An improvised similarity measure for generalized fuzzy numbers

Daud Mohamad, Noorlisa Sara Adlene Ramlan, Sharifah Aniza Sayed Ahmad
Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia

ABSTRACT

Similarity measure between two fuzzy sets is an important tool for comparing various characteristics of the fuzzy sets. It is a preferred approach as compared to distance methods as the defuzzification process in obtaining the distance between fuzzy sets will incur loss of information. Many similarity measures have been introduced but most of them are not capable to discriminate certain type of fuzzy numbers. In this paper, an improvised similarity measure for generalized fuzzy numbers that incorporate several essential features is proposed. The features under consideration are geometric mean averaging, Hausdorff distance, distance between elements, distance between center of gravity and the Jaccard index. The new similarity measure is validated using some benchmark sample sets. The proposed similarity measure is found to be consistent with other existing methods with an advantage of able to solve some discriminant problems that other methods cannot. Analysis of the advantages of the improvised similarity measure is presented and discussed. The proposed similarity measure can be incorporated in decision making procedure with fuzzy environment for ranking purposes.

Keywords: Decision making, Generalized fuzzy numbers, Geometric mean averaging, Hausdorff distance, Similarity measure

1. INTRODUCTION

The concept of similarity measure has been an essential element to be concerned as it is very useful in aggregating opinions and for comparison between fuzzy evaluation. Similarity measure effectively computes the degree of similarity of features of distinct generalized fuzzy numbers [1, 2]. It has been a better choice than using the defuzzification method in particular for comparing fuzzy numbers as it will preserve the fuzzy form of the fuzzy numbers. Numerous formulas have been proposed to evaluate the similarity between two fuzzy numbers from various perspectives. A distance based similarity measure between fuzzy numbers was introduced by [3] to aggregate the decision makers’ opinions. Later, [4] implemented similarity measure on trapezoidal fuzzy numbers in dealing with fuzzy opinions for group decision making by developing the formula using the distance of trapezoidal fuzzy numbers based on $l_p$ metric and the difference between minimum and maximum universe of discourse. A similarity measure was presented in [5] based on the distance of graded mean integration representation of trapezoidal and triangular fuzzy numbers which is useful to calculate disjoint fuzzy sets.

A new similarity measure was proposed in [2] by utilizing the Simple Center of Gravity Method (SCGM) for evaluating the Center of Gravity (COG) point of a generalized fuzzy number. It considers the confidence level of decision makers’ opinions in the evaluation and similarity comparison. Realizing there is a drawback in the similarity measure of [2], where the formula is not valid for singleton, [6] introduced a
similarity measure based on the radius of gyration (ROG) points. This is to solve the problem in COG based similarity measure and applied the method to a pattern recognition problem. Furthermore, [7] has improvised the formula of the COG distance in [2] by including a term that determines the calculation of spread of fuzzy numbers. In addition, [8] pointed out that there is a disadvantage in the similarity measure of [2] in determining the similarity measure between generalized fuzzy numbers with the same COG points. Thus [8] proved that the ROG produced a better result compared to the COG.

A similarity measure was presented in [9] by observing the distance of points between fuzzy numbers, meanwhile in [10] a similarity measure formula based on standard deviation operator for interval trapezoidal fuzzy numbers was proposed. The shape and spreads of the membership functions of fuzzy numbers including their relative distance are computed and compared. [11] made further improvement in COG based similarity measure introduced in [2] by incorporating an additional feature of Hausdorff distance. The characteristic of Hausdorff distance which is a nonlinear operator enables to measure the mismatch of two fuzzy numbers. In [12] a similarity measure was suggested based on the Jaccard index where it is useful to compare the similarity and diversity of sample sets and it has an advantage of its simplicity of calculation. Recently, a similarity measure based on left and right apex angles is proposed by [13]. In this method, the features of the COG, area, perimeter and height of generalized fuzzy numbers are included in the similarity measure. At present, there are vast applications of similarity measure in solving ranking and decision making problems [14-18].

A new similarity is proposed in this paper to overcome some of pertaining problems in comparing generalized fuzzy numbers, in particular the discriminant problem of certain types of fuzzy numbers. The new similarity measure has five important features namely the geometric mean averaging, Hausdorff distance, distance between elements, distance between center of gravity and the Jaccard index. The proposed similarity is then tested and validated on some benchmarking sets. An analysis and discussion is given for some sets under consideration that highlight advantages of the proposed similarity measure.

2. THE PROPOSED SIMILARITY MEASURE

We present some basic definitions and concepts that will be used in deriving the new similarity measure. A fuzzy set $A$ is represented by a set of ordered pairs,

$$A = \{(x, \mu_A(x)) | x \in A, \mu_A(x) \in [0,1] \}$$

where $\mu_A(x)$ is a membership function that specifies the degree to which $x$ belongs to $A$. A fuzzy number is a fuzzy set that is normal and convex where normal refers to a fuzzy set with height 1 and convex is when the fuzzy set satisfies

$$\mu_A(\lambda x + (1-\lambda)y) \geq \min(\mu_A(x), \mu_A(y))$$

for $(x,y)$ in $A$ and $\lambda \in [0,1]$. A generalized trapezoidal fuzzy number $A$ is represented as a 4-tuples $A=(a, b, c, d)$ with the membership function $\mu_A(x)$ is defined as

$$\mu_A(x) = \begin{cases} 
0 & , x < a \\
\frac{x-a}{b-a}, & a \leq x \leq b \\
\frac{w}{b-a}, & b < x \leq c \\
\frac{d-x}{d-c}, & c \leq x \leq d \\
0 & , x > d 
\end{cases}$$

where $w\in[0,1]$. When $w=1$, $A$ is just a standard trapezoidal fuzzy number and when $b=c$, it is a triangular fuzzy number.

The proposed similarity measure is an improvised version of the similarity measure introduced by [7]. In [7], only the three features of distance between elements, distance between centroid and the Jaccard index are included. However, this similarity has weaknesses, in particular unable to discriminate some fuzzy numbers with similar shape. [19] has proposed a new averaging operator which is known as Geometric Means Averaging (GMA) operators which is originated from the geometric mean formula to deal with the
drawbacks found in previous operators where it is unable to differentiate the similarity between fuzzy sets that have exactly or almost similar shape. The GMA operator is useful for dealing with intersection and union operations with fuzzy information. In addition, [20] highlighted the usefulness of the GMA in handling situations compromise of fuzzy aggregating and fuzzy decision making problems. Furthermore, it is more flexible to be used in handling MCDM problems [21]. The GMA formula [21] between two generalized trapezoidal fuzzy numbers \( \tilde{A} \) and \( \tilde{B} \) is defined as

\[
S(\tilde{A}, \tilde{B}) = \left( \frac{\sqrt[d]{\prod_{i=1}^{d} (2 - |a_i - b_i|)}}{2} - 1 \right)
\]

The Hausdorff distance is a measurement of how closed and bounded the sets of \( \tilde{A} \) and \( \tilde{B} \) between them [22] and is represented as;

\[
H(\tilde{A}, \tilde{B}) = \max \left\{ H^*(\tilde{A}, \tilde{B}), H^*(\tilde{B}, \tilde{A}) \right\}
= \max \left\{ |a_1 - b_1|, |a_2 - b_2| \right\}
\]

Evidently, \( H(\tilde{A}, \tilde{B}) \) is the maximal distance between two \( \alpha \) -cuts of set \( \tilde{A} \) and \( \tilde{B} \) whereas the \( H^*(\tilde{A}, \tilde{B}) \) and \( H^*(\tilde{B}, \tilde{A}) \) is the directed distance between \( \tilde{A} \) and \( \tilde{B} \).

Considering the the advantages of the GMA and Hausdorff distance, these two features are incorporated in the similarity measure introduced by [7] as an improved similarity measure of this paper. The GMA evidently can discriminate two similar generalized fuzzy number. The geometry information of the two generalized fuzzy numbers is obtained by considering the geometry shape of the membership functions using the Hausdorff distance. The improvised similarity measure is as follows. Let \( \tilde{A} = (a_1, a_2, a_3, a_4; w_A) \) and \( \tilde{B} = (b_1, b_2, b_3, b_4; w_B) \) be two generalized fuzzy numbers. The similarity measure between \( \tilde{A} \) and \( \tilde{B} \) is given as;

\[
S(\tilde{A}, \tilde{B}) = \left( \frac{\sqrt[d]{\prod_{i=1}^{d} (2 - |a_i - b_i|)}}{2} - 1 \right) \times \left( 1 + \left( H(\tilde{A}, \tilde{B}) + |w_A - w_B| \right) \right)^{-t} \times \left[ \left( 1 - \frac{1}{4} \sum_{i=1}^{4} |a_i - b_i| \right) \times \left( 1 - |\hat{x}_A - \hat{x}_B| \right) \right]^{\frac{1}{2}} \\
\times \left( \frac{\min(\hat{y}_A, \hat{y}_B)}{\max(\hat{y}_A, \hat{y}_B)} \right)
\]

where \( S(\tilde{A}, \tilde{B}) \in [0,1], t \in [0, \infty] \). The parameter \( t \) indicates the level of importance of the similarity measure on the Hausdorff distance. The value of \( t \) can be defined based on the real situation requirements. The similarity measure between fuzzy numbers is considered high when the value of \( S(\tilde{A}, \tilde{B}) \) is large.

It is obvious that the proposed similarity satisfies the following basic similarity measure properties:

i) Two generalized fuzzy numbers \( \tilde{A} = (a_1, a_2, a_3, a_4; w_A) \) and \( \tilde{B} = (b_1, b_2, b_3, b_4; w_B) \) are identical if and only if \( S(\tilde{A}, \tilde{B}) = 1 \).

ii) \( S(\tilde{A}, \tilde{B}) = S(\tilde{B}, \tilde{A}) \)

iii) If \( \tilde{A} = (a, a, a, a; 1) \) and \( \tilde{B} = (b, b, b, b; 1) \) are real numbers, then \( S(\tilde{A}, \tilde{B}) = \frac{\left( 1 - |a - b| \right)^{\frac{1}{2}}}{1 + |a - b|} \).

We shall prove iii) as the proofs of i) and ii) are straightforward.
Proof of iii)

We have \( \hat{y}_A = \hat{y}_B = \frac{1}{2} \) and thus

\[
S(\hat{A}, \hat{B}) = \left( \left( \prod_{i=1}^{n} \left( 2 - |a_i - b_i| \right) - 1 \right) \times \left( 1 + \left( H(\hat{A}, \hat{B}) + |w_A - w_B| \right) \right) \right)^{-1/2} \\
\times \left( 1 - \frac{4}{3} \sum_{i=1}^{n} |a_i - b_i| \times (1 - |a_i - b_i|) \right)^{1/2} \times \left( \frac{\min \left( \hat{y}_A, \hat{y}_B \right)}{\max \left( \hat{y}_A, \hat{y}_B \right)} \right)^{1/2} \\
\times \left( 1 - \frac{4}{3} \sum_{i=1}^{n} |a_i - b_i| \times |1 - |a_i - b_i|| \times \left( 1 - |a - b| \right) \right)^{1/2} \\
= \left( 1 - \frac{4}{3} \sum_{i=1}^{n} |a_i - b_i| \times (1 - |a_i - b_i|) \right)^{1/2} \times \left( 1 - |a - b| \right)^{1/2} \\
= \left( 1 - \frac{4}{3} \sum_{i=1}^{n} |a_i - b_i| \times (1 - |a_i - b_i|) \right)^{1/2} \times \left( 1 - |a - b| \right)^{1/2} \\
= \left( 1 - \frac{4}{3} \sum_{i=1}^{n} |a_i - b_i| \times (1 - |a_i - b_i|) \right)^{1/2} \times \left( 1 - |a - b| \right)^{1/2}.
\]

3. COMPARATIVE ANALYSIS

The data of 11 sets of fuzzy numbers in [23] are used to compare the similarity values of the proposed similarity measure with the existing similarity measures. These sets are commonly used by many [24-26] as a comparison and benchmarking tool. The similarity values obtained is as in Table 1.

The proposed similarity measure expectedly gives a consistent value as other similarity measures for Set 2 and 4. The similarity values of Set 1, Set 3 and Set 5 are slightly lower due to the effect of the extra features incorporated in the proposed similarity measure. A comparative analysis is made on the other six sets which are the pair of Set 6 and Set 7, the pair of Set 8 and Set 9 and the pair of Set 10 and Set 11. They are chosen for discussion due to their high resemblance of the fuzzy numbers’ shape. Thus, it will be clearer on how the geometric mean averaging and the Hausdorff distance will affect the similarity values between the generalized fuzzy numbers. We first investigate the similarity value of the pair of Set 6 and Set 7. The illustration of the Set 6 and Set 7 is given in Figure 1.

![Figure 1. Illustration of Set 6 and Set 7](image_url)

The Set 7 intuitively has a higher similarity compared to Set 6 due to high resemblance of shapes of two generalized fuzzy numbers in Set 7 as compared to in Set 6 even though both sets have the same relative distance between the generalized fuzzy numbers. The proposed method yields a slightly higher similarity measure for Set 7 (0.38) compared to Set 6 (0.35). Referring to the similarity value obtained from the existing similarity measure formula, most methods unable to clearly discriminate the degree of similarity of generalized fuzzy numbers when they have exactly the same or similar shapes. Furthermore, methods of [4] and [6] yield a contrast result in which the similarity value for Set 6 is higher than Set 7. Thus, the geometric mean averaging can specify the value of the degree of similarity for exact similar shapes between two sets of generalized fuzzy numbers. Set 8 and 9 are also compared. The shapes of generalized fuzzy numbers of Set 8

An improvised similarity measure for generalized fuzzy numbers (Daud Mohamad)
and Set 9 have some common similarities that worth for investigation and analysis. Figure 2 shows the illustration of the two sets.

Table 1. Comparison of the similarity values obtained from the existing and proposed similarity measure

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Set (1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tilde{A} = (0.10,0.20,0.30,0.40,1.00)$</td>
<td>0.98</td>
<td>0.97</td>
<td>1.00</td>
<td>0.84</td>
<td>0.80</td>
<td>0.85</td>
<td>0.84</td>
<td>0.79</td>
</tr>
<tr>
<td>$\tilde{B} = (0.10,0.25,0.25,0.40,1.00)$</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Set (2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tilde{A} = (0.10,0.20,0.30,0.40,1.00)$</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
<td>0.67</td>
</tr>
<tr>
<td>$\tilde{B} = (0.10,0.20,0.30,0.40,0.80)$</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Set (3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tilde{A} = (0.30,0.30,0.30,0.30,1.00)$</td>
<td>0.90</td>
<td>0.90</td>
<td>0.90</td>
<td>0.81</td>
<td>0.95</td>
<td>0.90</td>
<td>0.74</td>
<td></td>
</tr>
<tr>
<td>$\tilde{B} = (0.30,0.30,0.30,0.30,1.00)$</td>
<td>0.70</td>
<td>0.40</td>
<td>0.77</td>
<td>0.49</td>
<td>0.90</td>
<td>0.84</td>
<td>0.70</td>
<td>0.35</td>
</tr>
<tr>
<td>Set (5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tilde{A} = (0.20,0.20,0.20,0.20,1.00)$</td>
<td>0.70</td>
<td>0.25</td>
<td>0.77</td>
<td>0.49</td>
<td>0.78</td>
<td>0.84</td>
<td>0.70</td>
<td>0.38</td>
</tr>
<tr>
<td>$\tilde{B} = (0.30,0.30,0.30,0.30,1.00)$</td>
<td>0.70</td>
<td>0.50</td>
<td>0.77</td>
<td>0.49</td>
<td>0.49</td>
<td>0.49</td>
<td>0.70</td>
<td>0.38</td>
</tr>
<tr>
<td>Set (6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tilde{A} = (0.30,0.30,0.40,0.40,w)$</td>
<td>0.70</td>
<td>0.64</td>
<td>0.77</td>
<td>0.49</td>
<td>0.70</td>
<td>0.84</td>
<td>0.59</td>
<td>0.32</td>
</tr>
<tr>
<td>$\tilde{B} = (0.50,0.50,0.80,0.80,w)$</td>
<td>0.90</td>
<td>0.90</td>
<td>1.00</td>
<td>0.81</td>
<td>0.83</td>
<td>0.85</td>
<td>0.81</td>
<td>0.70</td>
</tr>
<tr>
<td>Set (9)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tilde{A} = (0.10,0.20,0.30,0.40,0.50)$</td>
<td>0.70</td>
<td>0.64</td>
<td>0.77</td>
<td>0.49</td>
<td>0.70</td>
<td>0.84</td>
<td>0.59</td>
<td>0.32</td>
</tr>
<tr>
<td>$\tilde{B} = (0.17,0.55,0.93,0.97)$</td>
<td>0.90</td>
<td>0.90</td>
<td>1.00</td>
<td>0.81</td>
<td>0.83</td>
<td>0.85</td>
<td>0.81</td>
<td>0.70</td>
</tr>
<tr>
<td>Set (11)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tilde{A} = (0.50,0.60,0.90,0.99)$</td>
<td>0.90</td>
<td>0.90</td>
<td>1.00</td>
<td>0.80</td>
<td>0.83</td>
<td>0.85</td>
<td>0.80</td>
<td>0.69</td>
</tr>
<tr>
<td>$\tilde{B} = (0.60,0.70,0.80,0.99)$</td>
<td>0.90</td>
<td>0.90</td>
<td>1.00</td>
<td>0.80</td>
<td>0.83</td>
<td>0.85</td>
<td>0.80</td>
<td>0.69</td>
</tr>
</tbody>
</table>

Note: * indicates that the method cannot calculate the degree of similarity between generalized fuzzy numbers

Figure 2. Illustration of Set 8 and Set 9
Apparently, the Set 8 has slightly higher similarity value (0.38) as compared to Set 9 (0.32) given by the proposed method. This is due to the significantly different shapes of the generalized fuzzy numbers in Set 9. Methods of [3, 5, 7], and [14] yield the same degree of similarity for Set 8 and Set 9. However, [4] and [6] produce contrast similarity values where Set 9 has higher similarity value than Set 8. The method of [3] is the only method that able to meet the expected similarity value as it contains the geometric mean averaging information. In this case, it also shows that by having the geometric mean averaging in the similarity measure formula, it helps to specify the value of the degree of similarity for exact similar shapes between the two generalized fuzzy numbers. Finally, a comparison is made between the Set 10 and set 11. The illustration of the two sets is given as in Figure 3.

![Figure 3. Illustration of Set 10 and Set 11](image)

Evidently, the shape of generalized fuzzy number $\tilde{A}$ of Set 10 is similar to the shape of generalized fuzzy number $\tilde{A}$ of Set 11 and their relative distance between generalized fuzzy numbers is equal. However, the height of generalized fuzzy number $\tilde{B}$ of Set 10 differs from the $\tilde{B}$ of Set 11. As the heights of the generalized fuzzy number $\tilde{B}$ in both sets are equally 1.00, the degree of similarity between generalized fuzzy numbers in Set 10 should be higher compared to Set 11. The proposed similarity measure gives a higher degree of similarity (0.70) to Set 10 as compared to Set 11 (0.69) and this is in agreement with the method of [7] and [18]. This indicates that the improvised similarity measure can discriminate the valuation of similar shapes of the fuzzy numbers in the two sets with different height of one of the generalized fuzzy number of those two sets.

4. CONCLUSION

We have improved the similarity measure introduced by [7] by adding the features of geometric mean averaging and Hausdorff distance in the formula. The new features that are introduced in the proposed similarity measure that are the geometric mean averaging and Hausdorff distance have some advantages. For instance, geometric mean averaging enables an exact computation of similarity value between generalized fuzzy numbers having a similar shape but with different height or spread of the bases. Geometric mean averaging is also able to clearly discriminate fuzzy numbers with similar shape. Most of the existing methods cannot distinctively discriminate the generalized fuzzy numbers in the case of similar shape and for the sets that having almost looked alike shape. Most methods yield the same similarity value for both cases. The advantage of Hausdorff distance is it can evaluate the mismatch of the geometric distance between fuzzy numbers by determining the maximal distance of generalized fuzzy numbers from the $\alpha$-level. The proposed similarity measure is able to evaluate the similarity value of a similar generalized fuzzy number while the existing methods cannot discriminate clearly the similarity value of the similar generalized fuzzy numbers. The proposed similarity measure is beneficial for comparing generalized fuzzy number in solving most decision making problems under fuzzy environment.

ACKNOWLEDGEMENTS

The authors would like to thank the Universiti Teknologi MARA for its financial support.
REFERENCES


