A Note on the New Similarity Measure for Fuzzy sets

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

Abstract: In this article, we intend to draw attention on the new Similarity measure for Fuzzy sets based on the extended definition of complementation. The old existing measures are based on traditional Zadehian Theory of Fuzzy sets where it is believed that there is no difference between Fuzzy membership function and Fuzzy membership value for the complement of a Fuzzy set which is already proved to be wrong. As a result, the previous Similarity measures have been proved illogical from the standpoints of new definition of complementation of Fuzzy set based on the fact that Fuzzy membership function and Fuzzy membership value for the complement of a Fuzzy set are two different things. Accordingly, we have already established a new Similarity measure with the help of extended definition of complementation using reference function. In this paper, an effort has been put forward to show the validity of the results obtained from our proposed measure with the help of traditional Hamming distance and Euclidean distance measures.

Keywords: Complement of a Fuzzy set, Fuzzy membership function, Fuzzy membership value, Fuzzy reference function, Fuzzy set, Similarity measure.

1. INTRODUCTION

Zadeh [1] introduced Fuzzy set 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 believed that the classical set theoretic axioms of exclusion and contradiction are not satisfied for Fuzzy sets. Regarding this, Baruah [2,3] proposed that two functions, namely Fuzzy membership function and Fuzzy reference function are necessary to represent a Fuzzy set. As a result, Baruah [2, 3] 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. 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, which are based on the traditional Zadehian definition of complementation, have appeared illogical. Accordingly, we have proposed a new Similarity measure[10] for Fuzzy sets using the extended definition of complementation[2,3,4] based on reference function so that it becomes free from any further controversy. In this article, our purpose is to prove that the results obtained from the application of our proposed measure are absolutely valid with respect to traditional Hamming distance and Euclidean distance measures.

The overall organization of this paper is as follows. In section 2 we discuss the new Similarity measure for Fuzzy sets based on extended definition of complementation. In section 3 we apply the new Similarity measure to calculate similarity measures of some collected data. In section 4 we verify the results obtained in section 3 with the results obtained by using Hamming distance and Euclidean distance measures. Finally, some conclusions are given in section 5.

2. THE NEW SIMILARITY MEASURE FOR FUZZY SETS

The new Similarity Measure [10] for Fuzzy sets with the extended definition of complementation is as follows:

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:

$$\text{Sim (A, B)} = \frac{l_{FS}(A,B)}{l_{FS}(A,B^C)} \qquad = \frac{a}{b} \qquad (1)$$

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,

$$0 \leqq \text{Sim (A, B)} \leqq \propto$$

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 $0 \leqq$ 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.
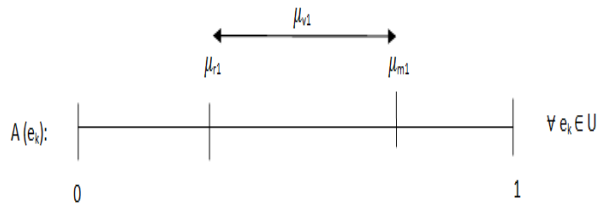

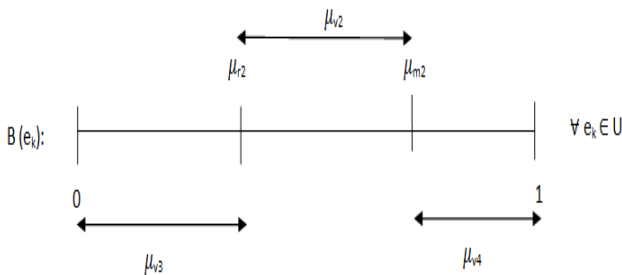**Figure 1. Representation of Fuzzy set A(e_k) in number line**

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

Where μ_{r1}, μ_{m1}, μ_{v1} ; μ_{r2}, μ_{m2}, μ_{v2} ; 0, μ_{r2}, μ_{v3} ; μ_{m2}, 1, μ_{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,

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

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

Case 2: when $\mu_{r2}=0$, $\mu_{m2} \neq 1$.
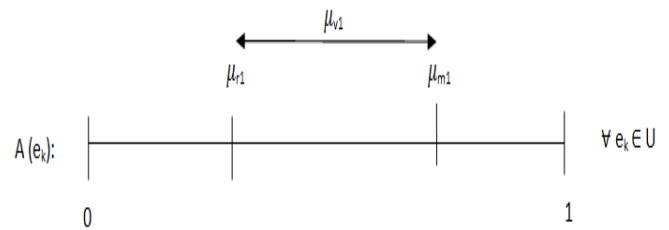
Case 2 can be visualized in Figure 3 and Figure 4.

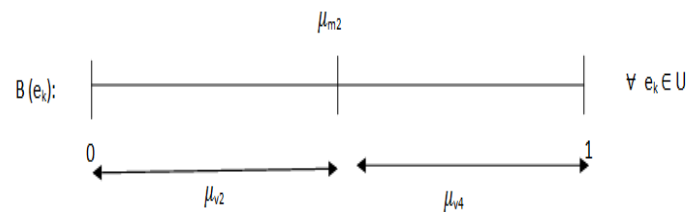**Figure 3. Representation of Fuzzy set A(e_k) in number line**

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

and Similarity Measure can be defined as,

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

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

Case 3: when $\mu_{r2} \neq 0$, $\mu_{m2}=1$.
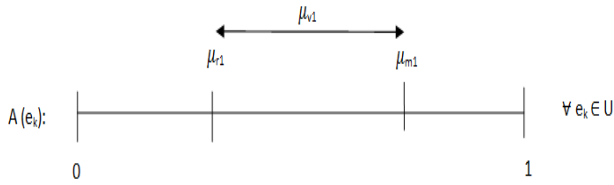
Case 3 can be visualized in Figure 5 and Figure 6.

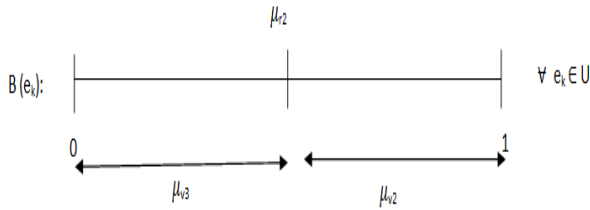**Figure 5. Representation of Fuzzy set A(e$_k$) in number line**

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

and Similarity Measure can be defined as,

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

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

3. APPLICATION OF THE NEW SIMILARITY MEASURE

We apply our proposed measure on some collected data[5].

Taking larger value (membership or non-membership function value) [5] as membership function value, μ_v and the smaller value[5] as reference function value, μ_r (since, always $0 \leq \mu_r(x) \leq \mu_v(x) \leq 1$), we can represent the collected data[5] as dataset1={A$_1$,A$_2$, A$_3$, A$_4$, A$_5$} and dataset2={B$_1$,B$_2$,B$_3$,B$_4$} defined on the same set of universe of discourse U={a, b, c, d, e} as given in Table 1 and Table 2.

Table 1: dataset 1 = {A₁, A₂, A₃, A₄, A₅}.

	A$_1$	A$_2$	A$_3$	A$_4$	A$_5$
a	(0.7,0.1)	(0.8,0.1)	(0.9,0.1)	(0.7,0.2)	(0.8,0.1)
b	(0.4,0.3)	(0.7,0.0)	(0.6,0.2)	(0.7,0.2)	(0.8,0.2)
c	(0.7,0.1)	(0.9,0.0)	(0.7,0.2)	(0.8,0.0)	(0.8,0.2)
d	(0.5,0.3)	(0.6,0.2)	(0.6,0.1)	(0.4,0.2)	(0.8,0.0)
e	(0.4,0.0)	(0.7,0.0)	(0.3,0.3)	(0.7,0.1)	(0.8,0.1)

Table 2: dataset 2 = {B₁, B₂, B₃, B₄}.

	B$_1$	B$_2$	B$_3$	B$_4$
a	(0.6,0.1)	(0.8,0.1)	(0.5,0.0)	(0.4,0.3)
b	(0.6,0.1)	(0.7,0.1)	(0.7,0.2)	(0.7,0.2)
c	(0.8,0.2)	(0.6,0.1)	(0.6,0.0)	(0.4,0.3)
d	(0.6,0.1)	(0.4,0.4)	(0.8,0.1)	(0.5,0.4)
e	(0.8,0.1)	(0.8,0.0)	(0.8,0.1)	(0.6,0.1)

Each element (a, b, c, d or e) ∈ U in Table1 and Table2 is described by: a reference function and a membership function value.

Now to calculate a similar set from the dataset 1 for a particular set in dataset 2, we proceed in the following way:

Step 1: At first we calculate the similarity measure $\frac{B_jA_i}{B_jA_i^C}$ for each set B$_j$ ∈ dataset 2,(where j=1,2,3,4) with every set A$_i$ ∈ dataset 1 ,where (i=1,2,3,4,5) separately ,assessing the similarity measure for the every corresponding elements of the two sets i.e. a,b,c,d,e ∈ U, the set of universe of discourse considered for the two datasets.

Step 2: Then we find out the smallest value from the obtained similarity measures between a set B$_j$ and every set A$_i$, we considered in Step 1.From that value we can decide which A$_i$ ∈ dataset 1 is similar to a particular set B$_j$ ∈ dataset 2 .

Now we calculate the similarity measure values between the dataset 1 and the dataset 2 and represent the calculated values in table 3.

Table 3: Similarity measure values between dataset 1 and dataset 2.

	A$_1$	A$_2$	A$_3$	A$_4$	A$_5$
B$_1$	0.49	0.23	0.28	0.19	0.15
B$_2$	0.41	0.24	0.37	0.20	0.33
B$_3$	0.54	0.28	0.35	0.21	0.12
B$_4$	0.45	0.61	0.49	0.38	0.57

Hence from table 3 we can conclude that,

Set B$_1$ is similar to set A$_5$, set B$_2$ is similar to set A$_4$, set B$_3$ is similar to set A$_5$ and set B$_4$ is similar to set A$_4$.

4. VERIFICATION

We can verify the results obtained in section 3 by using normalized Hamming Distance and Euclidean Distance measures.
For that purpose, we apply Hamming Distance measure and Euclidean Distance measure on the same collected data[5] , used in section 3.

4.1 Hamming Distance measure:

To measure similarity between dataset 1 and dataset 2 using Hamming Distance measure, we proceed in the following way:
Step 1: At first we calculate the Hamming Distance measure for each set B$_j$ ∈ dataset 2, (where j=1,2,3,4) with every set A$_i$ ∈ dataset1 ,(where i=1,2,3,4,5) separately ,assessing the

distance measure for the every corresponding elements of the two sets i.e. a,b,c,d,e ∈ U, the set of universe of discourse considered for the two datasets.

Step 2: Then we find out the smallest value from the obtained measures between a set B_j and every set A_i , we considered in Step 1. From that value we can decide which A_i ∈ dataset 1 is similar to a particular set B_j ∈ dataset 2 .

The normalized Hamming distance of the set A_i from the set B_j, for all a, b, c, d, e ∈ U is

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

Table 4: Hamming Distance measures between dataset 1 and dataset 2.

	A_1	A_2	A_3	A_4	A_5
B_1	0.26	0.18	0.24	0.16	0.14
B_2	0.26	0.20	0.34	0.20	0.24
B_3	0.32	0.24	0.32	0.20	0.14
B_4	0.34	0.42	0.42	0.28	0.42

From table 4 we conclude that

Set B_1 is similar to set A_5, set B_2 is similar to set A_2 and A_4, set B_3 is similar to set A_5 and set B_4 is similar to set A_4.

4.2 Euclidean Distance measure:

To measure similarity between dataset 1 and dataset 2, we proceed in the following way:

Step 1: At first we calculate the Euclidean Distance measure for each set B_j ∈ dataset 2, (where j=1,2,3,4) with every set A_i ∈ dataset1   ,(where i=1,2,3,4,5) separately ,assessing the distance measure for the every corresponding elements of the two sets i.e. a,b,c,d,e ∈ U, the set of universe of discourse considered for the two datasets.

Step 2: Then we find out the smallest value from the obtained measures between a set B_j and every set A_i , we considered in Step 1. From that value we can decide which A_i ∈ dataset 1 is similar to a particular set B_j ∈ dataset 2 .

The Euclidean distance of the set A_i from the set B_j, for all a, b, c, d, e ∈ U is

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

Table 5: Euclidean Distance measures between dataset 1 and dataset 2.

	A_1	A_2	A_3	A_4	A_5
B_1	0.32	0.23	0.31	0.21	0.17
B_2	0.33	0.25	0.42	0.25	0.29
B_3	0.39	0.31	0.40	0.26	0.18
B_4	0.42	0.52	0.52	0.34	0.53

From table 5 we conclude that

Set B_1 is similar to set A_5, set B_2 is similar to set A_2 and A_4, set B_3 is similar to set A_5 and set B_4 is similar to set A_4.

Therefore from table 4 and table 5, we have found that the set B_2 is similar to both the set A_2 and the set A_4. This is due to the limitation that both Hamming distance and Euclidean distance measures can calculate upto two decimal places only. Hence even if there is a fractional difference between two values after two decimal places, these two measures can not represent it. It has been observed that the results obtained from Hamming distance and Euclidean distance are exactly similar. Thus, we have seen that though there is no error in the results, the results obtained from these two traditional measures , sometimes, may not be clear which is not a problem at all in case of our proposed measure. From table 3, we can clearly find that set B_2 is actually similar to set A_4. Also, it has been observed that the results obtained from table 3, table 4 and table 5 are absolutely similar with just a little exception in the result for the set B_2 in case of the two traditional distance measures which determines that our proposed measure gives clearer result than the traditional measures. Therefore, it can be concluded that the results obtained by the application of the new Similarity measure are clear and valid with respect to the traditional Hamming distance and Euclidean distance measures and hence legally acceptable.

5. CONCLUSION

In this work, we have first gone through the new Similarity measure for Fuzzy sets based on the extended definition of complementation based on reference function. Again, the Similarity measure has been applied to evaluate some collected dataset. Also, the results obtained from the application are compared with the results found by Hamming distance and Euclidean distance measures. Finally it has been proved that the results obtained   from the application of the new Similarity measure for Fuzzy sets are much clear and absolutely valid with respect to the traditional distance measures.

6. ACKNOWLEDGEMENT

7. 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 , Dusmanta 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] Szmidt 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 Softaware 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 Complementation"International Journal of Soft Computing and Engineering, Volume 3 Issue 4, pp.203-207, September, 2013.