Studies on Fuzzy Information Measures

Conference Paper · September 2007

DOI: 10.1109/FSKD.2007.534 · Source: IEEE Xplore

Shifei Ding
China University of Mining and Technology

Zhongzhi Shi
Chinese Academy of Sciences

Studies on Fuzzy Information Measures

Shifei Ding[1,2], Zhongzhi Shi[2], Shixiong Xia[1], Fengxiang Jin[3]

[1]College of Computer Science and Technology, China University of Mining and Technology, Xuzhou 221008, China
[2]Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080, China
[3]College of Geoinformation Science and Engineering, Shandong University of Science and Technology, Qingdao 266510, China
dingsf@cumt.edu.cn or dingshifei@sina.com

Abstract

Fuzzy information measures play an important part in measuring the similarity degree between two pattern vectors in fuzzy circumstance. In this paper, two new fuzzy information measures are set up. Firstly, the classical similarity measures, such as dissimilarity measure (DM) and similarity measure (SM) are studied, an axiom theory about fuzzy entropy is surveyed, and all kinds of definitions of fuzzy entropy are discussed. Secondly, based on the idea of Shannon information entropy, two concepts of fuzzy joint entropy and fuzzy conditional entropy are proposed and the basic properties of them are given and proved. At last, two new measures, fuzzy absolute information measure (FAIM) and fuzzy relative information measure (FRIM), are set up, which can be used to measure the similarity degree between a fuzzy set A and a fuzzy set B. So, It provides a new research approach for studies on pattern similarity measure.

Keywords: similarity measure; dissimilarity measure; fuzzy entropy; fuzzy absolute information measure; fuzzy relative information measure.

1. INTRODUCTION

Pattern recognition is the scientific discipline whose goal is the classification of objects into a number of categories or classes. It is divided into two types called supervised and unsupervised pattern recognition. In unsupervised pattern recognition, our major concern now becomes to reveal the organization of patterns into sensible clusters (groups), which will allow us to discover similarities and differences among patterns and to derive useful conclusions about them. This idea is met in many fields, such as life sciences (biology, zoology), medical sciences (psychiatry, pathlogy), social sciences (sociology, archeology) earth sciences (geography, geology), and engineering. A major issue in unsupervised pattern recognition is that of defining the similarity between two pattern vectors and choosing an appropriate measure for it. Another issue of importance of various types of ground cover and cluster (group) the vectors on the basis of the adopted similarity measure. Generally speaking, there are two similarity measures, that is dissimilarity measure (DM) and similarity measure (SM). They play an important part in the pattern recognition and classification, the picture processing and unsupervised learning theories, but essentially they all make use of the DM or SM between the two pattern vectors under the certain environment. In practice, the people usually meet the uncertain environment, for example random environment, fuzzy environment etc [1-3].

It is well known that the concept of Shannon's entropy is the central role of information theory sometimes referred as measure of uncertainty [4]. The entropy of a random variable is defined in terms of its probability distribution and can be shown to be a good measure of randomness or uncertainty. The concept of information is too broad to be captured completely by a single definition. However, for any probability distribution, we define a quantity, called the entropy, or information entropy, which has many properties that satisfy the intuitive notion of an information measure [5]. This notion is extended to define mutual information, which is a measure of the amount of information one random variable contains about another. Entropy then becomes the self-information of a random variable. Mutual information is a special case of a more general quantity called relative entropy, which is a measure of the distance between two probability distributions. On the basis of information entropy, some researchers have studied and applied it in pattern recognition, clustering analysis, feature

Fourth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2007)
0-7695-2874-0/07 $25.00 © 2007 IEEE

extraction, system simulation and produce decision-making and so on [6-8].

In this paper, we study the fuzzy information measure based on the existing information pattern measures, and arrange them as follows: In section 2, we introduce the commonly concepts of similarity measures, such as DM and SM. Then we discuss the four axioms of fuzzy entropy, according to the definition of Shannon information entropy, the concepts of fuzzy joint entropy and fuzzy conditional entropy are given, and some properties of them are given and proved in section 3. Two novel information measures, called fuzzy absolute information measure (FAIM) and fuzzy relative information measurement (FRIM), are set up, which is measured the similarity degree between fuzzy sets A and B to some extent in Section 4. In section 5, we come to a conclusion.

2. SURVEYS ON SIMILARITY MEASURES

Let X be our data set, we begin with definitions concerning measures between vectors, and often use the following measures:

A dissimilarity measure (DM) d on X is a function as follows.

$$d : X \times X \to R$$

where R is the set of real numbers, such that

$$\exists d_0 \in R : -\infty < d_0 \le d(x,y) < +\infty, \quad \forall x, y \in X \qquad (1)$$

$$d(x,x) = d_0, \forall x \in X;\ d(x,y) = d(y,x), \forall x, y \in X \qquad (2)$$

If in addition

$$d(x,y) = d_0 \Leftrightarrow x = y \quad \forall x, y \in X \qquad (3)$$

and

$$d(x,z) \le d(x,y) + d(y,z), \quad \forall x, y, z \in X \qquad (4)$$

d is called a metric DM. In equality (4) is also known as the triangular inequality. Finally, equivalence (3) indecates that the minimum possible dissimilarity level value d_0 between and two pattern vectors in X is achieved when they are identical. Sometimes we will refer to the dissimilarity level as distance, where the term is not used in its strict mathematical sense.

A similarity measure (SM) s on X is defined as

$$s : X \times X \to R$$

Such that

$$\exists s_0 \in R : -\infty < s(x,y) \le s_0 < +\infty, \quad \forall x, y \in X \qquad (5)$$

$$s(x,x) = s_0, \forall x \in X;\ s(x,y) = s(y,x), \forall x, y \in X \qquad (6)$$

If in addition

$$s(x,y) = s_0 \Leftrightarrow x = y \quad \forall x, y \in X \qquad (7)$$

and

$$s(x,y)s(y,z) \le [s(x,y) + s(y,z)]s(x,z), \forall x, y, z \in X \qquad (8)$$

s is called a metric SM.

In the sequel, we extend the preceding definitions in order to measure similarity between subsets of X. Let U be a set containing subsets of X. Usually, the DM or SM between two sets D_i and D_j are defined in terms of DM or SM between elements of D_i and D_j. Intutively speaking, the preceding definitions show that the DMs are "opposite" to SMs. For example, it is easy to show that is d is a (metric) DM, with $d(x,y) > 0, \forall x, y \in X$, then s=a/d with a>0 is a (metric) SM. Also, $d_{\max} - d$ is a (metric) SM, where $d_{\max}$ denotes the maximum value of d among all pairs of elements of X. It is also easy to show that if d is a (metric) DM on a finite set X, such that $d(x,y) > 0, \forall x, y \in X$, then so are $-\ln(d_{\max} + k - d)$ and $kd/(1+d)$, where k is an arbitrary positive constant. On the other hand, if s is a (metric) SM with $s_0 = 1 - \varepsilon$, where ε is a small positive constant, then 1/(1-s) is also a (metric) SM. Similar comments are valid for the DM and SM between sets $D_i, D_j \in U$.

3. FUZZY ENTROPY

Fuzzy entropy is very important for measuring fuzzy information, it is the basic function of the fuzzy information processing, and it is used to measure the fuzzy degree between the two fuzzy sets of A and B. Generally speaking, fuzzy entropy, i.e. information entropy about fuzzy sets, is the following mapping:

$$\begin{aligned} H :\ & \xi(X) \to R^{+} \\ & A \mapsto H(A) \end{aligned} \qquad (9)$$

Where $\xi(X)$ is a set consisted of all fuzzy subset of discrete universe of discourse X, and $A \in \xi(X)$.

Usually, H satisfies the four axioms of De Luca and Termini, that is:

(a) $H(A) = 0$ if and only if $\mu_A(x) = 0$ or $1, \forall x \in X$;

(b) $H(A)$ takes the maximum value if and only if $\mu_A(x) = 1/2,\ \forall x \in X$;

(c) if $A \prec B$, then $H(A) \le H(B)$, where "$A \prec B$" means as follows

$$\begin{cases} 0 \le \mu_A(x) \le \mu_B(x) \le 1/2, & for \quad 0 \le \mu_B(x) \le 1/2 \\ 1/2 \le \mu_B(x) \le \mu_A(x) \le 1, & for \quad 1/2 \le \mu_B(x) \le 1 \end{cases}$$

(d) $H(A) = H(\overline{A})$, where $\overline{A}$ is complementary set of A.

The fuzzy entropy depicts the whole degree of fuzziness for a fuzzy set. As seen from above formula (a), we can understand easily that an ordinary set, i.e. crisp set, is not fuzzy, that is to say, fuzzy entropy equals zero. The formulae (b) and (c) show that the nearer 1 or 0 of the

Fourth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2007)
0-7695-2874-0/07 $25.00 © 2007 IEEE

membership degree of every element, the not fuzzier of X, and the nearer 0.5, the fuzzier of X. The formula (d) show that the sets of A and $\overline{A}$ have the same degree of fuzziness.

Zadeh proposed the concept of fuzzy entropy in 1968 firstly [8]. He gave the definition of information entropy of a fuzzy set with probability distribution $(p_1, p_2, \cdots, p_n)$ as follows

$$H_z(A) = -\sum_{i=1}^{n} \mu_A(x_i) p_i \log p_i \tag{10}$$

Where $\mu_A(x_i)$ being membership degree of x_i for A, and p_i being probability of x_i occurred. But it doesn't satisfy the four axioms of fuzzy entropy. Seen from the form, it is only a kind of weighted Shannon information entropy.

In 1975, Kaufman gave the following definition of information entropy of fuzzy set A with n elements [9].

$$H_k(A) = -\frac{1}{\log n} \sum_{i=1}^{n} \varphi_i \log \varphi_i \tag{11}$$

Where $\varphi_i = \mu_A(x_i) \Big/ \sum_{i=1}^{n} \mu_A(x_i) \; (i = 1, 2, \cdots, n)$.

The weakness of the formula (11) is not to absolute value $\mu_A(x_i)$, but adopt the relative value φ_i, so for the membership degree $\mu_A(x_i) = 0.2$, or 0.5, or 1 for all elements, the fuzzy entropy is equal. Obviously, it is not our expected.

Actually, an information entropy function, which can be used to measure the degree of fuzziness for a fuzzy set, is following fuzzy entropy built up on the foundation of Shannon function [10], namely

$$H_{dt}(A) = \frac{1}{n} \sum_{i=1}^{n} S(\mu_A(x_i)) \tag{12}$$

Where $S(x) = -x \log x - (1-x)\log(1-x)$ being Shannon function, i.e.

$$H_{dt}(A) = -\frac{1}{n} \sum \{\mu_A(x_i) \log \mu_A(x_i) + [1 - \mu_A(x_i)] \log[1 - \mu_A(x_i)]\}$$

(13)

It may be proved that the formula (12) satisfies above four axioms of De Luca- Termini.

Another type is called the fuzzy entropy of fuzzy exponent entropy associated with distances between fuzzy sets, and it can be seen fuzzy entropy induced by the distance measures. Kosko defined this fuzzy entropy as follows in 1986[11]:

$$H_k(A) = \frac{d(A, A^{near})}{d(A, A^{far})} \tag{14}$$

Where A^{near} and A^{far} are the nearest and the farthest crisp set from fuzzy set A in $\xi(X)$ respectively, and d is a distance measure in $\xi(X)$, such as Hamming distance, Euclidean distance, or Minkowski distance and so on. It has been proved that the formula (14) satisfies the four axioms of De Luca-Termini.

Pal and Pal made use of exponent function to define the following fuzzy entropy [12]:

$$H_p(A) = k \sum \{\mu_A(x_i) \exp[\mu_A(x_i)] + [(1 - \mu_A(x_i)] \exp[1 - \mu_A(x_i)]\} \tag{15}$$

Where k is a normalization factor. The formula (15) also satisfies the four axioms of De Luca-Termini.

Bhandari and Pal defined fuzzy entropy after educing the subset degree of fuzzy sets [13]:

$$H_{bp}(A) = \frac{D(A, A^{near})}{D(A, A^{far})} \tag{16}$$

Where D is a fuzzy deviation measure between two fuzzy sets. This definition, the formula (16), also satisfies the four axioms of De Luca-Termini. At the same time, they also defined the following fuzzy entropy:

$$H(A) = \frac{1}{k(1-\alpha)} \sum \log\{[\mu_A(x_i)]^{\alpha} + [1 - \mu_A(x_i)]^{\alpha}\} \tag{17}$$

Where $\alpha > 0, \alpha \neq 1$, and k is a normalization factor. The formula (17) satisfies the four axioms of De Luca-Termini.

4. FUZZY INFORMATION MEASURES

4.1 Fuzzy absolute information measures

In Shannon information theory, another important concept is mutual information. How much information does one random variable tell about another one? In fact, this is a central idea in Shannon information theory. Based on the ideas of Shannon information theory, and De Luca and Termini's fuzzy entropy theory, we study fuzzy mutual information content between two fuzzy sets as follows.

Definition 1 Suppose X is a discrete universe of discourse, and $A, B \in \xi(X)$, let

$$X^{+} = \{x \mid x \in X, \mu_A(x) \geq \mu_B(x)\}$$

$$X^{-} = \{x \mid x \in X, \mu_A(x) < \mu_B(x)\}$$

Then we have

(1) Fuzzy entropy:

Fourth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2007)
0-7695-2874-0/07 $25.00 © 2007 IEEE

$$H(A)=-\frac{1}{n}\sum_{x\in X}[\mu_A(x)\log\mu_A(x)+(1-\mu_A(x))\log(1-\mu_A(x))] \quad (18)$$

$$H(B)=-\frac{1}{n}\sum_{x\in X}[\mu_B(x)\log\mu_B(x)+(1-\mu_B(x))\log(1-\mu_B(x))] \quad (19)$$

(2) Fuzzy joint entropy:

$$\begin{aligned}H(A\cup B)&=-\frac{1}{n}\sum_{x\in X}\{[\mu_A(x)\vee\mu_B(x)]\log[\mu_A(x)\vee\mu_B(x)]\\&+[1-\mu_A(x)\vee\mu_B(x)]\log[1-\mu_A(x)\vee\mu_B(x)]\}\\&=-\frac{1}{n}\sum_{x\in X^+}\{\mu_A(x)\log\mu_A(x)-[1-\mu_A(x)]\log[1-\mu_A(x)]\}\\&-\frac{1}{n}\sum_{x\in X^-}\{\mu_B(x)\log\mu_B(x)-[1-\mu_B(x)]\log[1-\mu_B(x)]\}\end{aligned} \quad (20)$$

(3) Fuzzy conditional entropy:

$$\begin{aligned}H(A/B)&=-\frac{1}{n}\sum_{x\in X^+}\{\mu_A(x)\log\mu_A(x)-\mu_B(x)\log\mu_B(x)\\&+[1-\mu_A(x)]\log[1-\mu_A(x)]-[1-\mu_B(x)]\log[1-\mu_B(x)]\end{aligned} \quad (21)$$

$$\begin{aligned}H(B/A)&=-\frac{1}{n}\sum_{x\in X^-}\{\mu_B(x)\log\mu_B(x)-\mu_A(x)\log\mu_A(x)\\&+[1-\mu_B(x)]\log[1-\mu_B(x)]-[1-\mu_A(x)]\log[1-\mu_A(x)]\end{aligned} \quad (22)$$

It can be seen that although we use the idea of Shannon information entropy, they are completely different from the concept.

They have following properties:

Theorem 1 Suppose X is a discrete universe of discourse, and $A,B\in\xi(X)$, then

$$H(A/B)\le H(A) \quad (23)$$

$$H(B/A)\le H(B) \quad (24)$$

Proof:

$$\begin{aligned}&H(A)-H(A/B)\\&=-\frac{1}{n}\sum_{x\in X}[\mu_A(x)\log\mu_A(x)+(1-\mu_A(x))\log(1-\mu_A(x))]\\&+\frac{1}{n}\sum_{x\in X^+}\{\mu_A(x)\log\mu_A(x)-\mu_B(x)\log\mu_B(x)\\&+[1-\mu_A(x)]\log[1-\mu_A(x)]-[1-\mu_B(x)]\log[1-\mu_B(x)]\\&=-\frac{1}{n}\sum_{x\in X^+}[\mu_B(x)\log\mu_B(x)+(1-\mu_B(x))\log(1-\mu_B(x))]\\&-\frac{1}{n}\sum_{x\in X^-}[\mu_A(x)\log\mu_A(x)+(1-\mu_A(x))\log(1-\mu_A(x))]\\&\ge 0\end{aligned}$$

So

$$H(A/B)\le H(A)$$

Similarly

$$H(B/A)\le H(B)$$

Based on the idea of mutual information of Shannon information theory, we can get the concept of fuzzy mutual information (FMI) as follows.

Definition 2 Suppose X is a discrete universe of discourse, and $A,B\in\xi(X)$, then the absolute difference value, $H(A)-H(A/B)$, is called the FMI between fuzzy set A and fuzzy set B, denoted by $H(A\cap B)$, that is to say

$$\begin{aligned}&H(A\cap B)=H(A)-H(A/B)\\&=-\frac{1}{n}\sum_{x\in X^+}[\mu_B(x)\log\mu_B(x)+(1-\mu_B(x))\log(1-\mu_B(x))]\\&-\frac{1}{n}\sum_{x\in X^-}[\mu_A(x)\log\mu_A(x)+(1-\mu_A(x))\log(1-\mu_A(x))]\end{aligned} \quad (25)$$

Similarly, we define $H(B\cap A)$, namely

$$\begin{aligned}&H(B\cap A)=H(B)-H(B/A)\\&=-\frac{1}{n}\sum_{x\in X^+}[\mu_B(x)\log\mu_B(x)+(1-\mu_B(x))\log(1-\mu_B(x))]\\&-\frac{1}{n}\sum_{x\in X^-}[\mu_A(x)\log\mu_A(x)+(1-\mu_A(x))\log(1-\mu_A(x))]\end{aligned} \quad (26)$$

According to the theorem 1 and definition 2, we can see easily that the FMI satisfies basic properties, such as nonnegativity, symmetry, but doesn't satisfy triangular inequality, so the FMI is a kind of generalized and absolute fuzzy entropy measure, and is called the fuzzy absolute information measure (FAIM). The FAIM measures the similarity degree between two fuzzy sets A and B to some extent.

4.2 Fuzzy relative information measures

In actually applying, we sometimes use the relative fuzzy measure. Now, we give a kind of fuzzy relative information measure as follows.

Definition 3 Suppose X is a discrete universe of discourse, and $A,B\in\xi(X)$, then the relative difference value, denoted by *R(A,B)*, is called fuzzy relative information measure for B to A, namely

$$R(A,B)=\frac{H(A\cap B)}{H(A)}=\frac{H(A)-H(A/B)}{H(A)} \quad (27)$$

Similarly

$$R(B,A)=\frac{H(B\cap A)}{H(B)}=\frac{H(B)-H(B/A)}{H(B)} \quad (28)$$

Fourth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2007)
0-7695-2874-0/07 $25.00 © 2007 IEEE

is called fuzzy relative mutual information for the fuzzy set A to fuzzy set B.

$R(A,B)$ shows an influence degree of the fuzzy set A to fuzzy set B in fuzzy information processing aspects. It is a kind of generalized and relative fuzzy entropy measure, and called the fuzzy relative information measure (FRIM). The FRIM can be used for the clustering analysis, pattern recognition and picture processing etc.

5. CONCLUSIONS

According to the four axioms of fuzzy entropy, some definitions of fuzzy entropy are summarized. On this foundation, we give the definitions of the fuzzy joint entropy and fuzzy conditional entropy, and prove a basic property. The classical similarity measures, such as the DM and the SM are studied, and then the two lately similarity measures of the FAIM and FRIM are set up. The FAIM and FRIM can be used to measure the similarity degree between two fuzzy sets A and B to some extent, and so they are a further development and perfect for the information similarity measure.

6. Acknowledgements

This work is supported by the National Natural Science Foundation of China under grant no.60435010, no.60675010, no.40574001 and no.50674086, the National Grand Fundamental Research (973) Program under grant no.2003CB317004, the National High-Tech Program (863) under grant no.2006AA01Z128 and the Doctoral Foundation of Chinese Education Ministry under grant no.20060290508. We would like to thank the referees and the editors for their good suggestions.

7. References

[1] S. Theodoridis and K. Koutroumbas. Pattern Recognition (2nd). New York: Academic Press, 2003.

[2] M.R. Anderberg. Cluster Analysis for Applications. New York: Academic Press, 1973.

[3] R.O. Duda and P.E. Hart. Pattern Classification and Scene Analysis. New York: John Wiley and Sons, 1973.

[4] C.E. Shannon. A Mathematical Theory of Commu-nication. Bell System Technique Journal, 1948, 27: 379-423, 623-656.

[5] T.M. Cover and J.A. Thomas. Elements of Information Theory. New York: John Wiley and Sons, 1991.

[6] S.F. Ding and Z.Z. Shi. Studies on Incidence Pattern Recognition Based on Information Entropy. Journal of Information Science, 2005,31(6): 497-502.

[7] S.F. Ding and F.X. Jin. Information Characteristics of Discrete K-L Transform Based on Information Entropy. Transactions Nonferrous Metals Society of China, 2003,13(3): 729-734.

[8] M. Chan. System Simulation and Maximum Entropy. Operations Research, 1971, 19:1751-1753.

[8] L.A. Zadeh. Probability Measures of Fuzzy Events. Journal of Mathematical Analysis and Applications, 1968, 23(10): 421-427.

[9] A. Kaufman. Introduction to the Theory of Fuzzy Subsets---Fundamental Theoretical Elements. New York: Academic Press, 1975.

[10] A. De Luca and S. Termini. A Definition of Nonprobabilistic Entropy in the Setting of Fuzzy Sets Theory. Information and Control, 1972, 20(3):301-312.

[11] B. Kosko. Fuzzy Entropy and Conditioning. Information Sciences, 1986, 40(2): 165-174.

[12] N.R. Pal and S.K. Pal. Higher Order Fuzzy Entropy and Hybrid Entropy of a Set. Information Sciences, 1992, 61(3): 211-231.

[13] D. Bhandari and N.R. Pal. Some New Information Measures for Fuzzy Sets. Information Sciences, 1992, 67(3): 209-228.

0-7695-2874-0/07 $25.00