Int. J. Open Problems Compt. Math., Vol. 2, No. 2, June 2009

Fuzzy Similarity Measure for Generalized Fuzzy Numbers

B. Sridevi and R. Nadarajan

Department of Mathematics and Computer Applications
PSG College of Technology, Coimbatore
Tamilnadu-India
e-mail: srinichu@yahoo.com

Abstract

This paper proposes a new fuzzy similarity measure to calculate the degree of similarity of generalized fuzzy numbers (GFN's).The fuzzy similarity measure is developed by integrating the concept of centre of gravity (COG) points and fuzzy difference of distance of points of fuzzy numbers. A fuzzy description for difference of distances between fuzzy numbers in its turn exploits appropriate similarity measure between the pattern sets when compared with other measures available. It greatly reduces the influence of inaccurate measures and provides a very intuitive quantification. Several sets of pattern recognition problems and a fingerprint matching problem are taken to compare the proposed method with the existing similarity measures. Our approach gives a better and more robust similarity measure.

Keywords: *Centre of Gravity, Generalized Fuzzy Numbers, Generalized Trapezoidal Fuzzy Numbers, Generalized Triangular Fuzzy Numbers, Similarity measure.*

1 Introduction

In traditional theories world representations are forced to comply with extremely precise models, avoiding and rejecting imprecision as a perturbation factor. However, imprecision plays an important role in information representation in real processes where increase in precision would otherwise become unmanageable. Fuzzy set theory allows the formalization of approximate reasoning and preserves the original information contents of imprecision. The fuzzy sets defined on the set of real numbers are used in many applications of fuzzy theory. Their membership functions of the form A:R → [0, 1] have quantitative meaning and may be considered as fuzzy numbers or intervals when they capture the concept of approximate reasoning G.J.Klir [1].Since any imprecision factor is represented as fuzzy number the study of their similarity measure becomes very important in the research topic of pattern recognition. In pattern recognition we often want to measure the geometric properties of regions in an image that are not crisply defined. Many of the standard geometric properties and relationship among regions are generalized to fuzzy numbers. So to measure their similarity is very important in decision making .Automated fingerprint classification [2], [3] constitutes a complex problem in the pattern recognition domain. Fuzzy geometrical features of finger print images can be considered in the form of fuzzy numbers to handle the uncertainties in decision making process. Various similarity measures have been proposed to calculate the degree of similarity between fuzzy numbers. In this paper we review only those measures that are most related to our measure. However there are some drawbacks in the earlier ones, the current method overcomes the drawback.

1.1 Outline of the paper

Section 2 gives the basic definitions, section 3 discusses the existing similarity measures, section 4 presents the new fuzzy similarity measure, its relevant properties and describes the approach of applying the current method to a fingerprint matching problem. Section 5 compares and discusses the results of proposed measure with other measures through pattern sets, and the conclusion is given in section 6.

2 Generalized Trapezoidal Fuzzy Numbers (GTFN) [4],[5]

The membership function of GTFN A= (a,b,c,d;w) where $a \le b \le c \le d$, $0 < w \le 1$ is defined as

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

If w = 1, then GTFN A is a normal trapezoidal fuzzy number A = (a, b, c, d). If a = b and c = d, then A is a crisp interval .If b = c then A is a generalized triangular fuzzy number. If a = b = c = d and w = 1 then A is a real number. Compared to normal fuzzy number the GFN can deal with uncertain information in a more flexible manner because of the parameter w that represent the degree of confidence of opinions of decision maker's.

3 Existing similarity measures between fuzzy numbers

For any 2 trapezoidal fuzzy numbers A= (a_1, a_2, a_3, a_4) and B= (b_1, b_2, b_3, b_4)

3.1 The similarity measure S.M.Chen [6]

$$S(A, B) = 1 - \frac{\sum_{i=1}^{4} |a_i - b_i|}{4}$$

3.2 In C.H Hsieh [7]

$$\mathbf{S(A, B)} = \frac{1}{1 + d(A,B)} \quad \text{Where } d(A,B) = |P(A) - P(B)|$$

$$P(A) = \frac{a_1 + 2a_2 + 2a_3 + a_4}{6} \qquad P(B) = \frac{b_1 + 2b_2 + 2b_3 + b_4}{6}.$$

3.3 Simple center of gravity method (SCGM) Chen and Chen [8]

The SCGM is based on the concept of medium curve [9] .The SCGM method integrates the concepts of geometric distance and the COG distance of GFN's. If the GFN's are A= $(a_1, a_2, a_3, a_4; w_A)$ and B=($b_1, b_2, b_3, b_4; w_B$) $0 \le a_1 \le a_2 \le a_3 \le a_4 \le 1$ and $0 \le b_1 \le b_2 \le b_3 \le b_4 \le 1$.

COG(A)=(x_A^*, y_A^*), COG (B) = (x_B^*, y_B^*) then

$$S(A,B) = 1 - \frac{\sum_{i=1}^{4}|a_i - b_i|}{4} \left(1 - \left|x_A^* - x_B^*\right|\right)^{B(S_A, S_B)} \frac{\min(y_A^*, y_B^*)}{\max(y_A^*, y_B^*)}$$ Where

$$y_A^* = \begin{cases} \dfrac{w_A\left(\dfrac{a_3 - a_2}{a_4 - a_1} + 2\right)}{6} & if \ a_1 \ne a_4 \\ \dfrac{w_A}{2} & if \ a_1 = a_4 \end{cases} \tag{1}$$

$$x_A^* = \frac{y_A^*(a_3 + a_2) + (a_4 + a_1)(w_A - y_A^*)}{2w_A} \tag{2}$$

$$B(S_A, S_B) = \begin{cases} 1 & if \ S_A + S_B > 0 \\ 0 & if \ S_A + S_B = 0 \end{cases} \tag{3}$$

$$S_A = a_4 - a_1 \quad S_B = b_4 - b_1 \tag{4}$$

3.4 The Radius of gyration based similarity measure Deng Yong [10]

$$\text{S (A, B)} = 1 - \frac{\sum_{i=1}^{4}|a_i - b_i|}{4} \left(1 - \left|r_x^A - r_x^B\right|\right)^{B(S_A, S_B)} \frac{\min(r_y^A, r_y^B)}{\max(r_y^A, r_y^B)}$$

Where $r_x^A = \sqrt{\dfrac{(I_x)_1 + (I_x)_2 + (I_x)_3}{((a_3 - a_2) + (a_4 - a_1))\dfrac{w_A}{2}}}$ $\quad r_y^A = \sqrt{\dfrac{(I_y)_1 + (I_y)_2 + (I_y)_3}{((a_3 - a_2) + (a_4 - a_1))\dfrac{w_A}{2}}}$

$$(I_x)_1 = \frac{(a_2 - a_1)w_A^3}{12} \quad (I_x)_2 = \frac{(a_3 - a_2)w_A^3}{3} \quad (I_x)_3 = \frac{(a_4 - a_3)w_A^3}{12}$$

$$(I_y)_1 = \frac{(a_2-a_1)^3 w_A}{4} + \frac{(a_2-a_1)a_1^2 w_A}{2} + \frac{2(a_2-a_1)^2 a_1 w_A}{3}$$
$$(I_y)_2 = \frac{(a_3-a_2)^3 w_A}{3} + \frac{(a_3-a_2)a_2^2 w_A}{1} + \frac{(a_3-a_2)^2 a_2 w_A}{1}$$
$$(I_y)_3 = \frac{(a_4-a_3)^3 w_A}{12} + \frac{(a_4-a_3)a_3^2 w_A}{2} + \frac{2(a_4-a_3)^2 a_1 w_A}{3}$$

3.5 Similarity measure based on geometric mean averaging operator Shi-Jay Chen [11]

$$S(A,B) = \left[\sqrt[4]{\prod_{i=1}^{4} (2-|a_i-b_i|)} - 1 \right] \times \frac{\min(y_A^*, y_B^*)}{\max(y_A^*, y_B^*)}$$ where y_A^*, y_B^* are given by (1).

4 New fuzzy similarity measure for GFN's

Though the methods discussed in the previous section could predict the similarity of fuzzy numbers they fail to correctly give the similarity measure in certain situations. Here we present a new similarity measure based on fuzzy difference of distance of points of fuzzy numbers rather than geometric distances used by the existing methods. We see that from pattern sets given in section 5 the current fuzzy similarity measure not only overcomes the drawback of the earlier methods it also gives the similarity measure with better accuracy.

The membership function to measure the difference in distance of points of two GFN's is defined as

$$\mu_d(x) = \begin{cases} 1-\dfrac{x}{d}, & 0 \le x \le d \\ 0, & \text{otherwise} \end{cases}$$

where $0 < d \le 1$ and x $= |a_i - b_i|$. The degree of similarity of two GFN's A and B is defined as

$$\mathrm{S(A,B)} = \frac{1}{4}\sum_{i=1}^{4} \mu_d(x)\, (1-|x_A^* - x_B^*|)^{B(S_A,S_B)}\, \frac{\min(y_A^*, y_B^*)}{\max(y_A^*, y_B^*)} \quad (5)$$

$B(S_A, S_B)$ is 0 or 1 according as COG point is considered or not and $x_A^*, x_B^*, y_A^*, y_B^*$ are given in (1), (2), (3), (4). In pattern recognition problems like fingerprint matching, earthquake damage analysis, speech recognition, handwriting recognition, image analysis etc., the study of similarity measure becomes very important . When the absolute difference between the base of two fuzzy numbers is small the value of $\mu_d(x)$ is large. The choice of d, represents the degree of accuracy required to measure the similarity of the fuzzy numbers. In situations where high degree of accuracy is required in the similarity rating , the value of d can be chosen as low as possible.

Here we take a more popular example of biometric technology the fingerprint matching problem. Finger prints are perhaps what the majority of people immediately associate with biometrics. In automatic fingerprint identification systems the system will search for a matching print and may in fact produce a list of many potential matches. So to study the degree of similarity between the supplied fingerprint and those listed out as matches is a crucial problem. Many fingerprint recognition algorithm are based on minutiae matching because it is more reliable and discriminating feature .Two features are selected - the number of matched sample points n and the mean distance difference of the matched minutiae pairs m Xingjian Chen [12].The membership function for n and m are represented by Gaussian function .

The feature n is represented by fuzzy feature N as $\mu_N(n) = e^{-\left(\frac{n-N_g}{\|N_g-N_i\|}\right)^2}$

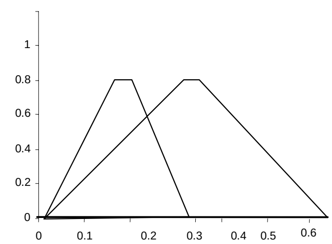
The feature m is represented by fuzzy feature M as $\mu_M(m) = e^{-\left(\frac{m-M_g}{\|M_g-M_i\|}\right)^2}$

Where N_i, - imposter match cluster centers of N ,N_g - genuine match cluster centers of N,M_i–imposter match cluster centers of M, M_g - genuine match cluster centers of M. We take a sample data of FVC 2002 DB1 with N_i = 18, N_g = 230, M_i = 4.8, M_g = 3.1. We approximate the Gaussian function as Trapezoidal function Min-You Chen [13] and obtain the GFN's as N (0, 0.31, 0.35, 0.64; 0.8), M (.07, 0.17, 0.18, 0.27; 0.8). The data is the finger prints of two different persons.

Figure 1 represents the fingerprint matching set for fuzzy features N and M. The value of d is taken as 0.5 for all the pattern sets discussed in this paper. For complicated sets where exactness is expected to be high the value of d can be taken accordingly. Using the proposed similarity measure we get

$\frac{1}{4}\sum_{i=1}^{4}\mu_d(x) =$ *.625 COG of N = (0.3234, 0.275) COG of M = (0.1717, 0.273)*

S (N, M) = .625 x .8483 x .9927 = 0.5263



N (0, 0.31, 0.35, 0.64; 0.8)

M (0.07, 0.17, 0.18, 0.27; 0.8)

Fig 1. Fingerprint matching pattern

4.1 Relevant Properties of the New Fuzzy Similarity Measure

Apart from the basic properties (reflexivity and symmetry) discussed in [8], the fuzzy similarity measure presented here satisfies other properties which reduces the computational work. The relevant properties we consider for the similarity measures depend on the usefulness within the domain of research but they are not considered as complete.

Property 1: If GFN's are real numbers i.e. if A=(a,a,a,a;1) and B(b,b,b,b;1) then S (A, B) = $1-\frac{|a-b|}{d}$

Proof: For real numbers $y_A^* = y_B^* = .5$

$S_A = S_B = 0 \quad \therefore S_A + S_B = 0\ \&\ B(S_A, S_B) = 0$

$S(A,B) = \frac{1}{4}\mu_d(x)(1-|x_A^* - x_B^*|)^0 \frac{0.5}{0.5}$ $\quad \mu_d|a-b| = 1-\frac{|a-b|}{d}$. *So* S (A, B) = $1-\frac{|a-b|}{d}$.

Property2: If A and B are same real numbers $a_1 = a_2 = a_3 = a_4 = b_1 = b_2 = b_3 = b_4; w_A \neq w_B$ or if they have same base but different weights i.e. $a_1 = b_1, a_2 = b_2, a_3 = b_3, a_4 = b_4; w_A \neq w_B$ we have $y_A^* = \frac{w_A}{w_B} y_B^*$ and $x_A^* = x_B^*$.

Proof:

Case I : If the GFN's are real numbers From (1) $y_A^* = \frac{w_A}{2}$, $y_B^* = \frac{w_B}{2}$ then $\frac{y_A^*}{y_B^*} = \frac{w_A}{w_B}$.

Hence $y_A^* = \frac{w_A}{w_B} y_B^*$.

From (2) $x_A^* = \dfrac{\frac{w_A}{2}(2a_1) + 2a_1(w_A - \frac{w_A}{2})}{2w_A} = a_1$

Similarly $x_B^* = a_1$. Therefore $x_A^* = x_B^*$.

Case II : If GFN's are have same base From (1) $y_A^* = \dfrac{w_A(\frac{a_3 - a_2}{a_4 - a_1} + 2)}{6}$ now since $a_i = b_i$ for all i we have

$$y_B^* = \frac{w_B(\frac{a_3 - a_2}{a_4 - a_1} + 2)}{6} \qquad \frac{y_A^*}{y_B^*} = \frac{w_A}{w_B} \quad , \quad y_A^* = \frac{w_A}{w_B} y_B^* \tag{6}$$

$$x_A^* = \frac{\frac{w_A}{w_B} y_B^* (a_3 + a_2) + (a_4 + a_1)(1 - \frac{y_B^*}{w_B}) w_A}{2w_A}$$

$$= \frac{y_B^* (a_3 + a_2) + (a_4 + a_1)(w_B - y_B^*)}{2w_B}$$

$$= \frac{y_B^* (b_3 + b_2) + (b_4 + b_1)(w_B - y_B^*)}{2w_B} = x_B^*$$

Property 3: If the GFN's are same real numbers or if they have same base and in either case different weights and in particular if $w_A = 1$ or $w_B = 1$ then S (A, B) = min (w_A, w_B).

Proof: If A and B are same real numbers or if they are GFN's with same base then according to property 2 we have $y_A^* = \frac{w_A}{w_B} y_B^*$ and $x_A^* = x_B^*$.

Let $w_A = 1$ and since $y_B^* \leq w_B$

$$\frac{\min(y_A^*, y_B^*)}{\max(y_A^*, y_B^*)} = \frac{\min(\frac{w_A}{w_B} y_B^*, y_B^*)}{\max(\frac{w_A}{w_B} y_B^*, y_B^*)} = \frac{\min(\frac{y_B^*}{w_B}, y_B^*)}{\max(\frac{y_B^*}{w_B}, y_B^*)}$$

$$= \frac{y_B^*}{\frac{y_B^*}{w_B}} = w_B$$

Also $\mu_d(x) = 4 \ \forall \ x$ So S (A, B) = 1 x 1 x w_B = min (w_A,w_B).

4 Comparison Of Fuzzy Similarity Measure With The Existing Methods

The concept of fuzzy difference of points and COG points of GFN's is incorporated in the new fuzzy similarity measure. Few pattern sets of generalized fuzzy numbers are taken to compare the proposed similarity method with the existing ones. The pattern sets are shown in figure 2. The results obtained are given in table 1.The similarity rate of fuzzy numbers depends on various aspects like the degree of confidence of decision maker, the differences in distance of the base of GFN's (even when they are of same shapes), the COG values of the GFN's etc.If the degree of confidence w is low then accordingly the similarity rate should also be low.

It is seen that in set 10 though the GFN's are of same shape their base values are different whereas in set 1 the GFN's are of different shapes but the base values are same .Set 1 is more similar than set 10, but Chen and Chen [8], Deng Yong [10] and Shi Jay Chen[11]depicts the similarity measure for set 10 higher than set 1.

In set 5 both the GFN's are same except only the value of w is different. Whereas in set 7 the GFN's are two different real lines but Chen and Chen [8], Deng Yong [10] and Shi-Jay Chen [11] rates set 7 with higher similarity than set 5 and our proposed method gives the similarity rate as .8 for both the cases which seems to be reasonable measure.

In set 16 the GFN's are with different base, different shapes and the values of w also differ, but the similarity values given by the existing methods are too high especially Deng Yong [10] gives the similarity rate as .7 which is too high. Also Shi-Jay Chen[11] predicts the similarity rate as .5927 for set 16 and for set 3 which is less similar than set 16 it gives higher value as .5997 wherein all the other methods rate set 16 more similar than set 3 .Our proposed method rates set

16 as .29 and set 3 as .24.A similar type of similarity results can be found in set 12 and set 23.

Also for set 13 and set 14 the similarity measures given by existing methods are very high. It is obviously seen that set 9 ,set 21 and set 22 are same type of GFN's all the methods gives the similarity rate as same except Deng Yong [10].

Compared to all the pattern sets given in fig 1 we can see that the GFN's in Set 17 are dissimilar because the value of w is only .3897and the base values of GFN's are very wide apart A (0, .225, .45; .225$\sqrt{3}$) and B (.45, .675, .9;.225$\sqrt{3}$). Our proposed method gives the similarity measure as .055 but the values given by Hsieh [7] and Chen [6] and Shi Jay [11] are not acceptable. Similarly set 18 has the GFN's with w value different as well as too low, they are of different shapes defined in different intervals accordingly we get the similarity rate also low as 0.1031.

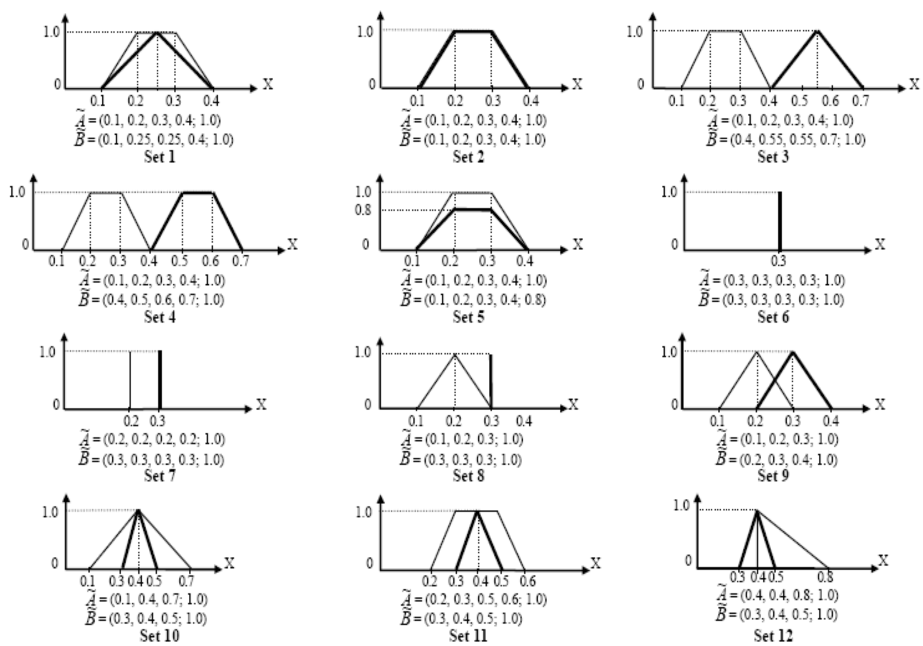
Deng Yong [10] argument was Chen and Chen [8] gives same similarity measure for those GFN's that have same COG points and their method rules out this flaw and gives different similarity rate. Though the problem is ruled out in [10] the similarity measures produced by it for certain GFN's are not acceptable for pattern sets like 13, 14, 15 and 16.

Of all the distinct GFN's given in fig 2, it is very well seen that set 1 is having higher similarity. The proposed method gives a high similarity rate for set 1 compare to other distinct GFN's discussed in fig 2. So the new similarity measure not only overcomes the drawbacks of the existing similarity measures it also gives a better similarity rating. From the results obtained in table 1 we see that the existing methods fail to predict the correct similarity measure in certain cases whereas the new fuzzy similarity method gives an accurate measure.

Table 1 Comparison of the Calculation Results of Fuzzy Similarity Measure with the existing methods

SET	CHEN[6]	HSIEH [7]	CHEN [8]	DENGYONG [10]	SHI [11]	PROPOSED METHOD
1	.975	1	.8357	.7954	.8356	.8143
2	1	1	1	1	1	1
3	.7	.7692	.42	.4028	.5997	.24
4	.7	7692	.49	.4931	.7	.28
5	1	1	.8	.8	.8	.8
6	1	1	1	1	1	1
7	.9	.909	.9	.81	.9	.8
8	.9	.909	.54	.5754	.5991	.48
9	.9	.909	.81	.8112	.9	.72
10	.9	1	.9	.8854	.8974	.8
11	.9	1	.72	.6914	.72	.64

12	.9	.9375	.78	.7744	.8959	.69
13	.7	.7692	.49	.8961	.6971	.28
14	.7	.7692	.49	.7781	.7	.28
15	.7	.7692	.49	.4931	.7	.28
16	.7	.7692	.49	.7004	.5927	.29
17	.55	.6897	.3025	.309	.55	.055
18	.55	.6897	.3025	.287	.3612	.1031
19	.8	.8333	.5486	.5905	.6854	.4110
20	.8	.8333	.5486	.5899	.6854	.4110
21	.9	.9091	.81	.8568	.9	.72
22	.9	.9091	.81	.8551	.9	.72
23	.9	1	.8077	.8255	.8077	.7179
24	.9	1	.8028	.8255	.8028	.71
25	.85	.8696	.6193	.6419	.728	.5094
26	.85	.8696	.6193	.6441	.728	.5094



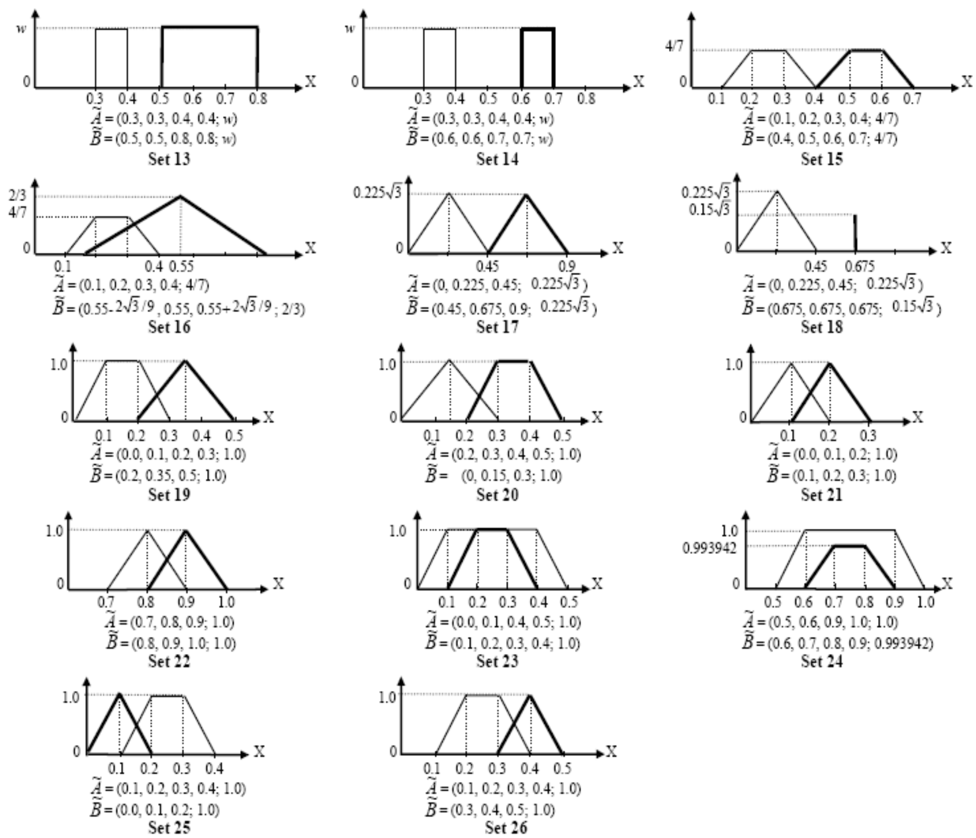


Fig 2 Few sets of GFN's

6 Conclusion

In this paper we presented a new fuzzy similarity measure for generalized fuzzy numbers. The proposed measure work successfully in situations where the GFN's have same COG points and overcomes the drawbacks of the existing methods. We see that giving a fuzzy definition for distances between points of fuzzy numbers very much improves the similarity measure than the geometric distances adopted by earlier methods. The measure greatly reduces the influence of inaccurate measures and provides a very intuitive quantification. The results obtained by our method reflect the significance of fuzzy representation rather than the crisp definition.

7 Open problem

Active research is going on in the study of the similarity measure of fuzzy numbers in various areas. Several novel measures are used to access similarity of fuzzy numbers but the current measure correlates better than the other measures For crudely categorizing pairs of fuzzy numbers as either similar or dissimilar all the measures performed well. But for distinguishing between degrees of similarity between exigent pairs certain measures were clearly superior and others were clearly inferior. However there is no serious attempt to validate the techniques through behavioral experiments. Some authors have mentioned that their technique work very well, but they do not provide appropriate data to support their claim. Future plan is to acquire validity for the behavior of the measure and scale up the experiment with larger database.

References

[1] G.J.Klir and B.Yuan, Fuzzy sets and Fuzzy Logic Theory and Applications, Upper Saddle River, NJ Prentice –Hall, 1995.

[2] H.M. Hsu and C.T.Chen, Aggregation of fuzzy opinions under group decision making, *Fuzzy Sets Syst.*, vol.79, no.3, pp.279-285, 1996.

[3] H.S.Lee, An optimal aggregation method for fuzzy, opinions of group decision, Proc.1999 *IEEE Int. Conf. Systems, Man, Cybernetics*, vol.3, pp.314- 319, 1999.

[4] S.H.Chen, Operations on fuzzy numbers with function principal, *Tamkang J. Manag.Sci.*,vol.6,no.1,pp.13 -25,1985.

[5] Ranking generalized fuzzy number with graded mean integration, *in proc.8th Int.Fuzzy Syst.Assoc.World Congr.,vol.2,Taipei*,Taiwan,Republic of China, 1999 , pp.899-902.

[6] Shyi Meng Chen, New methods for subjective mental workload assessment and fuzzy risk analysis, *Cyber. Syst.: Int.J.* vol.27, no.5, pp.449-472, 1996.

[7] C.H.Hsieh and S.H.Chen, Similarity of generalized fuzzy numbers with graded mean integration representation, in *Proc. 8th Int. Fuzzy Systems Association World Congr., vol.2, Taipe*i, Taiwan, Republic of China, 1999, pp.551-555.

[8] Si Jay Chen. Shyi Meng Chen., Fuzzy risk analysis based on similarity measures of generalized fuzzy numbers, *IEEE Trans. Fuzzy Systems* 11, pp. 45-56,2003.

[9] P.Subasic and K.Hirota, Similarity rules and gradual rules for analogical and interpolative reasoning with imprecise data, *Fuzzy Sets Syst.*, vol.96, no.1, pp.53-75, 1998.

[10] Deng Yong, Shi Wenkang , Du Feng , Liu Qi, A new similarity measure of generalized fuzzy numbers and its application to pattern recognition, *in Pattern Recognition Letters*,vol.25,pp 875-883,2004.

[11] Shi-Jay Chen , New Similarity measure of generalized fuzzy numbers based on geometric mean averaging operator, in *Proc.IEEE International Conference on Fuzzy Systems,Fuzz-IEEE 2006,Vancouver,*Canada.

[12] Xingjian Chen, Jie Tian, A new algorithm for distorted fingerprints matching based on normalized fuzzy similarity measure, IEEE Trans on Image Processing, Vol.15, no.3, March 2006

[13] Min-You Chen and D.A.Linkens, Rule-base self-generation and simplification for data driven fuzzy models ,Fuzzy Sets ans syst.,vol142,pp243-265,2004.