

The New Similarity Measure for Fuzzy Sets and its Application to Medical Diagnostic Reasoning

Pranamika Kakati

Department of Computer Application, Girijananda Chowdhury Institute of Management & Technology, Guwahati, Assam, India.

ABSTRACT

In this article we shall deal with the new Similarity measure for Fuzzy sets based on the extended definition of complementation. Measuring similarity between Fuzzy sets is very important in the application of Fuzzy set theory. The new Similarity measure is based on the fact that Fuzzy membership function and Fuzzy membership value for the complement of a Fuzzy set are two different things. In this paper, our purpose is to put forward an algorithm to show the effectiveness of this new Similarity measure in supporting Medical Diagnostic reasoning. Also, we shall demonstrate the application of the proposed algorithm evaluating some collected data and verify the obtained results with the help of traditional Hamming Distance and Euclidean Distance measures.

General Terms

Algorithm for Medical Diagnostic Reasoning.

Keywords

Fuzzy set, Complement of a Fuzzy set, Fuzzy membership value, Fuzzy membership function, Fuzzy reference function, Similarity measure, Medical Diagnostic reasoning.

1. INTRODUCTION

Zadeh [1] introduced Fuzzy set theory in 1965. Since Zadeh initiated Fuzzy sets, many approaches and theories treating imprecision and uncertainty have been proposed. Different researchers have proposed different Similarity measures for Fuzzy sets, all based on Zadehian concept. Zadeh defined Fuzzy set in the manner where it has been accepted that the classical set theoretic axioms of exclusion and contradiction are not satisfied for Fuzzy sets. In Zadehian theory of Fuzzy set, it has been believed that for a Fuzzy set A and its complement A^c , neither $A \cap A^c$ is null set nor $A \cup A^c$ is the universal set which is proved to be wrong by Baruah[2,3]. Baruah proposed that in the Zadehian definition of the complement of a Fuzzy set, Fuzzy membership function and Fuzzy membership value had been taken to be the same, which leads to the conclusion that the Fuzzy sets do not follow the set theoretic axioms of exclusion and contradiction. Therefore, Baruah has put forward an extended definition of Fuzzy set and reintroduced the notion of complement of a Fuzzy set in a way that the set theoretic axioms of exclusion and contradiction can be seen valid for Fuzzy sets also. According to Baruah, to define a Fuzzy set two functions namely- Fuzzy membership function and -Fuzzy reference function are necessary. Fuzzy membership value is the difference between Fuzzy membership function and Fuzzy reference function. Fuzzy membership function and Fuzzy membership value are two different things. Neog and Sut [4] have generalized the concept of complement of a Fuzzy set, introduced by Baruah[2,3], when the Fuzzy reference function is not zero and defined arbitrary Fuzzy union and intersection

extending the definition of Fuzzy sets given by Baruah [2, 3]. As a consequence of which, the previously existing similarity measures of Fuzzy set, based on the traditional Zadehian definition of complementation, have appeared illogical. Accordingly, we have established a new Similarity measure[10] for Fuzzy sets using the extended definition of complementation[2,3,4] based on reference function. Also, we have proved the validity[11] of the new Similarity measure[10] with the help of traditional Hamming Distance and Euclidean Distance measures. In this paper, our aim is to propose an algorithm based on the new Similarity measure[10] to support Medical Diagnostic reasoning. Also, we attempt to demonstrate the application of the proposed algorithm on some collected data and validate the results obtained with the help of Hamming Distance and Euclidean Distance measures.

The overall organization of this paper is as follows. Section 2 overview the new Similarity measure for Fuzzy sets based on extended definition of complementation. In section 3 we propose an algorithm based on the new Similarity measure to support Medical Diagnostic reasoning. In section 4 we show the application of the proposed algorithm on some collected data. We verify the results obtained in section 4 with the results obtained by Hamming distance and Euclidean distance measures in section 5. Finally, some conclusions are given in section 6.

2. THE NEW SIMILARITY MEASURE FOR FUZZY SETS

Let A and B be two elements belonging to a Fuzzy set (or sets). Now we can measure the similarity between A and B as below:

$$\text{Sim}(A, B) = \frac{I_{FS}(A, B)}{I_{FS}(A, B^c)} = \frac{a}{b} \quad (1)$$

where a is distance from $A(\mu_m, \mu_r, \mu_v)$ to $B(\mu_m, \mu_r, \mu_v)$ and b is a distance from $A(\mu_m, \mu_r, \mu_v)$ to $B^c(\mu_m, \mu_r, \mu_v)$ where μ_m, μ_r, μ_v are membership function, reference function and membership value respectively.

For this similarity measure, we have,

$$0 \leq \text{Sim}(A, B) \leq \alpha$$

Similarly we can calculate the Similarity between two Fuzzy sets:

Let A and B be two Fuzzy sets defined on the same set of universe of discourse. Now we can measure the similarity between A and B by assessing similarity of the corresponding elements belonging to A and B, as defined in the eqn (1).

Now using Baruah’s definition of Fuzzy set, for the Similarity measure of A and B, we can obtain the following 4 possibilities,

- A and B may be two exactly similar sets.
- or A and B^C may be two exactly similar sets.
- or A may be more similar to B than to B^C.
- or A may be more similar to B^C than to B.
- But A can never be similar to B and B^C together i.e. A=B=B^C is never possible according to the new definition of complementation of Fuzzy set [2, 3].

Therefore from the above analysis, for the Similarity measure of A and B, we can conclude four possible cases as follows:

- Case 1: Sim(A,B)=0 when A=B i.e. AB=0.
- Case 2: Sim(A,B)=∞ when A=B^C i.e. AB^C=0.
- Case 3: Sim(A,B) > 1 when AB > AB^C.
- Case 4: Sim(A,B) < 1 when AB < AB^C.

Hence to measure the similarity between the two Fuzzy sets A and B, one should be interested in the values $0 \leq \text{Sim}(A,B) < 1$.

Let us explain the above idea for the new Similarity measure into details:

Let A and B be two Fuzzy sets defined on the same set of universe of discourse $U = \{e_1, e_2, e_3, e_4, e_5\}$. Now we can calculate the similarity measure for A and B assessing the similarity measure for the every corresponding elements of A and B i.e. for the every element e_1, e_2, e_3, e_4, e_5 of the set of universe of discourse U, considered for A and B. This means similarity measure for A and B has to be calculated with respect to every $e_1, e_2, e_3, e_4, e_5 \in U$.

Now, based on the new definition of Fuzzy set, the similarity measure for the Fuzzy set A($e_k, k=1,2,3,4,5$) and the Fuzzy set B($e_k, k=1,2,3,4,5$) can be obtained under the 3 possible cases in the following manner:

We can visualize the Fuzzy set A(e_k) and the Fuzzy set B(e_k) in the number line in Figure 1 and Figure 2 respectively.

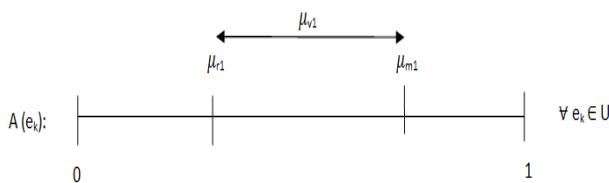


Figure 1. Representation of Fuzzy set A(e_k) in number line

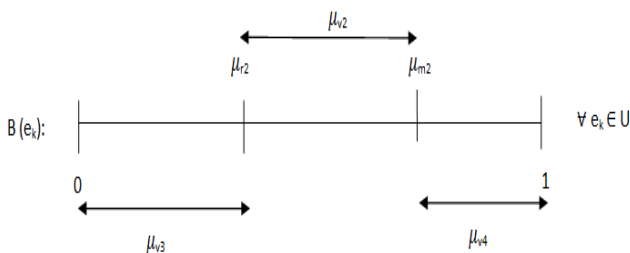


Figure 2. Representation of Fuzzy set B(e_k) in number line

Where $\mu_{r1}, \mu_{m1}, \mu_{v1}; \mu_{r2}, \mu_{m2}, \mu_{v2}; 0, \mu_{r2}, \mu_{v3}; \mu_{m2}, 1, \mu_{v4}$ are reference function, membership function and membership value of the Fuzzy set A, the Fuzzy set B and the two complement sets of B respectively for every $e_k \in U$.

Now the 3 possible cases are:

Case 1: when $\mu_{r2} \neq 0, \mu_{m2} \neq 1$.

Case 1 can be visualized in Figure 1 and Figure 2 and Similarity Measure can be defined as,

$$\frac{AB}{AB^C} = \frac{\text{Sim}(A, B)}{\frac{1}{5} \sum_{k=1}^5 \left(\frac{|A(s_k(\mu_{r1})) - B(s_k(\mu_{r2}))| + |A(s_k(\mu_{m1})) - B(s_k(\mu_{m2}))| + |A(s_k(\mu_{v1})) - B(s_k(\mu_{v2}))|}{|A(s_k(\mu_{r1})) - 0| + |A(s_k(\mu_{r1})) - B(s_k(\mu_{m2}))| + |A(s_k(\mu_{m1})) - B(s_k(\mu_{r2}))| + |A(s_k(\mu_{m1})) - 1| + |A(s_k(\mu_{r1})) - B(s_k(\mu_{v3}))| + |A(s_k(\mu_{v1})) - B(s_k(\mu_{v4}))|} \right)}$$

Case 2: when $\mu_{r2} = 0, \mu_{m2} \neq 1$.

Case 2 can be visualized in Figure 3 and Figure 4.

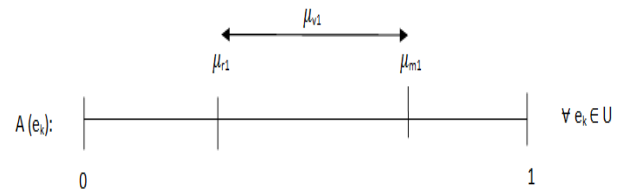


Figure 3. Representation of Fuzzy set A(e_k) in number line

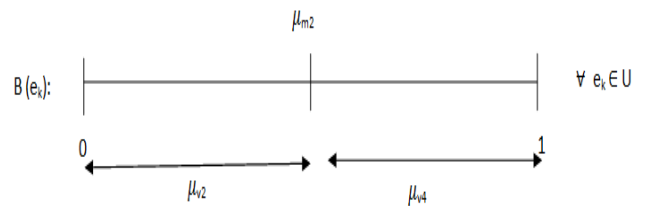


Figure 4. Representation of Fuzzy set B(e_k) in number line

and Similarity Measure can be defined as,

$$\frac{AB}{AB^C} = \frac{\text{Sim}(A, B)}{\frac{1}{5} \sum_{k=1}^5 \left(\frac{|A(s_k(\mu_{r1})) - 0| + |A(s_k(\mu_{m1})) - B(s_k(\mu_{m2}))| + |A(s_k(\mu_{v1})) - B(s_k(\mu_{v2}))|}{|A(s_k(\mu_{r1})) - B(s_k(\mu_{m2}))| + |A(s_k(\mu_{m1})) - 1| + |A(s_k(\mu_{r1})) - B(s_k(\mu_{v3}))|} \right)}$$

Case 3: when $\mu_{r2} \neq 0, \mu_{m2} = 1$.

Case 3 can be visualized in Figure 5 and Figure 6.

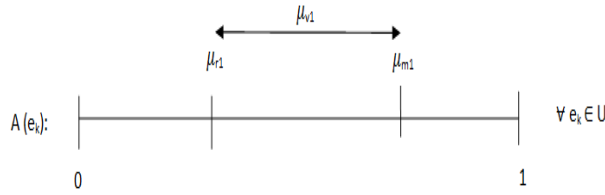


Figure 5. Representation of Fuzzy set $A(e_k)$ in number line

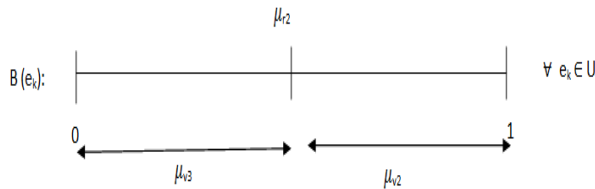


Figure 6. Representation of Fuzzy set $B(e_k)$ in number line

and Similarity Measure can be defined as,

$$\frac{AB}{AB^c} =$$

$$Sim(A, B) =$$

$$\frac{1}{5} \sum_{k=1}^5 \frac{(|A(S_k(\mu_{r1})) - B(S_k(\mu_{r2}))| + |A(S_k(\mu_{m1})) - 1| + |A(S_k(\mu_{v1})) - B(S_k(\mu_{v2}))|)}{(|A(S_k(\mu_{r1})) - 0| + |A(S_k(\mu_{m1})) - B(S_k(\mu_{r2}))| + |A(S_k(\mu_{v1})) - B(S_k(\mu_{v3}))|)}$$

3. ALGORITHM FOR MEDICAL DIAGNOSTIC REASONING

We design an algorithm on the basis of the new Similarity Measure discussed in section 2 in the following manner:

Step 1: Input Fuzzy sets P_i (for Patients) and D_j (for Diseases)
 $\forall i, j \in Z^+$.

Step 2: For each P_i (Patient Set)
Take each D_j (Disease Set)
and consider every S_k (Symptom characteristics),
 $k \in Z^+$.

To calculate: $Sim(P_i, D_j) = \frac{P_i D_i}{P_i D_j^c}$
 $\forall S_k \in S$ (Symptom Set)

Step 3: For each P_i

Find the smallest $\frac{P_i D_i}{P_i D_j^c} \forall D_j$

To determine the disease for the patient.

4. APPLICATION OF THE PROPOSED ALGORITHM

Now, we apply our proposed algorithm on some collected data[6].

Taking the lower bound value[6] as reference function value, μ_r and upper bound value[6] as membership function value, μ_v (since, always $0 \leq \mu_r(x) \leq \mu_v(x) \leq 1$), we can represent the collected data[6] as

dataset1= Disease Sets = {Fever(D_1), Malaria(D_2), Typhoid(D_3), Stomach Problem(D_4), Chest Problem(D_5)} and

dataset 2= Patient Sets={ P_1, P_2, P_3, P_4 }

defined on the set of universe of discourse, S ={Chest pain(C_p), Cough(C_o), Stomach pain(S_p), Headache(H_d), Temperature(T_r)} i.e. the set of symptoms.

Now we put the disease sets values and the patient sets values in table 1 and table 2 respectively.

Table 1: dataset 1 of Disease Sets

	D_1	D_2	D_3	D_4	D_5
C_p	(0.3,0.1)	(0.2,0.1)	(0.1,0.1)	(0.3,0.2)	(0.9,0.8)
C_o	(0.7,0.4)	(1.0,0.7)	(0.4,0.2)	(0.3,0.2)	(0.2,0.2)
S_p	(0.3,0.1)	(0.1,0.0)	(0.3,0.2)	(1.0,0.8)	(0.2,0.2)
H_d	(0.5,0.3)	(0.4,0.2)	(0.9,0.6)	(0.6,0.2)	(0.2,0.0)
T_r	(1.0,0.4)	(1.0,0.7)	(0.7,0.3)	(0.3,0.1)	(0.2,0.1)

Table 2: dataset 2 of Patient Sets

	P_1	P_2	P_3	P_4
C_p	(0.4,0.1)	(0.2,0.1)	(0.5,0.0)	(0.6,0.3)
C_o	(0.9,0.6)	(0.3,0.1)	(0.3,0.2)	(0.8,0.7)
S_p	(0.2,0.2)	(0.9,0.6)	(0.4,0.0)	(0.6,0.3)
H_d	(0.9,0.6)	(0.6,0.4)	(0.9,0.8)	(0.6,0.5)
T_r	(0.9,0.8)	(0.2,0.0)	(0.9,0.8)	(0.9,0.6)

Each symptom in Table1 and Table2 is described by: a reference function and a membership function value.

Now we derive a diagnosis for each patient P_i , $i=1,2,3,4$ using the new similarity measure .

1. Now at first we calculate the similarity measure $\frac{P_i D_i}{P_i D_j^c}$ for each patient P_i , $i=1,2,3,4$ with every disease D_j , $j=1,2,3,4,5$ separately, assessing the similarity measure for the every corresponding elements of the two sets i.e. the symptom characteristics C_p, C_o, S_p, H_d, T_r .

2. Then we find out the smallest value from the obtained similarity measures between a patient and every disease, we considered in Step 1. From that value we can decide from which disease the patient is suffering from.

Now we calculate the similarity measure values between the dataset of diseases and the dataset of patients and represent the calculated values in table 3.

Table 3: Similarity measure values for Each Patient with every Disease

	D_1	D_2	D_3	D_4	D_5
P_1	0.35	0.27	0.29	0.56	0.59
P_2	1.04	0.75	0.33	0.19	0.42
P_3	0.39	0.56	0.35	0.71	0.57
P_4	0.34	0.39	0.40	0.51	0.51

From Table 3 we can conclude:

Patient P_1 is suffering from disease D_2 , patient P_2 is suffering from disease D_4 , Patient P_3 is suffering from disease D_3 , Patient P_4 is suffering from disease D_1 .

5. VERIFICATION

Now we can verify the results of Table 3 using traditional Hamming Distance and Euclidean Distance measures.

5.1 Hamming distance measure

Now we derive a diagnosis for each patient P_i , $i=1,2,3,4$ using the Hamming Distance measure .

1. Now at first we calculate the Hamming Distance measure for each patient P_i , $i=1,2,3,4$ with every disease D_j , $j=1,2,3,4,5$ separately ,assessing the measure for the every corresponding elements of the two sets i.e. the symptom characteristics C_p, C_o, S_p, H_d, T_r .
2. Then we find out the smallest value from the obtained measures between a patient and every disease, we considered in Step 1. From that value we can decide from which disease the patient is suffering from.

The normalized Hamming distance of the set P_i from the set D_j , for all the symptom characteristics $C_p, C_o, S_p, H_d, T_r \in S$ is

$$HD(P_i, D_j) = \frac{1}{10} \sum_{k=1}^5 (|P_i(s_k(\mu_{r1})) - D_j(s_k(\mu_{r2}))| + |P_i(s_k(\mu_{m1})) - D_j(s_k(\mu_{m2}))| + |P_i(s_k(\mu_{v1})) - D_j(s_k(\mu_{v2}))|)$$

Table 4: Hamming Distance measures for Each Patient with every Disease

	D ₁	D ₂	D ₃	D ₄	D ₅
P ₁	0.28	0.24	0.28	0.54	0.56
P ₂	0.40	0.50	0.31	0.14	0.42
P ₃	0.38	0.44	0.32	0.50	0.55
P ₄	0.28	0.30	0.38	0.44	0.54

From table 4 we can conclude that

Patient P_1 is suffering from disease D_2 , patient P_2 is suffering from disease D_4 , Patient P_3 is suffering from disease D_3 , Patient P_4 is suffering from disease D_1 .

5.2 Euclidean distance measure

Now we derive a diagnosis for each patient P_i , $i=1,2,3,4$ using the Euclidean Distance measure .

1. Now at first we calculate the Euclidean Distance measure for each patient P_i , $i=1,2,3,4$ with every disease D_j , $j=1,2,3,4,5$ separately ,assessing the measure for the every corresponding elements of the two sets i.e. the symptom characteristics C_p, C_o, S_p, H_d, T_r .
2. Then we find out the smallest value from the obtained measures between a patient and every disease, we considered in Step 1. From that value we can decide from which disease the patient is suffering from.

The normalized Euclidean Distance measure of the set P_i from the set D_j , for all the symptom characteristics $C_p, C_o, S_p, H_d, T_r \in S$ is

$$ED(P_i, D_j) = \frac{1}{5} \sum_{k=1}^5 \sqrt{(|P_i(s_k(\mu_{r1})) - D_j(s_k(\mu_{r2}))|^2 + |P_i(s_k(\mu_{m1})) - D_j(s_k(\mu_{m2}))|^2 + |P_i(s_k(\mu_{v1})) - D_j(s_k(\mu_{v2}))|^2)}$$

Table 5: Euclidean Distance measures for Each Patient with every Disease

	D ₁	D ₂	D ₃	D ₄	D ₅
P ₁	0.36	0.30	0.36	0.68	0.72
P ₂	0.50	0.65	0.40	0.18	0.55
P ₃	0.47	0.56	0.40	0.62	0.72
P ₄	0.34	0.37	0.47	0.58	0.68

From table 5 we conclude that

Patient P_1 is suffering from disease D_2 , patient P_2 is suffering from disease D_4 , Patient P_3 is suffering from disease D_3 , Patient P_4 is suffering from disease D_1 .

Therefore from the results obtained from table 3, table 4 and table 5 we can state that the results found in the above Medical Diagnosis using the proposed algorithm are absolutely valid with respect to the traditional distance measures.

6. CONCLUSION

In this article, we have focused on the new Similarity measure for Fuzzy sets based on the extended definition of complementation using reference function where it is believed that Fuzzy membership function and Fuzzy membership value for the complement of a Fuzzy set are two different things. An algorithm based on the extended definition of complementation has been proposed to support Medical Diagnostic reasoning. Also we have demonstrated the application of the proposed algorithm on some collected data and verified the results obtained with the help of traditional Hamming Distance and Euclidean Distance measures. Finally, it has been observed that the proposed algorithm is clearly effective in supporting Medical Diagnostic reasoning.

7. ACKNOWLEDGMENT

The author would like to thank Hemanta K. Baruah, Professor, Department of Statistics, Gauhati University, for his valuable suggestions and guidance, in preparing this article.

8. REFERENCES

- [1] L.A.Zadeh, “ *Fuzzy Sets* ”, Information and Control, 8, pp. 338-353, 1965.
- [2] Hemanta K. Baruah, “ *Towards Forming A Field Of Fuzzy Sets* ” International Journal of Energy, Information and Communications, Vol.2, Issue 1, pp. 16-20, February 2011.
- [3] Hemanta K. Baruah, “ *The Theory of Fuzzy Sets: Beliefs and Realities* ”, International Journal of Energy, Information and Communications, Vol. 2, Issue 2, pp. 1-22, May, 2011.
- [4] Tridiv Jyoti Neog , Dushmantha Kumar Sut, “ *Theory of Fuzzy Soft Sets from a New Perspective* ”, International Journal of Latest Trends in Computing, Vol. 2, No 3, September, 2011.
- [5] Eulalia Szmidt, Janusz Kacprzyk, “ *Medical Diagnostic Reasoning Using a Similarity Measure for Intuitionistic Fuzzy Sets* ” Eighth Int. Conf. on IFSS, Varna, 20-21 June 2004, NIFS Vol. 10 (2004), 4, 61-69.

- [6] Hong-mei Ju , Feng-Ying Wang , “A Similarity Measure for Interval –valued Fuzzy Sets and Its Application in Supporting Medical Diagnostic Reasoning”, The Tenth International Symposium on Operation Research and Its Applications (ISORA 2011), Dunhuang, China, August 28-31, 2011.
- [7] Szmids E., Kacprzyk J. (2001), “Entropy for intuitionistic Fuzzy sets”. Fuzzy Sets and Systems, vol. 118, No. 3, pp. 467-477.
- [8] Hongmei Ju (2008). “Entropy for Interval-valued Fuzzy Sets”, Fuzzy information and engineering, Volume 1, 358-366.
- [9] Pranamika Kakati (2013). “A Study on Similarity Measure for Fuzzy Sets” International Journal of Advanced Research in Computer Science and Software Engineering, vol. 3 Issue 8, pp.97-103, August 2013.
- [10] Pranamika Kakati (2013) “A New Similarity Measure for Fuzzy Sets with the Extended Definition of Complementatation” International Journal of Soft Computing and Engineering, Volume 3 Issue 4, pp.203-207, September, 2013.
- [11] Pranamika Kakati (2013) “A Note on the New Similarity Measure for Fuzzy sets” International Journal of Computer Applications Technology and Research, Volume 2 Issue 5, pp.601-605, September-October, 2013.
- [12] Pranamika Kakati (2013) “ A New Architecture of Simplified Fuzzy ARTMAP with the Extended Definition of Complementatation” International Journal of Computer Technology and Applications, Volume 4 Issue 5, pp. 772-784, September-October, 2013.