The Tenth International Symposium on Operations Research and Its Applications (ISORA 2011)
Dunhuang, China, August 28–31, 2011

A Similarity Measure for Interval-valued Fuzzy Sets and Its Application in Supporting Medical Diagnostic Reasoning

Hong-mei Ju [1]　　Feng-Ying Wang[1]

1 School of Information, Beijing WUZI University, Beijing, China, 101149

Abstract We propose a new similarity measure for interval-valued fuzzy sets and show its usefulness in medical diagnostic reasoning. We point out advantages of this new concept over the most commonly used similarity measures which are the counterparts of distances. The measure we propose involves both similarity and dissimilarity.

Keywords Similarity measure; Interval-valued fuzzy sets; Medical diagnostic reasoning

1 Introduction

Interval-valued fuzzy sets (Zadeh [1]) can be viewed as a generalization of fuzzy sets (Zadeh [2]) that may better model imperfect information which is omnipresent in any conscious decision making. We present here interval-valued fuzzy sets as a tool for a more human consistent reasoning under imperfectly defined facts and imprecise knowledge, with an application in supporting medical diagnosis. More specifically, we have set of data, i.e. a description of a set of symptoms S , and a set of diagnoses D. We describe a state of a patient knowing results of his/her medical tests. We use the concept of an interval-valued fuzzy set that makes it possible to express many new aspects of imperfect information. For instance, in many cases information obtained cannot be classified due to lack of knowledge, discriminating power of measuring tools, etc. In such a case the use of a degree of membership and non-membership can be an adequate knowledge representation solution.

The proposed method of diagnosis involves a new measure of similarity for interval-valued fuzzy sets. For each patient the similarity measures for his or her particular set of symptoms and a set of symptoms that are characteristic for each diagnosis are calculated. The highest similarity points out a proper diagnosis.

In Section 2 we briefly overview interval-valued fuzzy sets. In Section 3 we propose a new measure of similarity for interval-valued fuzzy sets. In Section 4 we use the proposed similarity measure to single out a diagnosis for considered patients. We compare the solution obtained with a final diagnosis pointed out by looking for the smallest distance between symptoms characteristic for a patient and symptoms describing the illnesses. We give some conclusions in Section 5.

2 Brief Introduction to Interval-valued Fuzzy Sets

As opposed to a fuzzy set in X (Zadeh [2]), given by

$$A=\{x, A(x) \mid x\in X\} \tag{1}$$

where $A(x)\in[0,1]$ is the membership function of the fuzzy set A, an interval-valued fuzzy set (Zadeh [1]) A is given by

$$A=\{x,[A^-(x),A^+(x)] \mid x\in X\} \tag{2}$$

where $A^-,\ A^+ : X\to[0,1]$ such that

$$0\le A^-(x)\le A^+(x)\le 1,\ x\in X \tag{3}$$

and $[A^-(x),A^+(x)]$ is the interval degree of membership function of an element x to the set A. Obviously, each fuzzy set may be represented by the following interval-valued fuzzy set

$$A=\{x,\ [A(x),A(x)] \mid x\in X\} \tag{4}$$

For each interval-valued fuzzy set in X, we call

$$\pi_A(x)=A^+(x)-A^-(x) \tag{5}$$

an interval-valued fuzzy index (or a hesitation margin)of $x\in A$ and, it expresses a lack of knowledge of the degree x belongs to A.

A complement of an interval-valued fuzzy set A is

$$A^C=\{x,\ [1-A^+(x),1-A^-(x)] \mid x\in X\} \tag{6}$$

3 A New Similarity Measure

We propose here a new similarity measure for interval-valued fuzzy sets using a geometrical interpretation of interval-valued fuzzy sets given in Ju [3, 4], which implies that any combination of the parameters characteristic for elements belonging to an interval-valued fuzzy set can be represented inside triangle ABC (Figure 1).

In the simplest situations we assess a similarity of any two elements E and F belonging to an interval-valued fuzzy set (or sets). The proposed measure indicates if E is more similar to F or to F^C, where F^C is a complement of F, i.e. if E is more similar or more dissimilar to F (Figure 1).

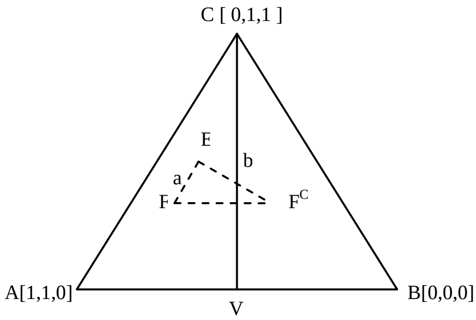

Fig. 1. Triangle ABC explaining a ratio-based measure of similarity

Definition 1.
$$Sim(E,F) = \frac{l_{IVFS}(E,F)}{l_{IVFS}(E,F^C)} = \frac{a}{b} \tag{7}$$

where: a is a distance (E,F) from $E\ [E^-,E^+,\pi_E]$ to $F\ [F^-,F^+,\pi_F]$, b is the distance (E,F^C) from $E\ [E^-,E^+,\pi_E]$ to $F^C\ [1-F^+,1-F^-,\pi_F]$. For (7) we have $0 < Sim(E,F) \le \infty$ and $Sim(E,F) = Sim(F,E)$.

Note that:

(1) $Sim(E,F) = 0$ means the identity of E and F,

(2) $Sim(E,F) = 1$ means that E is to the same extent similar to F and F^C (i.e., values bigger than 1 mean in fact a closer similarity of E and F^C to E and F),

(3) when $E = F^C$ (or $E^C = F$), i.e. $l_{IVFS}(E,F^C) = l_{IVFS}(E^C,F) = 0$ means the complete dissimilarity of E and F (or the identity of E and F^C), and then $Sim(E,F) \to \infty$,

(4) when $E = F = F^C$ means the highest entropy (see [4]) for both elements E And F i.e. the highest "fuzziness"–not too constructive a case when looking for compatibility (both similarity and dissimilarity).

So, when applying the measure (7) to analyze the similarity of two objects, the values are of interest. The proposed measure (7) was constructed for selecting objects which are more similar than dissimilar.

Now we show that a measure of similarity (7) between is more powerful than a simple distance, giving an example of medical diagnostic reasoning.

4 Medical Diagnostic Reasoning via Distances for Interval-valued Fuzzy Sets

To make a proper diagnosis D for a patient with given values of symptoms S , a medical knowledge base is necessary that involves elements described in terms of interval-valued fuzzy sets. We consider the data as: the set of diagnoses is D = {Fever, Malaria, Typhoid, Stomach problem, Chest problem}, and the set of symptoms is S = {Chest-pain, Cough, Stomach pain, Headache, Temperature}. The data are given in Table 1.

	Fever	Malaria	Typhoid	Stomach problem	Chest problem
Chest pain	[0.1, 0.3,0.2]	[0.1,0.2,0.1]	[0.1,0.1,0]	[0.2,0.3,0.1]	[0.8,0.9,0.1]
Cough	[0.4, 0.7,0.3]	[0.7,1,0.3]	[0.2,0.4,0.2]	[0.2,0.3,0.1]	[0.2,0.2,0]
Stomach pain	[0.1, 0.3,0.2]	[0,0.1,0.1]	[0.2,0.3,0.1]	[0.8, 1,0.2]	[0.2,0.2,0]
Headache	[0.3, 0.5,0.2]	[0.2,0.4,0.2]	[0.6,0.9,0.3]	[0.2,0.6,0.4]	[0, 0.2,0.2]
Temperature	[0.4, 1,0.6]	[0.7,1,0.3]	[0.3,0.7,0.4]	[0.1,0.3,0.2]	[0.1,0.2,0.1]

Table 1 Symptoms characteristic for the diagnoses considered

	Chest-pain	Cough	Stomach pain	Headache	Temperature
Li-hong	[0.1,0.4,0.3]	[0.6,0.9,0.3]	[0.2,0.2,0]	[0.6,0.9,0.3]	[0.8,0.9,0.1]
Wang-wei	[0.1,0.2,0.1]	[0.1,0.3,0.2]	[0.6,0.9,0.3]	[0.4,0.6,0.2]	[0,0.2,0.2]
Zhang-liang	[0,0.5,0.5]	[(0.2,0.3,0.1]	[0,0.4,0.4]	[0.8,0.9,0.1]	[0.8,0.9,0.1]
Sun-fang	[0.3,0.6,0.3]	[0.7,0.8,0.1]	[0.3,0.6,0.3]	[0.5,0.6,0.1]	[0.6,0.9,0.3]

Table 2 Symptoms characteristic for the patients considered

Each symptom is described by：a interval degree of membership and a hesitation degree. For example, for malaria the temperature is high [0.7, 1, 0.3], for a chest problem the temperature is low [0.1, 0.2, 0.1], etc. The set of patients is P = {Li-hong, Wang-wei, Zhang-liang, Sun-fang}. The symptoms are given in Table2.

Now we derive a diagnosis for each patient p_i, $i=1,\ldots,4$ using the proposed similarity measure (7). We propose:

i to calculate for each patient p_i a similarity measure (7) between his or her symptoms (Table 2) and symptom $s=1,\ldots,5$ characteristic for each diagnosis d_k, $k=1,\ldots,5$ (Table 1),

ii to single out the lowest numerical value from the obtained similarity measures

	Fever	Malaria	Typhoid	Stomach problem	Chest problem
Li-hong	0.75	1.19	1.31	3.27	∞
Wang-wei	2.1	3.73	1.1	0.35	∞
Zhang-liang	0.87	1.52	0.42	2.61	∞
Sun-fang	0.95	0.77	1.67	∞	2.56

Table 3. Similarities of symptoms for each patient to the considered set of possible diagnoses

From Definition 1, similarity measure (7) for patient p_i-between his or her symptoms and the symptoms characteristic for diagnosis d_k is

$$\mathrm{Sim}(s(p_i),d_k)=\frac{1}{5}\sum_{j=1}^{5}\frac{\left|s_j^-(p_i)-d_k^-(s_j)\right|+\left|s_j^+(p_i)-d_k^+(s_j)\right|+\left|\pi_j(p_i)-\pi_k(s_j)\right|}{\left|s_j^-(p_i)+d_k^+(s_j)-1\right|+\left|s_j^+(p_i)+d_k^-(s_j)-1\right|+\left|\pi_j(p_i)-\pi_k(s_j)\right|} \quad (8)$$

For example, for Li Hong, the similarity measures for his temperature and the temperature characteristic for the chest problem is

$$\mathrm{Sim}_T(s(p_1),d_5)=\frac{|0.8-0.1|+|0.9-0.2|+|0.1-0.1|}{|0.8-0.8|+|0.9-0.9|+|0.1-0.1|}\to\infty \quad (9)$$

Similarly, for all five symptoms, from (8) we get the similarity measure indicating how close are Li Hong’s symptoms to symptoms for a chest problem, and obtain ∞ because of (9), i.e. that at least one of symptoms is opposite to that for a chest problem. This indication cannot be obtained while considering just distances between symptoms.

In fact, many people may ask the question that why we define a new similarity measure for interval-valued fuzzy sets using (7) while not just using the widely received measure of Hamming distance.

For example, we still want to make a judgment that a patient should belong to a certain diagnose. We propose to solve the problem in the following way:

i to calculate for each patient p_i a similarity measure using the normalized Hamming distance between his or her symptoms (Table 2) and symptom $s = 1,...,5$ characteristic for each diagnosis d_k, $k = 1,...,5$ (Table 1),

ii to single out the lowest numerical value from the obtained similarity measures .

The normalized Hamming distance for all symptoms of patient i-th from diagnosis k is

$$l(s(p_i), d_k) = \frac{1}{10}\sum_{j=1}^{5}\left|s_j^{-}(p_i) - d_k^{-}(s_j)\right| + \left|s_j^{+}(p_i) - d_k^{+}(s_j)\right| + \left|\pi_j(p_i) - \pi_k(s_j)\right| \quad (10)$$

	Fever	Malaria	Typhoid	Stomach problem	Chest problem
Li-hong	0.28	0.24	0.28	0.54	0.56
Wang-wei	0.4	0.5	0.31	0.14	0.42
Zhang-lian	0.38	0.44	0.32	0.50	0.55
Sun-fang	0.28	0.30	0.38	0.44	0.54

Table 4. The normalized Hamming distances for each patient to the considered set of possible diagnoses

The distances (10) for each patient from the set of possible diagnoses are given in Table 4. The lowest distance points out a diagnosis: Li-hong suffers from malaria, Wang-wei from a stomach problem, Zhang-liang from typhoid, whereas Sun-fang from fever.

Now let's do a comparison between the results for the considered patients in Table 3 and Table 4. These results (Table 3) are different than those when we just considered the distances (Table 4). From Table 4, Wang-wei suffers from stomach problems, Zhang-liang from typhoid, but Li-hong from fever (not from malaria), and Sun-fang suffers from malaria (not from fever). These differences are because the similarity measure (7) can be small but at the same time the distance between the symptom of the patient and the complementary symptom characteristic for the examined illness can be smaller. We can easily see the advantages of this new concept over the most commonly used similarity measures which are the counterparts of distances.

5 Conclusions

The method proposed, performing diagnosis on the basis of the calculation of a new similarity measure for interval-valued fuzzy sets, makes it possible to avoid drawing conclusions about strong similarity between interval-valued fuzzy sets on the basis of the small distances between these sets.

Acknowledgements

This research work is supported by Funding Project for Academic Human Resources Development in Institutions of Higher Learning under the Jurisdiction of Beijing Municipality (PHR20101026) and Funding Project for Youth Scientific Research of Beijing Wuzi University (2010-2012)

References

[1] L.A.Zadeh (1975). The concept of a linguistic variable and its application to approximate reasoning—I, Inform. Sci. 8 199–249.

[2] L. A. Zadeh (1965). Fuzzy sets, Information and Control 8, 338–353.

[3] Hongmei Ju (2007). Distances between interval-valued fuzzy sets. Management Science and System Dynamics, Volume 3, 941-948**.**

[4] Hongmei Ju (2008). Entropy for Interval-valued Fuzzy Sets, Fuzzy information and engineering, Volume1，358-366.

[5] Hongmei Ju, Application of interval-valued Fuzzy Sets in Medical Diagnosis [J]. Proceedings of 2006 Asian Fuzzy Systems Society International Conference (2006)126-130.