International Journal of Fuzzy Logic and Intelligent Systems, vol. 5, no. 4, December 2005 pp. 367-371

Similarity Measure Construction with Fuzzy Entropy and Distance Measure

Sang-Hyuk Lee

School of Electrical and Computer Engineering, Pusan National University

Abstract

The similarity measure is derived using fuzzy entropy and distance measure. By the relations of fuzzy entropy, distance measure, and similarity measure, we first obtain the fuzzy entropy. And with both fuzzy entropy and distance measure, similarity measure is obtained. We verify that the proposed measure become the similarity measure.

Key Words : Similarity measure, fuzzy entropy, distance measure

1. Introduction

Similarity between two sets can be applied to the pattern classification field. Similarity measure has been noticed as the complementary meaning of the distance measure, i.e, $s+d=1$, where d and s are distance and similarity measure respectively. In the above, 1 means the sum of similarity and dissimilarity. In the previous literatures, fuzzy entropy of a fuzzy set represents a measure of fuzziness of the fuzzy set[1-7]. Hence, well-defined distance measure represents the fuzzy entropy. By the summing relation, we can notice that the similarity measure can be illustrated through distance measure and fuzzy entropy function. Well known-Hamming distance is usually used to construct fuzzy entropy, so we compose the fuzzy entropy function through Hamming distance measure. Using the relation of distance measure and similarity measure, we construct and prove the similarity measure with fuzzy entropy, and similarity measure is also constructed through distance measure. In the next section, the axiomatic definitions of entropy, distance measure and similarity measure of fuzzy sets are introduced and fuzzy entropy is constructed through distance measure. In Section 3, similarity measures are constructed and proved through fuzzy entropy and the distance measure. Used distance measure is proposed by considering support average. Conclusions are followed in Section 4.

Notations: Through out this paper, $R^+=[0,\infty)$, $F(X)$, and $P(X)$ represent the set of all fuzzy sets and crisp sets on the universal set X respectively. $\mu_A(x)$ is the membership function of $A\in F(X)$, and the fuzzy set A, we use A^C to express the complement of A, i.e. $\mu_{A^c}(x)=1-\mu_A(x)$, $\forall x\in X$. For fuzzy sets A and B, $A\cup B$, the union of A and B is defined as $\mu_{A\cup B}(x)=max(\mu_A(x),\mu_B(x))$, $A\cap B$, the intersection of A and B is defined as $\mu_{A\cap B}(x)=min(\mu_A(x),\mu_B(x))$. A fuzzy set A^* is called a sharpening of A, if $\mu_{A^*}(x)\geq\mu_A(x)$ when $\mu_A(x)\geq 1/2$ and $\mu_{A^*}(x)\leq\mu_A(x)$ when $\mu_A(x)<1/2$. For any crisp sets D, A_{near} and A_{far} of fuzzy set A are defined as

$$\mu_D(x)=\begin{cases}\frac{1}{2} & x\in D\\ 0 & x\notin D\end{cases},\qquad \mu_{A_{near}}(x)=\begin{cases}1 & \mu_A(x)\geqq\frac{1}{2}\\ 0 & \mu_A(x)<\frac{1}{2}\end{cases}$$

$$\mu_{A_{far}}(x)=\begin{cases}0 & \mu_A(x)\geqq\frac{1}{2}\\ 1 & \mu_A(x)<\frac{1}{2}\end{cases}.$$

2. Preliminary

In this section, we introduce some preliminary results and also discuss induced results. Liu suggested three axiomatic definitions of fuzzy entropy, distance measure and similarity measure as follows [4]. By these definitions, we can propose entropy, and compare it with the result of Liu.

2.1 Some definitions of fuzzy entropy

In this subsection, we introduce some preliminary results about fuzzy entropy, distance measure, similarity measure, and related properties.

Definition 2.1 (Liu, 1992) A real function e: $F(X)\to R^+$ or $e: P(X)\to R^+$ is called an entropy on $F(X)$ or $P(X)$ if e has the following properties:

(E1) $e(D)=0,\ \forall D\in P(X)$
(E2) $e([1/2])=max_{A\in F(X)}e(A)$
(E3) $e(A^*)\leq e(A)$, for any sharpening A^* of A

Manuscript received Jul. 4, 2005; revised Aug. 17, 2005.
This work has been supported by KESRI(R-2004-B-129), which is funded by MOCIE(Ministry of commerce, industry and energy).

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

where $[1/2]$ is the fuzzy set in which the value of the membership function is $1/2$.

Definition 2.2 [Liu, 1992] A real function d: $F^2 \rightarrow R^+$ is called a distance measure on $F(X)$ or $P(X)$ if d satisfies the following properties:

(D1) $d(A, B) = d(B, A), \forall A, B \in F(X)$

(D2) $d(A, A) = 0$, $\forall A \in F(X)$

(D3) $d(D, D^c) = max_{A,B \in F}\, d(A, B)$, $\forall D \in P(X)$

(D4) $\forall A, B, C \in F(X)$, if $A \subset B \subset C$, then $d(A, B) \leq d(A, C)$ and $d(B, C) \leq d(A, C)$.

Definition 2.3 (Liu, 1992) A real function s: $F^2 \rightarrow R^+$ or $P^2 \rightarrow R^+$ is called a similarity measure, if s has the following properties:

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

(S2) $s(A, A^c) = 0$ $\forall A \in F(X)$

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

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

Liu also pointed out that there is an one-to-one relation between all distance measures and all similarity measures, $d + s = 1$. Fuzzy normal similarity measure on F is also obtained by the division of $\max_{C,D \in F} s(C, D)$. If We divide universal set X into two parts D and D^C in $P(X)$, then the fuzziness of fuzzy set A be the sum of the fuzziness of $A \cap D$ and $A \cap D^C$. By this idea, following definition is followed.

Definition 2.4. (Fan and Xie, 1999) Let e be an entropy on $F(X)$. Then for any $A \in F(X)$,

$$e(A) = e(A \cap D) + e(A \cap D^c)$$

is σ-entropy on $F(X)$.

Definition 2.5. (Fan and Xie, 1999) Let d be a distance measure on $F(X)$. Then for any $A, B \in F(X)$, and $D \in P(X)$,

$$d(A, B) = d(A \cap D, B \cap D) + d(A \cap D^c, B \cap D^c)$$

be the σ-distance measure on $F(X)$.

Definition 2.6. (Fan and Xie) Let s be a similarity measure on $F(X)$. Then for any $A, B \in F(X)$, and $D \in P(X)$,

$s(A, B) = s(A \cap D, B \cup D^c) + s(A \cap D^c, B \cup D)$

be the σ-similarity measure on $F(X)$.

From definition 2.4-6, we can focus interesting area of universal set and extend the theory of entropy, distance measure and similarity measure of fuzzy sets. Fan and Xie derived new entropy via defined entropy, which is introduces by $e' = e/(2 - e)$, where e is an entropy on $F(X)$.

2.2 Fuzzy entropy with distance measure

In this section, we propose entropy that is induced by the distance measure. Among distance measures, Hamming distance is commonly used σ-distance measure between fuzzy sets A and B,

$$d(A, B) = \frac{1}{n}\sum_{i=1}^{n} | \mu_A(x_i) - \mu_B(x_i) |$$

where $X = \{x_1, x_2, \cdots x_n\}$, $| k |$ is the absolute value of k. Next Proposition shows that the distance relation of between fuzzy set and crisp sets.

Proposition 2.1 (Fan and Xie, 1999). Let d be a σ-distance measure on $F(X)$, then

(i) $d(A, A_{near}) \geq d(A^*, A_{near})$

(ii) $d(A, A_{far}) \leq d(A^*, A_{far})$.

Fan, Ma and Xie proposed the following theorem [7].

Theorem 2.1 (Fan, Ma, and Xie, 2001) Let d be a σ-distance measure on $F(X)$, if d satisfies

(i) $d(\frac{1}{2}D, [0]) = d(\frac{1}{2}D, D)$, $\forall D \in P(X)$,

(ii) $d(A^C, B^C) = d(A, B)$, $A, B \in F(X)$,

then $e(A) = D(A, A_{near}) + 1 - d(A, A_{far})$ is a fuzzy entropy.

Now we propose another fuzzy entropy induced by distance measure which is different from Theorem 3.1 of Fan, Ma and Xie [7]. Proposed entropy needs only A_{near} crisp set, and it has the advantage in computation of entropy.

Theorem 2.2 Let d be a σ-distance measure on $F(X)$; if d satisfies

$$d(A^C, B^C) = d(A, B), A, B \in F(X),$$

then

$$e(A) = 2d((A \cap A_{near}), [1]) + 2d((A \cup A_{near}), [0]) - 2 \tag{3}$$

is a fuzzy entropy.

Proofs of (3) are satisfied if (3) satisfy the Definition 2.1, so it is illustrated in [9]. Theorem 2.2 uses only A_{near} crisp set, hence we can consider another entropy. Which considers only A_{far}, and it has more compact form than Theorem 2.2.

Theorem 2.3 Let d be a σ-distance measure on $F(X)$; if d satisfies

$$d(A^c, B^c) = d(A, B), A, B \in F(X),$$

then

$$e(A) = 2d((A \cap A_{far}), [0]) + 2d((A \cup A_{far}), [1]) \tag{4}$$

is a fuzzy entropy.

In a similar way we can prove from (E1) to (E4) of

Definition 2.1, it is also found in [9].

Proposed entropies Theorem 2.2 and 2.3 have some advantages to the Liu's, they don't need assumption (i) of Theorem 2.1 to prove (3) and (4). Furthermore (3) and (4) use only one crisp sets A_{near} and A_{far}, respectively. Later we check the proposed entropy of Theorem 2.2 and 2.3 are the σ-entropy on $F(X)$ for any $A \in F(X)$, satisfying $e(A) = e(A \cap D) + e(A \cap D^C)$. Next, we apply Theorem 2.2 and 2.3 to detect reliable phase current among the 3-phases faulted induction motor. Proposed fuzzy entropies are more succinct than those of Fan *et.al*'s, and they need easier assumptions than previous results.

3. Derivation of Similarity Measure

We obtain the fuzzy entropy with the distance measure in previous section. Generally, fuzzy entropy is expressed through distance measure, i.e., $e(A) = e(d(A))$. In our result, entropy is represented distance measure itself, $e(A) = d(A)$. Hence, by the result of Liu's,

$$d(A) + s(A) = 1 \tag{6}$$

we modify the similarity measure as $s(A) = 1 - e(A)$, that means fuzzy set A matches to the crisp set A_{near} nearly as $s(A)$ approaches to 0. We illustrate the similarity measure with the entropy function in subsection 3.1 and the similarity measure construction using the distance measure in the subsection 3.2.

3.1 Similarity measure using the entropy function

We propose the similarity measure in the following theorems. Theorem 3.1 is obtained by considering Theorem 3.2.

Theorem 3.1 For fuzzy set $A \in F(X)$, if d satisfies distance measure, then

$$\begin{aligned} s(A, A_{\neq ar}) = 4 &- 2d((A \cap A_{near}), [1]) \\ &- 2d((A \cup A_{near}), [0]) \end{aligned} \tag{7}$$

is the similarity measure between fuzzy set A and crisp set A_{near}.

proof. We prove that the eq. (7) satisfies the Definition 2.3. (S1) means the commutativity of set A and A_{near}, hence it is clear from (7) itself. From (S2), $s(A, A^c) = 0$ is shown as

$$\begin{aligned} s(A, A^c) &= 4 - 2d((A \cap A^c), [1]) - 2d((A \cup A^c), [0]) \\ &= 4 - 2d([0], [1]) - 2d([1], [0]) \\ &= 4 - 2 \cdot 1 - 2 \cdot 1 = 0. \end{aligned}$$

For all $A, B \in F(X)$, inequality of (S3) is proved by

$$\begin{aligned} s(A, B) &= 4 - 2d((A \cap B), [1]) - 2d((A \cup B), [0]) \\ &\leqq 4 - 2d((D \cap D), [1]) - 2d((D \cup D), [0]) \\ &= s(D, D). \end{aligned}$$

Inequality is satisfied from
$d((A \cap B), [1]) \geqq d((D \cap D), [1])$ and
$d((A \cup B), [0]) \geqq d((D \cup D), [0])$.
Finally, (S4) is $\forall A, B, C \in F(X)$, $A \subset B \subset C$,

$$\begin{aligned} s(A, B) &= 4 - 2d((A \cap B), [1]) - 2d((A \cup B), [0]) \\ &= 4 - 2d(A, [1]) - 2d(B, [0]) \\ &\geqq 4 - 2d(A, [1]) - 2d(C, [0]) \\ &= s(A, C) \end{aligned}$$

also

$$\begin{aligned} s(B, C) &= 4 - 2d((B \cap C), [1]) - 2d((B \cup C), [0]) \\ &= 4 - 2d(B, [1]) - 2d(C, [0]) \\ &\geqq 4 - 2d(A, [1]) - 2d(C, [0]) \\ &= s(A, C) \end{aligned}$$

is satisfied. Inequality is also satisfied with $d(B, [0]) \leqq d(C, [0])$ and $d(B, [1]) \leqq d(A, [1])$.

Therefore proposed similarity measure (7) satisfy Definition 2.3. Similarly, we propose another similarity measure in the following theorem.

Theorem 3.2 For fuzzy set $A \in F(X)$ and distance measure d,

$$\begin{aligned} s(A, A_{\neq ar}) = 2 &- 2d((A \cap A^c_{near}), [0]) \\ &- 2d((A \cup A^c_{near}), [1]) \end{aligned} \tag{8}$$

is the similarity measure of fuzzy set A and crisp set A_{near}.

proof. Proofs are shown similarly as Theorem 3.1. Commutativity of (S1) is clear from (8). To show the property of (S2),

$$\begin{aligned} s(A, A^c) = 2 &- 2d((A \cap (A^c)^c), [0]) \\ &- 2d((A \cup (A^c)^c), [1]) \\ &= 2 - 2(d(A, [0]) + d(A, [1])) \\ &= 2 - 2 \cdot 1 = 0 \end{aligned}$$

is clear. (S3) is clear from the relation

$$\begin{aligned} s(A, B) &= 2 - 2d((A \cap B^c), [0]) - 2d((A \cup B^c), [1]) \\ &\leqq 2 - 2d((D \cap D^c), [0]) - 2d((D \cup D^c), [1]) \\ &= s(D, D), \end{aligned}$$

where the inequality is proved by
$d((A \cap B^c), [0]) \geqq d((D \cap D^c), [0])$ and
$d((A \cup B^c), [1]) \geqq d((D \cup D^c), [1])$.
Finally, $\forall A, B, C \in F(X)$ and $A \subset B \subset C$ imply

$$\begin{aligned} s(A, B) = 2 &- 2d((A \cap B^c), [0]) \\ &- 2d((A \cup B^c), [1]) \\ &= 2 - 2d([0], [0]) - 2d((A \cup B^c), [1]) \\ &\geqq 2 - 2d(A \cap C^c, [0]) - 2d(A \cup C^c, [1]) \end{aligned}$$

$$= 2 - 2d([0],[0]) - 2d(A \cup C^c,[1])$$
$$= s(A, C).$$

Also

$$s(B, C) = 2 - 2d((B \cap C^c),[0]) - 2d((B \cup C^c),[1])$$
$$= 2 - 2d([0],[0]) - 2d((B \cup C^c),[1])$$
$$\geq 2 - 2d(A \cap C^c,[0]) - 2d(A \cup C^c,[1])$$
$$= 2 - 2d([0],[0]) - 2d(A \cup C^c,[1])$$
$$= s(A, C)$$

is satisfied with
$d((A \cup B^c),[1]) \geq d((A \cup C^c),[1])$ and
$d((B \cup C^c),[1]) \geq d((A \cup C^c),[1])$.

We have proposed the similarity measure that are induced from fuzzy entropy. Those fuzzy entropy is also induced from distance measure. Hence to obtain more implicit result, we consider distance measure directly. In the following subsection, we suggest similarity measure which is constructed using distance measure.

3.2 Similarity measure using distance measure

We have represented the relation $s(A) = 1 - d(A)$ in (6), that means the sum of similarity measure and distance measure is constant. Hence, we solve the similarity measure directly from distance measure in this subsection. To obtain the similarity, proper distance measure is needed. First, we consider the Hamming distance $d(A, B) = \frac{1}{n}\sum_{i=1}^{n} |\mu_A(x_i) - \mu_B(x_i)|$. By the relation of (6),

$$s(A, B) = 1 - d(A, B) = 1 - \frac{1}{n}\sum_{i=1}^{n} |\mu_A(x_i) - \mu_B(x_i)| \quad (9)$$

is the similarity measure? We check the definition (S1) to (S4). (S1) is clear from the eq. (9). (S2) and (S3) are also obtained from (D3) and (D2) of definition 2.2. Finally, (S4) is proved with the help of (D4). Naturally we conclude that if the distance measure is constructed properly, then the similarity measure can be obtained. Now we consider the membership functions type1 and 2 in figure 1 and 2. In the Fig. 1 and 2, area between μ_A and μ_B are the same, value of Hamming distance of Fig. 1 and 2 are same. Then which case is more similar ?

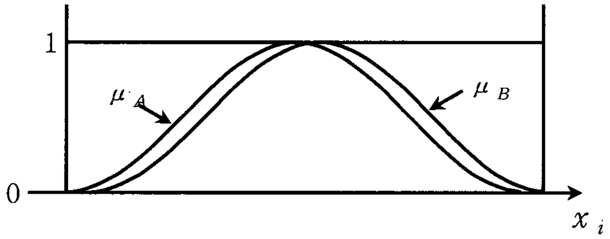


Fig.1 Membership functions type 1

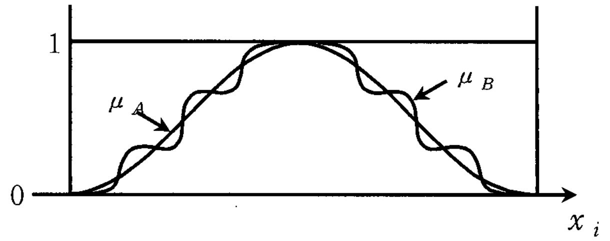


Fig. 2 Membership functions type 2

Consider the function

$$d(A, B) = \frac{1}{n}\sum_{i=1}^{n} |\mu_A(x_i) - \mu_B(x_i)| + |support_A(x_i) - support_B(x_i)| \quad (10)$$

where $support_A(x_i) = \frac{1}{n}\sum_{i=1}^{n} |x_i|$, $x_i \in A$ and $support_B(x_i) = \frac{1}{n}\sum_{i=1}^{n} |x_i|$, $x_i \in B$ are the average values of support. In this case, the first part of (10) is the Hamming distance, hence the value of Fig. 1 and Fig. 2 are the same. However the last part of (10) represents the difference of average support. Now we check whether the eq. (10) satisfy the distance measure definition or not. For (D1), commutativity is clear from (10). $d(A, A) = 0$ is also clear, and the first part of distance between D and D^c represents the sum 1. (D4) is clear because Fig. 2 has similar average value compare to Fig. 1. Eq. (10) is the distance measure, thus we can propose similarity measure as

$$s(A, B) = 1 - \frac{1}{n}\sum_{i=1}^{n} |\mu_A(x_i) - \mu_B(x_i)| - |support_A(x_i) - support_B(x_i)|. \quad (11)$$

proof of (11). Proof is similar to theorem 3.1 and 2, commutativity of (S1) is proved by (11). (S2) is also proved by $support_A(x_i) = support_B(x_i)$. Also $d(D, D) = 0$, hence $s(D, D) = 1$ is proved. (S4) is finally proved by the (D4). Proof verify that eq. (11) represent similarity measure.

4. Conclusions

We introduce the distance measure, similarity measure and fuzzy entropy, fuzzy entropy can be represented by the function of distance measure. By the one to one correspondence of distance measure and similarity measure, we construct the similarity measure using distance measure. As we noted before, fuzzy entropy is the function of distance measure. Hence similarity measure is constructed through the fuzzy entropy, and we prove. And similarity measure is also induced through distance measure. We verify that the proposed measure is the similarity measure.

References

[1] D. Bhandari, N. R. Pal, "Some new information measure of fuzzy sets," Inform. Sci. 67, 209-228, 1993.

[2] A. Ghosh, "Use of fuzziness measure in layered networks for object extraction: a generalization," Fuzzy Sets and Systems, 72, 331-348, 1995.

[3] B. Kosko, Neural Networks and Fuzzy Systems, Prentice-Hall, Englewood Cliffs, NJ, 1992.

[4] Liu Xuecheng, "Entropy, distance measure and similarity measure of fuzzy sets and their relations," Fuzzy Sets and Systems, 52, 305-318, 1992.

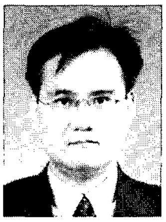
[5] N.R. Pal, S.K. Pal, "Object-background segmentation using new definitions of entropy," IEEE Proc. 36, 284-295, 1989.

[6] J. L. Fan, W. X. Xie, "Distance measure and induced fuzzy entropy," Fuzzy Set and Systems, 104, 305-314, 1999.

[7] J. L. Fan, Y. L. Ma, and W. X. Xie, "On some properties of distance measures," Fuzzy Set and Systems, 117, 355-361, 2001.

[8] S. Aksoy and R.M. Haralick, "Feature normalization and likelihood-based similarity measures for image retrieval", Pattern Recognition Letters, 22, 563-582, 2001.

[9] S.H. Lee, K.B. Kang and S.S. Kim, "Measure of fuzziness with fuzzy entropy function", Journal of Fuzzy Logic and Intelligent Systems, vol. 14, no. 5, 642-647, 2004.



Sang-Hyuk Lee

He received the B.S. degree in electrical engineering from Chungbuk National University, Cheongju, Korea, in 1988, the M.S. and Ph.D. degrees in electrical engineering from Seoul National University, Seoul, Korea, in 1991 and 1998, respectively. He also received M.S. degree in mathematics from Chungnam National University, Daejeon, Korea, in 2003. He served as a Research Fellow from 1996 to 1999 in the HOW Co. Ltd. Since 2000, he has been with Pusan National University. Currently he is a Chaired Professor in Specialized Group in Industrial Automation, Information, and Communication. His research interests include robust control theory, game theory, and fuzzy theory.