퍼지 및 지능시스템학회 논문지 2001, Vol. 11, No. 3, pp. 270-274

A similarity measure of fuzzy sets

Soon H. Kwon

School of EECS, Yeungnam Univ., 214-1 Dae-dong, Kyongsan, Kyongbuk 712-749, Korea

요 약

지금까지 제안된 유사도 척도는 첫째, 기하학적 유사도 척도, 둘째, 집합론적 유사도 척도, 그리고 마지막으로 일치 함수를 이용한 유사도 척도와 같이 세 종류로 분류될 수 있다. 본 논문에서는 이러한 기존의 유사도 척도가 갖는 여러 가지 성질에 근거하여 퍼지 집합에 관한 새로운 유사도 척도를 제안하고 이의 성질을 알아본다. 마지막으로, 예제를 통하여 제안된 유사도 척도와 기존의 유사도 척도의 특성을 비교한다.

Abstract

Conventional similarity measures suggested so far can be classified into three categories: (i) geometric similarity measures, (ii) set-theoretic similarity measures, and (iii) matching function-based similarity measures. On the basis of the characteristics of the conventional similarity measures, in this paper, we propose a new similarity measure of fuzzy sets and investigate its properties. Finally, numerical examples are provided for the comparison of characteristics of the proposed similarity measure and other previous similarity measures.

Key Words : Fuzzy sets, Fuzzy models, Rule simplification, Similarity measure.

1. Introduction

In recent years, traditional mathematical model-based techniques for modeling of complex and uncertain systems are being replaced by a linguistic approach using fuzzy sets. The basic idea of fuzzy modeling is to represent the system by a set of fuzzy if-then rules that maps fuzzy partitioned input space to outputs, that is, approximates the fuzzy partitioned system by a simple linguistic model.

In a fuzzy model developed by using expert knowledge or numerical data, the redundancy and thus complexity due to similar fuzzy sets representing compatible concepts is inevitable. This results in a computationally inefficient and linguistically intractable description of the system. Several methods (e.g., similarity measure-based simplification [6] and the statistical information criterion-based model construction [11]) have been proposed for the construction of adequate but parsimonious fuzzy models.

Since Zadeh's work [15], a lot of attentions have been paid for the development of new similarity measures and their applications. Similarity measures suggested so far can be classified into three categories [1]: (i) geometric similarity measures,(ii) set-theoretic similarity measures and (iii) matching function-based similarity measures. Properties of those have been investigated for providing some useful information to select a suitable similarity measure in applications of fuzzy sets by many authors [1,5,8]. It has been argued that geometric similarity measures representing similarity as proximity of fuzzy sets are best suited for measuring similarity among distinct fuzzy sets, while the set-theoretic similarity measures are the most suitable for capturing similarity among overlapping fuzzy sets [8].

Under the background, in this paper, we propose a new similarity measure and investigate its properties. Two numerical examples are provided for the comparison between the proposed measure and other previous similarity measures.

접수일자 : 2001년 2월 17일
완료일자 : 2001년 5월 14일
This research was supported by the Yeungnam University research grants in 2000.

2. Similarity measures

In the following, let A and B be fuzzy sets of the discrete universe of discourse $X=\{x_1, x_2, \ldots, x_n\}$. Let $a=\{a_1, a_2, \ldots, a_n\}$ and $b=\{b_1, b_2, \ldots, b_n\}$ be the vector representations of the fuzzy sets A and B, respectively, where a_i and b_i are membership values $\mu_A(x_i)$ and $\mu_B(x_i)$ with respect to x_i and x_j (i, j = 1, 2,...,n), respectively. The support supp(A) of a fuzzy set A is the crisp set of all $x_i \in X$ such that $\mu_A(x_i) > 0$:

$$\mathrm{supp}(A)=\{x_i|\mu_A(x_i)>0\}.$$

Furthermore, let F(X) be the class of all fuzzy sets of X. $A^c \in F(X)$ is the complement of $A \in F(X)$. The $\wedge$ and $\vee$ operators denote the minimum and maximum, respectively. $|\cdot|$ denotes the cardinality of a fuzzy set

defined as

$$|A| = \sum_{x_i \in X} \mu_A(x_i).$$

A one-parameter Minkowski class $d_r(a, b)$ of distance functions is defined as follows:

$$d_r(a, b) = [\sum_{i=1}^{n} |a_i - b_i|^r]^{1/r}, \quad r \geq 1.$$

The concept of similarity is generally interpreted as having the same shape, but not the same size or position. It may be interpreted in different ways depending on the context and situations. In fuzzy set theory and applications, the concept of similarity is generally interpreted as a generalization of the concept of equality. On the basis of this interpretation, we define similarity between fuzzy sets as the degree to which the fuzzy sets are equal.

The similarity measure is a function assigning a similarity value to the pair of fuzzy sets (A,B) that indicates the degree to which A and B are equal or how similar they are. It is also required that the similarity measures should satisfy the following properties [4]:

(P1) $S(A, B) = S(B, A)$, $A, B \in F$.

(P2) $S(D, D^c) = 0$, if D is a crisp set.

(P3) $S(E, E) = \max_{A,B \in F} S(A, B)$, for all $E \in F$.

(P4) If $A \subseteq B \subseteq C$ for all $A, B, C \in F$
then $S(A, B) \geq S(A, C)$ and $S(B, C) \geq S(A, C)$.

Under these backgrounds, several different measures of similarity have been proposed for measuring similarity among fuzzy sets [1-3, 7, 9, 10, 12-14]:

(i) Measures based on the geometric distance model

$$L(A, B) = 1 - d_\infty(a, b) = 1 - \max_i (|a_i - b_i|) \quad (1)$$

$$W(A, B) = 1 - \frac{\sum_{i=1}^{n} d_1(a, b)}{n} = 1 - \frac{\sum_{i=1}^{n} |a_i - b_i|}{n} \quad (2)$$

$$S_2(A, B) = 1 - \frac{\sum_{i=1}^{n} |a_i - b_i|}{\sum_{i=1}^{n(a_i + b_i)}}. \quad (3)$$

Geometric similarity measures represent similarity as proximity of fuzzy sets, and not as a measure of equality. Thus, the geometric similarity measures are best suited for measuring similarity among distinct fuzzy sets.

(ii) Measures based on the set-theoretic approach

$$M(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{\sum_{i=1}^{n} (a_i \wedge b_i)}{\sum_{i=1}^{n} (a_i \vee b_i)} \quad (4)$$

$$T(A, B) = \sup_{x_i \in X} \mu_{A \cap B}(x_i) = \max(\mu_{A \cap B}(x_1), \mu_{A \cap B}(x_2), \ldots, \mu_{A \cap B}(x_n)) \quad (5)$$

$$S_4(A, B) = \frac{1}{n} \sum_{i=1}^{n} \frac{a_i \wedge b_i}{a_i \vee b_i} \quad (6)$$

$$S_r(A, B) = \frac{\sum_{i=1}^{n} a_i \wedge b_i}{\max(\sum_{i=1}^{n} a_i, \sum_{i=1}^{n} b_i)} \quad (7)$$

$$S_J(A, B) = \frac{M(A \cap B)}{M(A) + M(B) - M(A \cap B)} \quad (8)$$

where M(A) is the size of fuzzy set A

$$M(A) = \int_{-\infty}^{\infty} A(x)\, dx.$$

The interpretation of similarity as approximate quality can better be represented by a set-theoretic similarity measures based on set-theoretic operations such as union and intersection. Therefore, the set-theoretic similarity measures are suitable for capturing similarity among overlapping fuzzy sets.

(iii) Measures based on the matching function and correlation

$$P(A, B) = S(A, B) = \frac{a \cdot b}{\max(a \cdot a, b \cdot b)} \quad (9)$$

$$k(A, B) = \frac{C(A, B)}{\sqrt{T(A) \cdot T(B)}} \quad (10)$$

where

$$T(A) = \sum_{i=1}^{n} [a_i^2 + (1 - a_i)^2]$$

and

$$C(A, B) = \sum_{i=1}^{n} [a_i \cdot b_i + (1 - a_i) \cdot (1 - b_i)]$$

The larger the values of the above all similarity measures, the more the similarity between the fuzzy sets A and B.

Under the background, we will define a similarity measure S_K: $F^2 \to R^+$ between fuzzy sets A and B as follows.

Definition 1. For $A, B \in F(X)$ we define

$$S_K(A, B) = \frac{a \cdot b}{a \cdot a + b \cdot b - a \cdot b} = \frac{\sum_{i=1}^{n} a_i \cdot b_i}{\sum_{i=1}^{n} a_i^{2+} \sum_{i=1}^{n} b_i^{2+} \sum_{i=1}^{n} a_i \cdot b_i} \quad (11)$$

and we call it the similarity measure of A and B.

Proposition 2. For $A, B \in F(X)$ we have $S_K(A,B) = S_K(B, A)$.

The proof is obvious.

Proposition 3. $S_K(D, D^c)=0$, if D is a crisp set.
The proof is obvious.

Proposition 4. $S_K(E,E)=\max_{A,B\in F} S_K(A,B)$, for all $E\in F(X)$.
The proof is obvious.

Proposition 5. If $A\subseteq B\subseteq C$ for all A, B, $C\in F(X)$, then $S_K(A,B)\geq S_K(A,C)$ and $S_K(B,C)\geq S_K(A,C)$.

Proof. Since $A\subseteq B\subseteq C$ implies $a_i\leq b_i\leq c_i$,

$\sum_{i=1}^n a_i\leq \sum_{i=1}^n b_i\leq \sum_{i=1}^n c_i$,

$\sum_{i=1}^n a_i\cdot b_i\leq \sum_{i=1}^n a_i\cdot c_i\leq \sum_{i=1}^n b_i\cdot c_i$ and

$\sum_{i=1}^n a_i^2\leq \sum_{i=1}^n b_i^2\leq \sum_{i=1}^n c_i^2$, then

$$
\begin{aligned}
&S_K(A,B)-S_K(A,C)\\
&=\frac{\sum_{i=1}^n a_i\cdot b_i}{\sum_{i=1}^n a_i^2+\sum_{i=1}^n b_i^2-\sum_{i=1}^n a_i\cdot b_i}-\frac{\sum_{i=1}^n a_i\cdot c_i}{\sum_{i=1}^n a_i^2+\sum_{i=1}^n c_i^2-\sum_{i=1}^n a_i\cdot c_i}\\
&=\frac{\sum_{i=1}^n a_i\cdot(c_i-b_i)\cdot(b_ic_i-a_i^2)}{\left(\sum_{i=1}^n (a_i-b_i)^2+\sum_{i=1}^n a_i\cdot b_{iRIGHT}\right)\left(\sum_{i=1}^n (a_i-c_i)^2+\sum_{i=1}^n a_i\cdot c_{iRIGHT}\right)}\\
&\geq\frac{\sum_{i=1}^n a_i\cdot(c_i-b_i)\cdot(b_i^2-a_i^2)}{\left(\sum_{i=1}^n (a_i-b_i)^2+\sum_{i=1}^n a_i\cdot b_{iRIGHT}\right)\left(\sum_{i=1}^n (a_i-c_i)^2+\sum_{i=1}^n a_i\cdot c_i\right)}\\
&\geq 0.
\end{aligned}
$$

Futhermore

$$
\begin{aligned}
&S_K(B,C)-S_K(A,C)\\
&=\frac{\sum_{i=1}^n b_i\cdot c_i}{\sum_{i=1}^n b_i^2+\sum_{i=1}^n c_i^2-\sum_{i=1}^n b_i\cdot c_i}-\frac{\sum_{i=1}^n a_i\cdot c_i}{\sum_{i=1}^n a_i^2+\sum_{i=1}^n c_i^2-\sum_{i=1}^n a_i\cdot c_i}\\
&=\frac{\sum_{i=1}^n c_i\cdot(b_i-a_i)\cdot(c_i^2-a_i\cdot b_i)}{\left(\sum_{i=1}^n (b_i-c_i)^2+\sum_{i=1}^n b_i\cdot c_{iRIGHT}\right)\left(\sum_{i=1}^n (a_i-c_i)^2+\sum_{i=1}^n a_i\cdot c_{iRIGHT}\right)}\\
&\geq\frac{\sum_{i=1}^n c_i\cdot(b_i-a_i)\cdot(c_i^2-a_i^2)}{\left(\sum_{i=1}^n (b_i-c_i)^2+\sum_{i=1}^n b_i\cdot c_{iRIGHT}\right)\left(\sum_{i=1}^n (a_i-c_i)^2+\sum_{i=1}^n a_i\cdot c_i\right)}\\
&\geq 0.
\end{aligned}
$$

Hence $S_K(A,B)\geq S_K(A,B)$ and $S_K(B,C)\geq S_K(A,C)$.

Proposition 6. For A, $B\in F(X)$ we have if A=B, then $S_K(A,B)=1$.

The proof is obvious.

Proposition 7. For A, $B\in F(X)$ we have $0\leq S_K(A,B)\leq 1$.

Proof. First we will show that the first inequality $S_K(A,B)\geq 0$ holds. Since $a_i\geq 0$, $b_i\geq 0$ implies $a\cdot b=\sum_{i=1}^n a_i\cdot b_i\geq 0$ we obtain

$$
\begin{aligned}
S_K(A,B)&=\frac{\sum_{i=1}^n a_i\cdot b_i}{\sum_{i=1}^n a_i^2+\sum_{i=1}^n b_i^2-\sum_{i=1}^n a_i\cdot b_i}\\
&\geq\frac{\sum_{i=1}^n a_i\cdot b_i}{\sum_{i=1}^n a_i^2+\sum_{i=1}^n b_i^2-2\sum_{i=1}^n a_i\cdot b_i}\\
&=\frac{\sum_{i=1}^n a_i\cdot b_i}{\sum_{i=1}^n (a_i-b_i)^2}\geq 0.
\end{aligned}
$$

Furthermore, we have

$$
\begin{aligned}
&\sum_{i=1}^n a_i^2+\sum_{i=1}^n b_i^2-\sum_{i=1}^n a_i\cdot b_i-\sum_{i=1}^n a_i\cdot b_i\\
&=\sum_{i=1}^n a_i^2+\sum_{i=1}^n b_i^2-2\sum_{i=1}^n a_i\cdot b_i\\
&=\sum_{i=1}^n (a_i-b_i)^2\geq 0
\end{aligned}
$$

which proves the second inequality and completes the proof.

Proposition 8. For A, $B\in F(X)$, $S_K(A,B)=1$ if and only if A=B.

Proof. The sufficiency is obvious.
Necessity: Suppose

$$\sum_{i=1}^n a_i^2+\sum_{i=1}^n b_i^2-\sum_{i=1}^n a_i\cdot b_i=\sum_{i=1}^n a_i\cdot b_i$$

i.e,

$$\sum_{i=1}^n a_i^2+\sum_{i=1}^n b_i^2-2\sum_{i=1}^n a_i\cdot b_i=\sum_{i=1}^n (a_i-b_i)^2=0.$$

This implies $a_i-b_i=0$, i.e., $a_i=b_i$ for all i(i=1, ·, n) and the proof is completed.

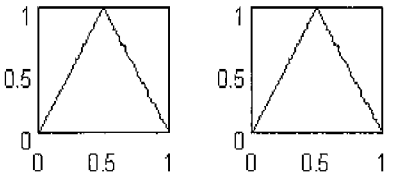
3. Numerical examples

In this section we provide two numerical examples showing the characteristics of the proposed similarity measure $S_K(A,B)$ and other similarity measures, i.e., i) measures based on the geometric distance model: L(A,B), W(A,B), and S2(A,B), ii) measures based on the set-theoretic approach: M(A,B), T(A,B),S4(A,B), the similarity ratio Sr(A,B), and SJ(A,B), iii) Measures based on the matching function and correlation: P(A,B) and the correlation coefficientk(A,B). The first one is an example showing the behavior of similarity measures for fuzzy sets with varying degree of height. The second one is an

example showing the behavior of similarity measures for fuzzy sets with varying degree of overlap.

In computer implementation, continuous domains have been discretized. Each result has been obtained by using the MATLAB on the discretizeddomains.

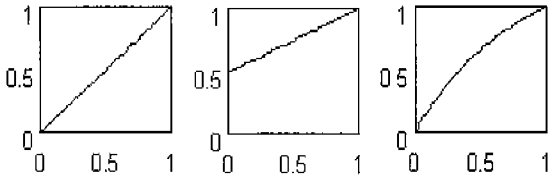
Example 1. Consider two fuzzy sets A and B defined as shown in Fig. 1 (a) and (b). The behavior of the similarity measures for a fixed fuzzy set A and a fuzzy set B with varying degree of height from 0 to 1 is shown in Fig. 2. The abscissa andordinate of Fig. 2 represent the height of the fuzzy set B and the similarity degree between A and B, respectively.



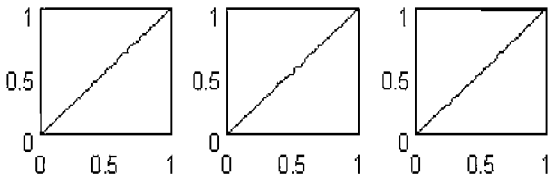
(a) fuzzy set A(b) fuzzy set B

Fig. 1. Fuzzy sets A and B.

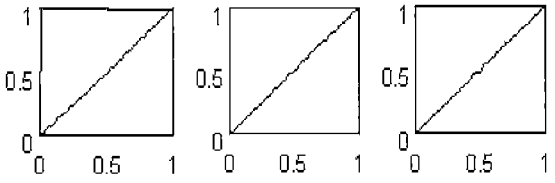
It may be difficult to say which similarity measure



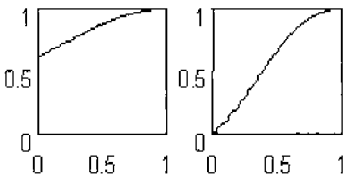
(a) L(A,B) (b) W(A,B) (c) S2(A,B)



(d) M(A,B) (e) T(A,B) (f) S4(A,B)



(g) Sr(A,B) (h) SJ(A,B) (i) P(A,B)



(j) k(A,B) (k) SK(A,B)

Fig 2. The behavior of the similarity measures.

is the best (in the sense of the reliablility) from Fig. 2. But we have the following remarks: We can see that the behavior of the similarity measures S2(A,B), k(A,B) and the proposed SK(A,B) is nonlinearbut the other similarity measures investigated in this example shows the linear behavior as the height of the fuzzy set B varies from 0 to 1.

Example 2. Consider two fuzzy sets A and B defined as shown in Fig. 3. The behavior of the similarity measures for a fixed fuzzy set A and a fuzzy set B with varying degree of overlap is shown in Fig. 4. The abscissa andordinate of Fig. 2 represent the degree of overlap of the fuzzy sets A and B, and the similarity degree between A and B, respectively.

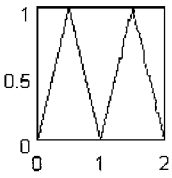
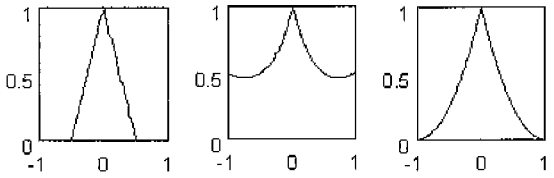
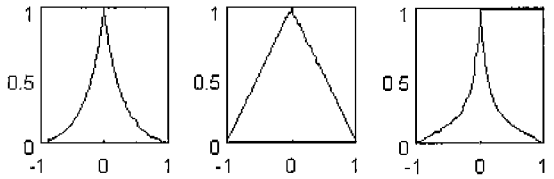


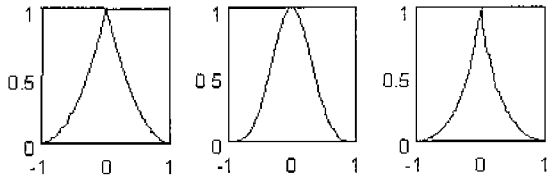
Fig. 3. Fuzzy sets A and B.



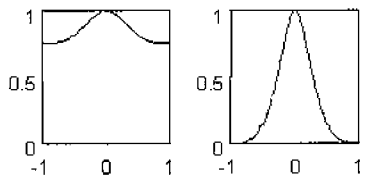
(a) L(A, B) (b) W(A, B) (c) S2(A, B)



(d) M(A, B) (e) T(A, B) (f) S4(A, B)



(g) Sr(A, B) (h) SJ(A, B) (i) P(A, B)



(j) k(A, B) (k) SK(A, B)

Fig 4. The behavior of the similarity measures.

From Fig. 3 and 4, we have the following remarks. By intuitional observation, we can see that the similarity measures W(A, B) and k(A, B) providing much larger

values than those of the others seems not reliable for us.

From the above discussion on the properties and numerical examples, we can conclude that the proposed similarity measure SK(A,B) can be useful in the classification of some similar sets even though it is hard to say that the proposed similarity measure is the best.

4. Conclusions

We have proposed a new similarity measure of fuzzy sets and examine its properties. Numerical examples have been provided to compare the characteristics of the proposed similarity measure and other similarity measures.

Future work is to research on the investigation of other properties and characteristics of the proposed measure and its applications to modeling of fuzzy systems.

References

[1] S. M. Chen, M. S. Yeh and P. Y. Hsiao, A comparison of similarity measures of fuzzy values, *Fuzzy Sets and Systems*, vol. 72, pp. 79-89, 1995.

[2] T. Gerstenkorn and J. Manko, Correlation of intuitionistic fuzzy sets, *Fuzzy Sets and Systems*, vol. 44, pp. 39-43, 1991.

[3] L.-K. Hyung, Y. S. Song and K. M. Lee, Similarity measure between fuzzy sets and between elements, *Fuzzy Sets and Systems*, vol. 62, pp 291-293, 1994.

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

[5] C. P. Pappis and N. I. Karapapilidis, A comparative assessment of measures of similarity of fuzzy values, *Fuzzy Sets and Systems*, vol. 56, pp. 171-174, 1993.

[6] M. Setnes, R. Babuska, U. Kaymak and H. R. van N. Lemke, Similarity measures in fuzzy rule base simplification, *IEEE Trans. Systems*, Man, Cybern. Part B, vol. 28, no. 3. pp. 376-386, 1998.

[7] W. J. Wang, New similarity measures on fuzzy sets and on elements, *Fuzzy Sets and Systems*, vol. 85, pp. 305-309, 1997.

[8] R. Zwick, E. Carlstein and D. V. Budescu, Measures of similarity among fuzzy concepts: a comparative analysis, Int. J. Approximate Reasoning, vol. 1, pp. 221-242, 1987.

[9] E. Chouraqui and C. Inghilterra, A model of case-based reasoning for solving problems of geometry in a tutoring system, in J-M Laborde (Ed.), Intelligent Learning Environments: The case of geometry, Springer-Verlag, Berlin, pp. 1-16, 1996.

[10] E. Castineira, S. Cubillo and E. Trillas, On a similarity ratio, Proc. 1999 EUSFLAT-ESTYLF Joint Conf., pp. 417-420, 1999.

[11] J. Yen and L. Wang, Application of statistical information criteria for optimal fuzzy model construction, *IEEE Trans. Fuzzy Systems*, vol. 6, no. 3, pp. 362-372, 1998.

[12] Y. Jin, W. von Seelen, and B. Sendhoff, On generating flexible, complete, consistent and compact fuzzy rule systems from data using evolution strategies, *IEEE Trans. Systems, Man, Cybern.* Part B vol. 29, no. 6, pp. 829-845, 1999.

[13] X. Wang, B. D. Baets, and E, Kerre, A comparative study of similarity measures, *Fuzzy Sets and Systems*, vol. 73, pp. 259-268, 1995.

[14] C. Yu, Correlation of fuzzy numbers, *Fuzzy Sets and Systems*, vol. 55, pp. 303-307, 1993.

[15] L. A. Zadch, Similarity relations and fuzzy orderings, Inform. Sci., vol. 3, pp. 177-200, 1971

저 자 소 개

권순학 (Soon H. Kwon)

1983년 : 서울대학교 제어계측공학과 (학사)

1985년 : 서울대학교 대학원 제어계측공학과 (공학석사)

1984년~1986년 : 삼성전자㈜ 주임연구원

1986년~1991년 : 한국과학기술연구원 연구원

1992년~1995년 : 동경공업대학 (공학박사)

1996년~현재 : 영남대학교 전자정보공학부 부교수

관심분야 : 지능시스템의 모델링 및 제어

E-mail : shkwon@yu.ac.kr