near}, [1]_X)$$

$A \cap B$ and $A \cup B$ are expressed the minimum and maximum value, expressions are commonly used in fuzzy set theory. The distance is defined by

$$d(A \cap B) = \frac{1}{n} \sum_{i=1}^{n} | \mu_A(x_i) - \mu_B(x_i) |. \qquad A_{near}$$

represents the crisp set "near" to the fuzzy set A. A_{near} can be utilized by various variable as $0 \le near \le 1$. For example, the value of crisp set $A_{0.5}$ has one when $\mu_A(x) \ge 0.5$, and is zero otherwise. Above fuzzy entropies are represent the degree of uncertainty between fuzzy set and corresponding deterministic ordinary set A_{near}.

Next, similarity measures between two data sets are also followed [5,6].

$$s(A, B) = d(A \cap B, [0]_X) + d(A \cup B, [1]_X)$$
$$s(A, B) = 2 - d(A \cap B, [1]_X) - d(A \cup B, [0]_X)$$

Equations of fuzzy entropy and similarity can be also explained by graphical point of view. Fuzzy entropy means the degree of uncertainty or the dissimilarity between two data sets, fuzzy set and corresponding ordinary set generally. Hence, it can be designed through many ways satisfying Definition 2.1. Similarity measure represents the degree of similarity

between all kinds of data sets. Fuzzy entropy and similarity can be explained by graphical illustration in Fig. 1. Shaded area representsthe common information of two fuzzy sets with membership functions. Hence, regions C and Dsatisfy the definition of similarity measure. Except region of C and D satisfy the dissimilarity between two data sets.

The relation between similarity measure and dissimilarity measure can be derived as follows

$$D(A,B) + s(A,B) = 1 \qquad (1)$$

By the comparison with (1) and Fig. 1 it is clear that $s(A,B)$ is represented by graphical summation of C and D.
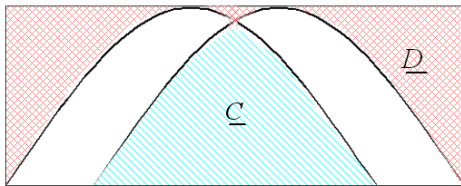

[Fig. 1] Gaussian type two membership functions

In which the total information of two fuzzy set membership functions are represented by the summation of results similarity and dissimilarity measure.

2.2 Fuzzy Entropy and Similarity Measure Application

Calculation of uncertainty and certainty for datacan be applied to the various fields such as data classification, pattern recognition. Next examples show the reliable data selection problem. Reliable data selection problem can be solved using fuzzy entropy and similarity measure [7]. In Fig. 2, Gaussian distribution is considered as the fuzzy membership function, and the chosen 5 student scores are also shown in. 5 students' scores are chosen randomly. In Fig. 2(a), 5 students have 50, 52, 55, 57, and 59 points.

Whereas, 12, 46, 53, 55, and 91 points are illustrated in Fig. 2(b).Two figures seemed clear to identify which one represents middle level or average level students by heuristic approach. However two data sets show unclear by calculation of fuzzy entropy even more numerical calculation of each average.
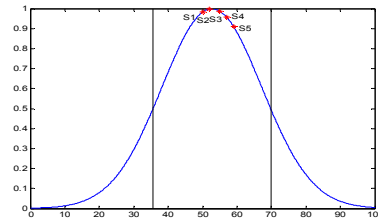
(a) Five students selection
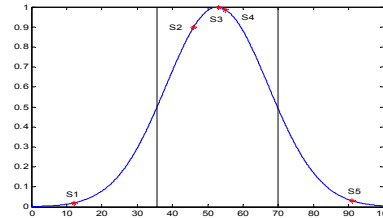
(b) Five students selection

[Fig. 2] Membership functions of 5 students

Consider fuzzy entropy as follows:

$$e(A, A_{near}) = 2d(A, A \cap A_{near}) + 2d(A_{near}, A \cap A_{near}) \quad (2)$$

The average level student's points are between 37 and 71, i.e. $\mu_{A_{0.5}}(x) = 1$ when $37 \leq x \leq 71$, $\mu_{A_{0.5}}(x) = 0$ otherwise. In the view of fuzzy entropy computation, both cases are calculated for the problem of how much they are in the average level.

Computation results say that

$$e(A, A_{0.5}) = 2d(A, A \cap A_{0.5}) + 2d(A_{0.5}, A \cap A_{0.5})$$

$$= \frac{2}{5}(|1 - 0.983| + |1 - 0.999|$$

$$+ |1 - 0.987| + |1 - 0.957| + |1 - 0.91|)$$

$$= 0.0656 .$$

In the above, $d(A, A \cap A_{0.5})$ has to be deleted because of distance between same points. Similarly, Fig. 2(b) shows that

$$e(A, A_{0.5}) = 2d(A, A \cap A_{0.5}) + 2d(A_{0.5}, A \cap A_{0.5})$$

$$= \frac{2}{5}|0.019 - 0| + |0.031 - 0|)$$

$$+ \frac{2}{5}|1 - 0.899| + |1 - 1| + |1 - 0.987|)$$

$$= 0.0656 .$$

Hence, the fuzzy entropy results indicate that two trials have same degree of uncertainty. Furthermore, they show good certainty because of small entropy value. However, their data points are not proper to represent middle level. The reason for the same fuzzy entropy values of two trials is originated from the property of complementary. This drawback was overcome through similarity measure [7].

[Table 2] Sample, fuzzy entropy, and similarity measure

	Data Information		
	Sample	Fuzzy entropy	Similarity measure
Fig. 2(a)	50, 52, 55, 57, 59	0.0656	0.9832
Fig.2(b)	12, 46, 53, 55, 91	0.0656	0.5872

With the results, similarities are calculated with designed similarity measure by 0.9832 and 0.5872, respectively. The first trial has the higher similarity value than Fig. 2(b), hence it can be determined that the result is the nearest average level 5 students with only similarity measure. From this decision, with only similarity measure provides which trial is the most reliable data selection for this problem. To obtain same result fuzzy entropy calculation is needed more statisticalinformation. Whereas compared to those results of fuzzy entropy, similarity measure has explicit advantage for reliable data selecting.

3. Applications of Fuzzy Entropy and Similarity Measure to Management Problems

With fuzzy entropy and similarity measure, accessibility to the decision theory, system modeling or system management are introduced.

3.1 Decision Theory

For decision making, building partial consequence and objective compatibility have been designed through fuzzy set theory [11]. In order to design necessity and possibility of decision it is necessary to formulate objective and consequence as fuzzy membership function.

Compatibility level is composed with necessity and possibility as following formulation:

$$k_{i,j} = (1 - \alpha)\Pi(\mu_j, \pi_{ij}) + \alpha N(\mu_j, \pi_{ij}) , \qquad (4)$$

where $\Pi(\mu_j, \pi_{ij})$ and $N(\mu_j, \pi_{ij})$ are denoted as possibility and necessity of decision. Furthermore, μ_j and π_{ij} are objective and consequence for considering fact, respectively. Considering fuzzy membership functions μ_j and π_{ij} areneeded to be small entropy, because low entropy value guarantee more certain to the fact. Furthermore possibility is greater than the necessity if the similarity between objective and consequence membership functions become greater. In example of [11], the fuzzy objective $\mu_j(x)$ corresponds to

$$\mu_j(x) = \frac{450 - x}{75}, \text{ if } 375 \le x \le 450,$$

$$\mu_j(x) = 1, \text{ if } 0 \le x \le 375,$$

$$\mu_j(x) = 0, \text{ if } x \ge 450.$$

Consequence functions satisfies

$$\pi_{ij}(x) = \frac{x - 375}{25}, \text{ if } 375 \leq x \leq 400,$$

$$\pi_{ij}(x) = \frac{425 - x}{25}, \text{ if } 400 \leq x \leq 425,$$

$$\pi_{ij}(x) = 0, \text{ if } x \leq 375 \text{ and } x \geq 425.$$

Similarity measure between μ_j and π_{ij} has the following structure.

$$s(\mu_j, \pi_{ij}) = d(\mu_j \cap \pi_{ij}, [0]_X) + d(\mu_j \cup \pi_{ij}, [1]_X) \quad (5)$$

It is clear that similarity measure value is proportional to the $\Pi(\mu_j, \pi_{ij})$ and $N(\mu_j, \pi_{ij})$ by the graphical presentation of pairs μ_j and π_{ij}. Therefore similarity modification is also applicable to the decision theory.

3.2 Characteristics of Relative Information Measure

Definition of relative information has not been formulated by researchers. In [12], they just proposed fuzzy relative information measure $R[A, B]$ as the fuzzy relative information measure of B to A. Hence, definition of fuzzy relative information measure will be presented through analyzing the definition of $R[A, B]$.

Proposition 3.1 Fuzzy relative information measure $R[A, B]$ satisfies following properties:

(i) $R[A, B] = 0$ if and only if there is no intersection between A and B, or A, B are ordinary sets.

(ii) $R[A, B] = R[B, A]$ if and only if $H(A) = H(B)$.

(iii) $R[A, B]$ takes maximum value and

$R[A, B] \geq R[B, A]$ if and only if A is contained in B, i.e, $\mu_A(x) \leq \mu_B(x)$ for $\forall x \in X$.

(iv) If $A \subset B \subset C$, then $R(B, A) \geq R(C, A)$ and $R(A, B) = R(A, C) = R(B, C)$.

Liu insisted that entropy can be calculated from the similarity measure and dissimilarity measure, which is denoted by $s + d = 1$ [2]. With this concept relative information measure can be designed via similarity measure. By the definition of entropy for certain fact, $H(A \cap B)$ and $H(A)$ satisfy $H((A \cap B), (A \cap B)_{near})$ and $H(A, A_{near})$, respectively. Where, $(A \cap B)_{near}$ satisfies the same definition of A_{near}. Roughly, it can be satisfied that

$$R[A, B] = \frac{1 - s((A \cap B), (A \cap B)_{near})}{1 - s(A, A_{near})} \quad (6)$$

Where, $s((A \cap B), (A \cap B)_{near}) = 1 - H((A \cap B), (A \cap B)_{near})$ and $s(A, A_{near}) = 1 - H(A, A_{near})$.

This measure also satisfies Proposition 3.1. Next, another relative information measure satisfying Proposition 3.1 without virtual ordinary sets $(A \cap B)_{near}$ and A_{near} is considered.

3.3 Fuzzy Coaliation in Game Theory

Coalition vectors $\alpha \in [0,1]^N$ are chosen inbetween zero and one, where N is a set of players. Each fuzzy coalition is identified with a point in the hypercube $\alpha \in [0,1]^N$, while an ordinary coalition is regarded as a vertex of this hypercube, a point in $\alpha \in [0,1]^N$. Hence, optimal choice of fuzzy coalition vector to minimized payoff function is needed. Whereas

opponents also try to make minimized other side payoff function [13].

$$\min_u U(x_i, s_j, u) = \min_u U(x_i, s_j)$$

$$\min_d V(x_i, s_j, d) = \min_d V(x_i, s_j)$$

Where, $i \in N$ and $j \in M$ are number of players and strategies, respectively. Furthermore, $u = f(x, s)$ and $d = g(x, s)$ are inputs to minimize payoff functions. In order to determine input variable player participation degree is determined by adjusting coalition vector. Problem can be transformed to determine is to determine α_i, which constitutes

$$u = \sum_{i=1, j=1}^{N, M} \alpha_i x_i s_j$$

, and it minimize $U(x_i, s_j)$. Here, $\alpha_i x_i = x(a_i)$ are considered as the fuzzy set with membership values. Also strategies are considered as the ordinary set elements. Then, it is possible to calculate the similarity measure between μ_x and fixed values. It was also verified that the calculation of similarity measure between fuzzy set and single datum [9]. Hence, similarity measure is applicable to determine the coalition vector of fuzzy game theory.

4. Conclusions

Data groups havecorrelation between the degree of similarity and dissimilarity. These relations are expressed by fuzzy entropy and similarity measure. First, fuzzy entropy and similarity are introduced, and discussed their meaning and application. Two measures are applied to the reliable data selection problem. Fuzzy set analysis can be also applied to decision theory or system management problem, especially in fuzzy game theory. For decision making considered objective and consequence are needed. Decision tools, necessity and possibility, are proportional to the similarity measure between objective and consequence membership function. Hence, the conventional decision procedure, designing compatibility level, can be replaced with similarity measure.

References

[1] H. T Nguyen, "Fuzzy sets and probability", *Fuzzy Sets and Systems*, vol. 90, pp. 129–132, 1997. doi:10.1016/S0165- 0114(97)00078-X.

[2] L. Xuecheng, "Entropy, distance measure and similarity measure of fuzzy sets and their relations", *Fuzzy Sets and Systems*, vol. 52, pp. 305–318, 1992. doi: 10.1016/0165-0114(92)90239-Z

[3] D. Bhandari and N.R. Pal, "Some new information measure of fuzzy sets", *Inform. Sci.* vol. 67, pp. 209 –228, 1993. doi: 10.1016/0020-0255(93)90073-U.

[4] Ghosh, "Use of fuzziness measure in layered networks for object extraction: a generalization", *Fuzzy Sets and Systems*, vol. 72, pp. 331‑348, 1995. doi: 10.1016/0165-0114(94)00291-E.

[5] S.H. Lee, W. Pedrycz, and Gyoyong Sohn, "Design of Similarity and Dissimilarity Measures for Fuzzy Sets on the Basis of Distance Measure", *International Journal of Fuzzy Systems,* vol. 11, pp. 67–72, 2009.

[6] S.H. Lee, K.H.Ryu, G.Y. Sohn, "Study on Entropy and Similarity Measure for Fuzzy Set", *IEICE Trans. Inf. & Syst.*, vol. E92-D, Sep. pp. 1783–1786, 2009.

[7] S.H. Lee, Y.T. Kim, S.P. Cheon, and S.S. Kim, "Reliable data selection with fuzzy entropy", *Lecture Notes in Artificial Intelligence*, vol. 3613, pp. 203‑212, 2005. doi:10.1007/11539506.

[8] S.H. Lee, S. J. Kim, N. Y. Jang, "Design of Fuzzy Entropy for Non Convex Membership Function", *Communications in Computer and Information Science*, vol. 15, pp. 55‑60, 2008. doi:10.1007/ 978-3-540-85930-7.

[9] S. H. Lee, H. J. Park, W. J. Park, "Similarity computation between fuzzy set and crisp set with similarity measure based on distance", *Lecture*

Notes in Computer Science, vol. 4993, pp. 644‐649, 2008. doi:10.1007/978-3-540-68636-1_77.

[10] S.J. Chen and S.M. Chen, ″Fuzzy risk analysis based on similarity measures of generalized fuzzy numbers,″ *IEEE Trans. on Fuzzy Systems*, vol. 11, no. 1, pp. 45-56, 2003. doi:10.1109/TFUZZ.2002. 806316.

[11] M. Roubens, ″Fuzzy sets and decision analysis″, *Fuzzy Sets and Systems*, vol. 90, pp. 199‐206, 1997. doi:10.1016/S0165-0114(97)00087-0.

[12] S.F. Ding, S,H. Xia, F.X. Jin, Z.Z. Shis, ″Novel fuzzy information proximity measures″, *Journal of Information Science*, vol. 33, no. 6, pp. 678-685, 2007. doi: 10.1177/0165551507076332.

[13] Y.A. Hwang, ″Fuzzy games: A characterization of the core″, *Fuzzy Sets and Systems*, vol. 158, pp. 2480-2493, 2007. doi:10.1016/j.fss.2007.03.009.

저자소개

HongSiang Chua

· 2002. June : Uni. of Manchester Institute Science and Tech., UK (Bachelor of Eng. (Hons))
· 2006. Jan. : The Uni. of Manchester, UK (PhD)
· 2007. Jan.~ Present: Lecturer of Faculty of Engineering, Computing, and Science, Swinburne Uni. of Tech. Sarawak Campus
· E-Mail : hschua@swinburne.edu.my
<Research Interest> : Sensor network security, Signal Processing

Sanghyuk Lee

· 1988. Feb. : Chungbuk National University, Korea, Electrical Engineering, (B. Eng.)
· 1991. Feb: Seoul National University, Korea, Electrical Engineering, (M.S.)
· 1998. Feb: Seoul National University, Korea, Electrical Engineering, (Ph.D.)
· 2011. Aug.~Present: Professor of Dept. of Electrical Engineering, XI'anJiaotong-Liverpool University, China
· E-Mail : Sangyuk.Lee@xjtlu.edu.cn