Годишник на секция “Информатика”
Съюз на учените в България
Том 4, 2011, 90-94

Annual of “Informatics” Section
Union of Scientists in Bulgaria
Volume 4, 2011, 90-94

A NOTE ON SIMILARITY MEASURES DEFINED OVER INTUITIONISTIC FUZZY SETS

Peter Vassilev

Institute of Biophysics and Biomedical Engineering
Bulgarian Academy of Sciences
105 Acad. G. Bonchev Str., 1113 Sofia, Bulgaria,
E-mail: peter.vassilev@gmail.com

Abstract: In previous research, the author formulated the so called $\xi, \eta-$ similarity measures (obtained with the use of normalized distances). Here a possible novel approach for generating intuitionistic fuzzy sets from a specific universe is proposed. The application of the $\xi, \eta-$ similarity measures over these intuitionistic fuzzy sets is considered.
Keywords: Intuitionistic fuzzy sets, Distance, Metric, Pseudometric, Similarity measure.

1. Introduction

Intuitionistic fuzzy sets (IFS) were introduced by K. Atanassov (see [1]) as a generalization and extension of the concept of fuzzy sets. We will briefly remind some of the basic definitions and notions.

Let X be a universe set (i.e. the set of all the (relevant) elements that will be considered) and let $A \subset X$. An intuitionistic fuzzy set is a set

$$A^* \stackrel{\text{def}}{=} \{\langle x, \mu_A(x), \nu_A(x)\rangle | x \in X\}$$

where the functions

$$\mu_A : X \to [0,1] \quad \text{and} \quad \nu_A : X \to [0,1]$$

reflect the degree of membership (belongingness) and non-membership (non-belongingness) of the element $x \in X$ to the set A, respectively, and for every $x \in X$ it is fulfilled that:

$$0 \leq \mu_A(x) + \nu_A(x) \leq 1 \tag{1}$$

The function $\pi_A : X \to [0,1]$, which is given by

$$\pi_A(x) \stackrel{\text{def}}{=} 1 - \mu_A(x) - \nu_A(x) \; \forall x \in X \tag{2}$$

defines the degree of uncertainty of the membership of the element x to the set A.

2. Generating intuitionistic fuzzy sets from a universe set

Here, we will discuss the possibility to algorithmically extract such membership, non-membership and uncertainty values in some special cases. Let us consider the finite sequence

$$\{a_n\}_{n=1}^{1000} = \underbrace{b\ldots b}_{n\text{-times}}$$

Here we consider each of these entities as a "string", i.e. we will only be interested in the string's length and the symbols it contains (a more general approach allowing the considerations of multi-symbol "strings" will be detailed in future study, but here we are only using this as an illustrative example of such a possibility). Thus, as the formal definition above states, let X be the set of all these "strings" and let A be the subset of all strings with odd lengths, and let B the subset with even lengthed strings. In order not to complicate the notation we will use $J, J \subset X$, to define the general approach. Let us denote the length of x by $l(x)$. Now for any element $x \in X$ there are only three possible cases:

1) $x \in J$, then $\mu_J(x) = 1$, $\nu_J(x) = 0$

2) $x \notin J$ &$\exists y \in J : l(y) > l(x)$ $\mu_J = \frac{l(x)}{l(y')}$, $\nu_J(x) = 0$, where y' denotes the element in J for which the difference of lengths $l(y') - l(x)$ is minimal.

3) $x \notin J$ &$\forall y \in J : l(x) > l(y)$ $\mu_J = \frac{l(y')}{l(x)}$, $\nu_J(x) = \frac{l(x)-l(y')}{l(x)}$, where y' denotes the element in J for which the difference of lengths $l(x) - l(y')$ is minimal.

The motivation for this definition is simple. When we have **Case 2)**, x is a substring of y, i.e. it is partially contained in y. Among all those elements we find the one closest in length to x. The remaining part is attributed to uncertainty and not to non-membership, since the eventual continuation of the substring x may potentially be complemented to y. In **Case 3)**, however, this is impossible, i.e. the string cannot be reduced to a substring, hence the remainder is attributed to non-membership.

Remark 1. The above procedure, in this special case, obviously produces for every subset J a unique intuitionistic fuzzy set.

This observation is important for two reasons:

First, we can test how different similarity measures and distances over IFS perform in this case (see e.g. [2, 3, 4])

Second, we can apply the apparatus developed in [5] to generate a distance/similarity measure that performs better.

3. Summary of previous results

In a previous paper [5], the author defined a way of generating new distances and similarity measures. Here we will briefly recall (omitting the proofs) the conceptual idea. Further, we will denote the class of all IFSs defined over a universe set X by $IFS(X)$.

The following theorems are fundamental in this approach, so we give them here in full:

Theorem 1. Let n and k be arbitrary positive integers. Let $f^{(i)} : IFS(X) \times IFS(X) \rightarrow [0, +\infty), i = \overline{1, n}$ be metrics and $g^{(j)} : IFS(X) \times IFS(X) \rightarrow [0, +\infty), j = \overline{1, k}$ be pseudometrics. Then, for every fixed real numbers $\alpha_i \in (0, 1], i = \overline{1, n}$ and $\beta_j \in (0, 1], j = \overline{1, k}$, the mapping $t_{\alpha,\beta} : IFS(X) \times IFS(X) \rightarrow [0, +\infty)$ that is given by:

$$t_{\alpha,\beta}(A, B) = \sum_{i=1}^{n} f_{\alpha_i}^{(i)}(A, B) + \sum_{j=1}^{k} g_{\beta_j}^{(j)}(A, B), \tag{3}$$

where $\alpha \stackrel{\text{def}}{=} (\alpha_1, \ldots, \alpha_n)$, $\beta \stackrel{\text{def}}{=} (\beta_1, \ldots, \beta_k)$ is a metric.

If $A, B, C \in IFS(X)$ are such that $A \subseteq B \subseteq C$ and we have:

$$f^{(i)}(A, B) \leq f_i(A, C);\ f^{(i)}(B, C) \leq f^{(i)}(A, C),\ \overline{i = 1, n};$$

$$g^{(j)}(A, B) \leq g^{(j)}(A, C);\ g^{(j)}(B, C) \leq g^{(j)}(A, C),\ \overline{j = 1, k},$$

then it is true:

$$t_{\alpha,\beta}(A, B) \leq t_{\alpha,\beta}(A, C);\ t_{\alpha,\beta}(B, C) \leq t_{\alpha,\beta}(A, C)$$

Theorem 2. Let $\xi > 0$ and $\eta \geq 1$ be fixed real numbers and $d : IFS(X) \times IFS(X) \rightarrow [0, +\infty)$ be an arbitrary metric (pseudometric). Then, $d^* : IFS(X) \times IFS(X) \rightarrow [0, 1]$ given by

$$d^*(A, B) \stackrel{\text{def}}{=} \frac{d(A, B)}{\xi + \eta d(A, B)} \tag{4}$$

is a normalized metric (pseudometric). Moreover, if d is a non-Archimedean metric, then d^* is a non-Archimedean metric, too.

Hence,

Corollary 1. The similarity measure introduced by

$$s_{\alpha,\beta}^*(A, B) \stackrel{\text{def}}{=} 1 - t_{\alpha,\beta}^*(A, B),$$

where

$$t_{\alpha,\beta}^*(A, B) = \frac{t_{\alpha,\beta}(A, B)}{\xi + \eta t_{\alpha,\beta}(A, B)} \quad \forall A, B \in IFS(X)$$

is well defined.

In short we will summarize the above as follows: new distances may be introduced by adding (weighted) pseudometrics and distances. These in turn may be normalized to the interval $[0, 1]$ by the use of two parameters ξ and η. The normalized distance can then be used to define a ξ, η-similarity measure (for fixed ξ and η).

4. Application of the results

Let $X = \{1, 11, 111, 1111, 11111\}$ be our universe set. Choosing

$$A = \{111, 11111\} \text{ and } B = \{11, 111, 1111\}$$

we obtain the following two intuitionistic fuzzy sets (according to the three cases discussed above).

$$\widehat{A} = \{\langle 1, \frac{1}{3}, 0\rangle, \langle 11, \frac{2}{3}, 0\rangle, \langle 111, 1, 0\rangle, \langle 1111, \frac{4}{5}, 0\rangle, \langle 11111, 1, 0\rangle$$

$$\widehat{B} = \{\langle 1, 1/2, 0\rangle, \langle 11, 1, 0\rangle, \langle 111, 1, 0\rangle, \langle 1111, 1, 0\rangle, \langle 11111, \frac{4}{5}, \frac{1}{5}\rangle$$

Let us consider now the distance defined by:

$$d(\widehat{A}, \widehat{B}) = \frac{1}{10} \sum_{i=1}^{5} |\mu_A(x) - \mu_B(x)| + |\nu_A(x) - \nu_B(x)|$$

From it we will obtain that $d(\widehat{A}, \widehat{B}) = 11/100$. This implies that the two sets are relatively close, since $s(\widehat{A}, \widehat{B}) \stackrel{\text{def}}{=} 1 - d(\widehat{A}, \widehat{B})$ is close to 1, which is contrary to our intuitive notions since the two sets are quite different. Now let us improve this distance by adding a pseudometric to it, thus obtaining a new distance. We will use the following pseudometric:

$$\varrho_2(\widehat{A}, \widehat{B}) = \frac{1}{5} \sum_{i=1}^{5} |(\mu_A(x) - \nu_A(x))^2 - (\mu_B(x) - \nu_B(x))^2| = \frac{1309}{4500}$$

We define the new distance, as follows (choosing $\xi = \eta = 1$):

$$u(\widehat{A}, \widehat{B}) = \frac{10d(\widehat{A}, \widehat{B}) + 5\varrho_2(\widehat{A}, \widehat{B})}{1 + 10d(\widehat{A}, \widehat{B}) + 5\varrho_2(\widehat{A}, \widehat{B})} = \frac{2299}{3199} \approx 0.7186$$

The corresponding similarity measure $s_u(\widehat{A}, \widehat{B}) = 1 - u(\widehat{A}, \widehat{B}) = 1 - \frac{2299}{3199} \approx 0.29$ is much further away from 1, and thus it represents more accurately our intuitive notion of the situation.

5. Conclusion

A new way for introducing intuitionistic fuzzy sets from a universe of specific type was introduced. The previously defined distances and similarity measures were applied to the newly received sets, and “improvement” in the obtained values in accordance with the intuitive notions was observed. Further research will be conducted, especially in the area of constructive (in the sense of automatic) generation of membership and non-membership functions over universe sets of specific types.

Acknowledgments

This paper is partially supported by grant DID-2-29 “Modelling Processes with Fixed Development Rules” of the Bulgarian National Science Fund and grant BG051PO001-3.3.04/40 of the Operative Program “Human Resource and Development”.

References

[1] Atanassov, K. *Intuitionistic Fuzzy Sets*, Springer Physica-Verlag, Heidelberg, 1999

[2] Wang, Xin. Distance measure between intuitionistic fuzzy sets. *Pattern Recognition Letters* 26 (2005), 2063-2069.

[3] Szmidt, E., J. Kacprzyk, 2000. Distances between intuitionistic fuzzy sets. *Fuzzy Sets Systems* 114 (3), 505-518.

[4] Ye, J. Cosine similarity measures for intuitionistic fuzzy sets and their applications. *Mathematical and Computer Modelling* 53 (2011), 91-97.

[5] Vassilev, P. A note on distance and similarity measures between intuitionistic fuzzy sets. (in press)