SEARCHING FOR A CONSENSUS SIMILARITY FUNCTION FOR GENERALIZED TRAPEZOIDAL FUZZY NUMBERS

E. VICENTE, A. MATEOS* and A. JIMÉNEZ-MARTÍN

Decision Analysis and Statistics Group, Universidad Politécnica de Madrid, Campus de Montegancedo s/n, Boadilla del Monte, 28660-Madrid, Spain
* E-mail: amateos@fi.upm.es

There is controversy regarding the use of the similarity functions proposed in the literature to compare generalized trapezoidal fuzzy numbers since conflicting similarity values are sometimes output for the same pair of fuzzy numbers. In this paper we propose a similarity function aimed at establishing a consensus. It accounts for the different approaches of all the similarity functions. It also has better properties and can easily incorporate new parameters for future improvements. The analysis is carried out on the basis of a large and representative set of pairs of trapezoidal fuzzy numbers.

1. Introduction

A great variety of similarity functions have been proposed in the literature to compare fuzzy numbers. The usual parameters are distance, shape and size, which are then aggregated in different ways leading to different functions.

However, there is controversy regarding the use of similarity functions since different similarity functions sometimes output conflicting similarity values for the same pair of trapezoidal fuzzy numbers. Moreover, their performance is often analyzed based on biased sets of fuzzy numbers, clearly highlighting the benefits but disguising the associated drawbacks.

In this paper we propose a similarity function for generalized trapezoidal fuzzy numbers aimed at establishing a consensus between the similarity functions reported in the literature, whose performance is analyzed on the basis of a large and representative set of pairs of trapezoidal fuzzy numbers.

In Section 2 we review the most representative similarity functions in

the literature and the set of pairs of fuzzy numbers commonly used for analyzing their performance. In Section 3 we develop a method for deriving a representative set of pairs of fuzzy numbers. We also propose a stable consensus similarity function. Finally, we conclude with the advantages of the proposed consensus function in Section 4.

2. Similarity Functions for Generalized Trapezoidal Fuzzy Numbers

The similarity measures of fuzzy numbers are usually based on a comparison of various associated parameters, such as distance, shape or size. One of the first measures of similarity was proposed by Chen[1] in 1996. It used the geometric distance between the considered fuzzy numbers. However, this function is not applicable for generalized fuzzy numbers.

Chen[2] himself extended his similarity measure in 2003 to generalized trapezoidal fuzzy numbers, incorporating the distance between the centers of gravity of the compared numbers to the similarity function. Another measure proposed by Chen[3] in 2009 accounted for the perimeter and height of the considered generalized trapezoidal fuzzy numbers.

Gomathi and Sivaraman[4] proposed using the geometric mean rather than the average distance between vertices. Wen et al.[5] added the area of the considered fuzzy numbers to the similarity function.

In 2009 Sridevi and Nadarajan[6] replaced Chen's[3] geometrical distance by the fuzzy distance $1 - \sum_{i=1}^{4} \frac{\mu_d(x_i)}{4}$, with $\mu_d(x_i) = 1 - (x_i/d)$ if $0 \leq x_i \leq d$ (0 otherwise), $d \in (0, 1]$ and $x_i = | a_i - b_i |$. If $d = 1$ we have Chen's similarity function. In the same year, Wei and Chen[7] proposed a measure adding the perimeter and height of generalized trapezoidal fuzzy numbers to the geometric distance.

Xu et al.[8] proposed the geometrical distance and the distance between the centers of gravity of the compared numbers in 2010. However, this function had a serious drawback since $S(\widetilde{a}, \widetilde{b}) \neq 1 - |a - b| \, ; \forall a, b \in \mathbb{R}$.

The most recent proposals are by Zhu and Xu[9] in 2012 and Vicente et al.[10] in 2013. However, Zhu and Xu's function is not well defined since the similarity between $\widetilde{a} = (0.01, 0.01, 0.01, 0.01; 0.5)$ and $\widetilde{b} = (1, 1, 1, 1; 1)$ is $S(\widetilde{a}, \widetilde{b}) = \sqrt{-0.0021}$, whereas the function by Vicente et al. is too demanding because it imposes a penalty of $(1 - \alpha - \beta)$ when the compared numbers do not share area.

Arbitrary sets of fuzzy numbers are usually used to compare the proposed similarity functions, in such a way that the benefits of the proposed

function are clearly shown but the associated drawbacks remain hidden. Chen originally proposed a set with only 15 pairs of fuzzy numbers, which was enlarged afterwards by other authors, such as Sridevi and Nadarajan[6] (26 pairs), Xu et al.[8] (30 pairs) or Vicente-Cestero et al.[11] (35 pairs), but other authors, such as Gomathi and Sivaraman[4], continue to use small sets of pairs (5 pairs). Besides, some sets are clearly deficient. For example, Xu et al. and Gomathi and Sivaraman use fuzzy numbers with zero height, which does not make sense.

Figure 1 shows the performances of the above similarity functions on the basis of the set of pairs proposed by Chen and Chen[2] (15 pairs). The similarity output by the different functions varies enormously for some pairs, especially, 3, 4, 8, 10, 11 and 15.
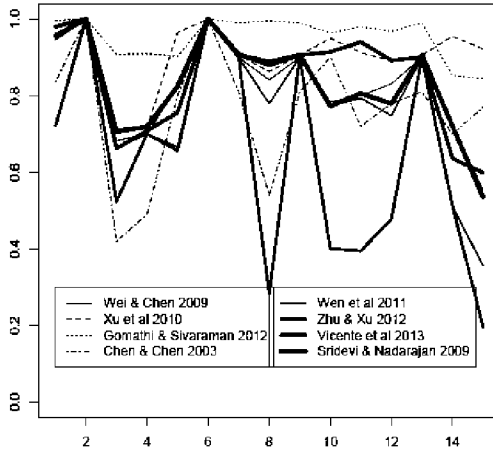

Fig. 1.   Similarity values on the set of Chen and Chen

Different criteria could be used to search for a consensus among the considered similarity functions. We have considered the Pearson correlation coefficient, i.e., the consensus function should maximize the Pearson correlation coefficient regarding the other similarity functions. The functions by Gomathi and Sivaraman and by Zhu and Xu will not be considered henceforth since they generate atypical data associated with incorrectly processed pairs. An analytical expression of a similarity function that satisfies these criteria is derived in Section 3.

3. Deriving a Consensus Similarity Function

The most used parameters in similarity functions are the arithmetic and geometric distances, distances between the centers of gravity, differences between the heights, perimeters and areas. We consider the following families of similarity functions:

$$S_1(\widetilde{a}, \widetilde{b}) = 1 - (w_1 d_1 + w_2 d_2 + w_3 d_3),$$

$$S_2(\widetilde{a}, \widetilde{b}) = 1 - (w_1 d_1 + w_2 d_2 + w_3 d_3) \left(\frac{d_4}{d_5}\right)^{v_1} \left(\frac{d_6}{d_7}\right)^{v_2} \left(\frac{d_8}{d_9}\right)^{v_3} \left(\frac{d_{10}}{d_{11}}\right)^{v_4} \left(\frac{d_{12}}{d_{13}}\right)^{v_5},$$

$$S_3(\widetilde{a}, \widetilde{b}) = 1 - (w_1 d_1 + w_2 d_2 + w_3 d_3) \prod_{i=1}^{n} \left(\frac{v_{1i} d_4 + v_{2i} d_6 + v_{3i} d_8 + v_{4i} d_{10} + v_{5i} d_{12}}{v_{1i} d_5 + v_{2i} d_7 + v_{3i} d_9 + v_{4i} d_{11} + v_{5i} d_{13}}\right),$$

$$S_4(\widetilde{a}, \widetilde{b}) = [1 - (w_1 d_1 + w_2 d_2 + w_3 d_3)] \prod_{i=1}^{n} \left(\frac{v_{1i} d_4 + v_{2i} d_6 + v_{3i} d_8 + v_{4i} d_{10} + v_{5i} d_{12}}{v_{1i} d_5 + v_{2i} d_7 + v_{3i} d_9 + v_{4i} d_{11} + v_{5i} d_{13}}\right),$$

with $\widetilde{a} = (a_1, a_2, a_3, a_4; w_1)$ and $\widetilde{b} = (b_1, b_2, b_3, b_4; w_2)$, $d_1 = (\sum |a_i - b_i|)/4$, the Euclidean distance; $d_2 = d\left((X_{\widetilde{a}}, Y_{\widetilde{a}}), (X_{\widetilde{b}}, Y_{\widetilde{b}})\right)$, the distance between their centers of gravity; $d_3 = (\prod_{i=1}^{4} |a_i - b_i|)^{1/4}$, the geometric mean; $d_4 = min\{P_{\widetilde{a}}, P_{\widetilde{b}}\}$ ($d_5 = max\{P_{\widetilde{a}}, P_{\widetilde{b}}\}$), the minimum (maximum) of their perimeters; $d_6 = min\{A(\widetilde{a}), A(\widetilde{b})\}$ ($d_7 = max\{A(\widetilde{a}), A(\widetilde{b})\}$), the minimum (maximum) of their areas; $d_8 = min\{w_{\widetilde{a}}, w_{\widetilde{b}}\}$, $d_9 = max\{w_{\widetilde{a}}, w_{\widetilde{b}}\}$, $d_{10} = min\{Q_{\widetilde{a}}, Q_{\widetilde{b}}\}$, $d_{11} = max\{Q_{\widetilde{a}}, Q_{\widetilde{b}}\}$, with $Q_{\widetilde{a}} = \sqrt{(a_2 - a_1)^2 + (a_3 - a_2)^2 + (a_4 - a_3)^2 + w_{\widetilde{a}}^2}$; and $d_{12} = \int_0^1 \mu_{\widetilde{a} \cap \widetilde{b}}(x) dx$ ($d_{13} = \int_0^1 \mu_{\widetilde{a} \cup \widetilde{b}}(x) dx$), the intersection area (union), $w_1, w_2, w_3, v_j, v_{ji} \in [0, 1]$, with $w_1 + w_2 + w_3 = 1$, $\sum v_j = 1$ and $\sum v_{ji} = 1$.

Note that the second family generalizes to the first family when we consider $v_i = 0 \ \forall i$. The third family generalizes to the second family when we consider $n = 5$ and $(v_i = 1, v_{ji} = 0 \ \forall i \neq j) \ \forall i = 1, ..., 5$.

We have found using Monte Carlo simulation that the functions that best fit the consensus are from the fourth family ($S_4(\widetilde{a}, \widetilde{b})$) with $n = 1$. Therefore, we focus on this family with $n = 1$ to identify the consensus function.

Then, the problem is to compute the values $w_i, v_j \in [0, 1]$ that maximizes the minimum Pearson correlation coefficient R_i between the function S_4 and each considered similarity function on the basis of 48 pairs of fuzzy numbers taken from Xu et al., Gomathi and Sivaraman, Sridevi and Nadarajan and Wei and Chen:

$$max \ z = \min_i \{R_i\}$$
$$s.\,t. \quad w_1 + w_2 + w_3 = 1, \qquad 0 \leq w_i \leq 1, \forall i$$
$$v_1 + v_2 + v_3 + v_4 + v_5 = 1, \ 0 \leq v_j \leq 1, \forall j$$

However, the above optimization problem is complex and it is necessary to use metaheuristics to solve it. The optimum solution using *simulated annealing*[12] is w_1^*=0.234, w_2^*=0.744, w_3^*=0.022, v_1^*=0.109, v_2^*=0.123, v_3^*=0.15, v_4^*=0.604 and v_5^*=0.013. Note that the most relevant parameters in the optimal solution are w_2^*, associated with the distance between the centers of gravity; and v_4^*, associated with elements $Q_{\tilde{a}}$ and $Q_{\tilde{b}}$ proposed by Gomathi and Sivaraman to reduce computational time for the perimeter assessment.

The correlation coefficients of the consensus function regarding the similarity functions by Wei and Chen, Chen, Xu et al, Wen et al, Vicente et al and Sridevi and Nadarajan are 0.96, 0.92, 0.90, 0.91, 0.96 and 0.95, respectively. Figure 2 shows how the fit of the consensus function compared with the considered similarity functions.
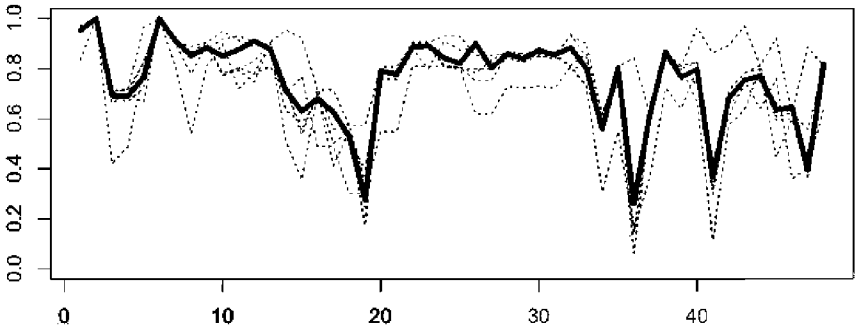

Fig. 2.   Comparing the values output by the consensus function

4. Conclusion

We have proposed a consensus function on the basis of different similarity functions reported in the literature for generalized trapezoidal fuzzy numbers. The proposed function has better properties than some similarity functions reported in the literature, such as those proposed by Xu et al, which does not properly measure the similarity between two real numbers, or by Zhu and Xu, which is not fully defined.

The proposed function aggregates the approaches of different similarity functions whose outputs vary enormously, making a consensus necessary. Besides, it can easily incorporate new perspectives (parameters) for future

improvements. The appropriateness of parameters can be evaluated, and parameters with low weights can be discarded, leading to a refined function with fewer parameters.

Acknowledgments

References

1. S. M. Chen, New methods for subjective mental workload assessment and fuzzy risk analysis, *Cybernetics and Systems*, *27*, 449-472 (1996).
2. S. J. Chen, S. M. Chen, Fuzzy risk analysis based on similarity measures of generalized fuzzy numbers, *IEEE Transactions on Fuzzy Systems*, *11*, 45-56, (2003).
3. S. M. Chen, J.-H. Chen, Fuzzy risk analysis based on similarity measures between interval-valued fuzzy numbers and interval-valued fuzzy number arithmetic operators, *Expert Systems with Applications*, *36*, 6309-6317, (2009).
4. V. L. Gomathi Nayagam, G. Sivaraman, A novel similarity measure between generalized fuzzy numbers, *International Journal of Computer Theory and Engineering*, *4*, 448-450, (2012).
5. J. Wen, X. Fan, D. Duanmu, D. Yong, A modified similarity measure of generalized fuzzy numbers, *Procedia Engineering*, *15*, 2773-2777, (2011).
6. B. Sridevi, R. Nadarajan, Fuzzy similarity measure for generalized fuzzy numbers, *International Journal of Open Problems in Computer Science and Mathematics*, *2*, 111-116, (2009).
7. S. H. Wei, S. M. Chen, A new approach for fuzzy risk analysis based on similarity measures of generalized fuzzy numbers, *Expert Systems with Applications*, *36*, 589-598, (2009).
8. Z. Xu, S. Shang, W. Qian, W. Shu, A method for fuzzy risk analysis based on the new similarity of trapezoidal fuzzy numbers, *Expert Systems with Applications*, *37*, 1920-1927, (2010).
9. L. Zhu, R. Xu, *Fuzzy risk analysis based on similarity measure of generalizard fuzzy numbers*, Fuzzy Enginering and Operations Research. Berlin/Heidelberg: Springer, 569-587, (2012).
10. E. Vicente, A. Mateos, A. Jiménez, A new similarity function for generalized trapezoidal fuzzy numbers, *Lecture Notes in Artificial Intelligence*, *7894*, 400-411, (2013).
11. E. Vicente-Cestero, A. Mateos, A. Jiménez-Martín, Similarity functions for generalized trapezoidal fuzzy numbers: An improved comparative analysis, *Journal of Intelligence and Fuzzy Systems*, under revision, (2014).
12. S. Kirkpatrick, C. D. Gelatt, C. D. Vecchi, Optimization by simulated annealing, *Science*, *220*, 671-680, (1983).