Proceedings of the 3rd World Congress on Mechanical, Chemical, and Material Engineering (MCM'17)
Rome, Italy – June 8 – 10, 2017
Paper No. ICMIE 104
ISSN: 2369-8136
DOI: 10.11159/icmie17.104

A Centroid Based Correlation Coefficient of Fuzzy Numbers

Junhu Ruan[1,2,3], Felix T. S. Chan[2], Fangwei Zhu[3], Yan Shi[4], Yumeng Wang[5]
[1]College of Economics and Management, Northwest A&F University
No. 3, Taicheng Road, Yangling, China
rjh@nwafu.edu.cn
[2]Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University
Hung Hom, Hong Kong
f.chan@polyu.edu.hk
[3]Faculty of Management and Economics, Dalian University of Technology
No. 2, Linggong Road, Dalian, China
zhufw@dlut.edu.cn
[4]General Education Center, Tokai University
9-1-1, Toroku, Kumamoto, Japan
yanshi@ktmail.tokai-u.jp
[5]School of Agricultural Economics and Rural Development, the Renmin University of China
No. 59, Zhongguancun Street, Beijing, China
wymmyw@ruc.edu.cn

Abstract - Classic methods have been well reported to measure the correlation coefficient of crisp observed data. However, the uncertainty in the real world sometimes makes crisp data unavailable, especially for linguistic variables. Under this situation, fuzzy numbers are often involved in the observed data, but classic statistical methods cannot be directly used to calculate the correlation coefficient of fuzzy observed data. Motivated by this observation, we integrate the centroid technique with Pearson's correlation coefficient to propose a simple method for measuring the correlation coefficient of fuzzy data. The centroid based method is applied into the measurement of correlation coefficient between technology and management. The comparison with extant results shows the effectiveness and advantage of our method.

Keywords**:** Fuzzy observed data; Correlation coefficient; Centroid-based method; Technology and management

1. Introduction

Correlation is one important and basic relationship for researchers to analyze the intertwined world. Well-known methods to measure the correlation have been widely used, such as correlation diagram and correlation coefficient. Especially, Pearson's correlation coefficient is the most common measure. These classic statistical techniques are suitable to crisp observed data. However, in the real world, we sometimes cannot obtain crisp variable values due to the uncertainty and ambiguity. For example, it is an acceptable way to use fuzzy numbers to quantify linguistic variables. Then, how to determine the correlation coefficient of fuzzy observed data is a facing problem. Classic statistics does not give the answer.

In order to deal with the problem, some works have been reported in the literature. Gerstenkorn and Manko [1] gave an early interrelation coefficient of intuitionistic fuzzy sets. This interrelation coefficient was proved with good properties to measure the correlation of fuzzy sets, but it cannot differentiate the negative or positive correlation. Then, Chiang and Lin [2] used the conventional statistics to present a formula for measuring the fuzzy correlation coefficient among fuzzy data. Chiang and Lin' formula extended the correlation of fuzzy sets into [-1, 1] which is consistent with classic statistical techniques. However, Liu and Kao [3] argued that the correlation coefficient of fuzzy numbers should be a fuzzy number, and then used the α-cut technique to present a fuzzy correlation coefficient. Recently, more improved methods and applications are reported such as Hsu and Wu [4], Ye [5], Liao *et al.* [6], Yang [7], Ban *et al.* [8], and Şahin and liu [9].

We find that most extant methods are with high computation complexity which is not convenient to deal with practical problems as a simple way. The centroid technique is a simple and effective way to compare fuzzy numbers [10-13], but few reports are found to apply the centroid technique into the problem of determining the correlation coefficient of fuzzy data.

Motivated by this observation, we integrate the centroid technique with Pearson's correlation coefficient to propose a simple method for measuring the correlation coefficient of fuzzy data. The centroid based method is applied into the measurement of correlation coefficient between technology and management. The comparison with extant results shows the effectiveness and advantage of our method.

The organization of the rest is as follows. In Section 2, we present the proposed centroid based measure of fuzzy correlation coefficient. In Section 3, an application example is given to show the effectiveness of the proposed method. Section 4 concludes the work.

2. The Proposed Method

2.1. The problem description

A correlation coefficient is an index to measure the degree of correlation of observed data. Pearson's correlation coefficient is one of the most common indices for measuring the degree of correlation. Given two crisp datasets both with n observations, that is, $X=\{x_1,x_2,...,x_n\}$ and $Y=\{y_1,y_2,...,y_n\}$ where x_i and y_i denote the ith observed crisp values in these two datasets respectively, the Pearson's correlation coefficient of these two datasets (denoted as $r_{X,Y}$) is [14]:

$$r_{X,Y}=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2\sum_{i=1}^{n}(y_i-\bar{y})^2}} \tag{1}$$

where $\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$ and $\bar{y}=\frac{1}{n}\sum_{i=1}^{n}y_i$ are the means of X and Y, respectively. This is the classic index for measuring the correlation of two crisp datasets.

However, due to the uncertainty and variability, the observed data are sometimes not expressed by crisp values. For example, in order to observe the correlation between the freshness of in-transit fruits and the temperature of carrying vehicles, it is difficult to use crisp values to express the freshness as well as other linguistic variables. Then, it is better to use fuzzy numbers. In addition, interval or fuzzy values are often used to express the temperature range. In this case, how to measure the degree of correlation between two fuzzy variables is a difficulty. That is, given two fuzzy observed datasets, $\tilde{X}=\{\tilde{x}_1,\tilde{x}_2,...,\tilde{x}_n\}$ and $\tilde{Y}=\{\tilde{y}_1,\tilde{y}_2,...,\tilde{y}_n\}$ where $\tilde{x}_i$ and $\tilde{y}_i$ denote the ith observed fuzzy values in these two datasets respectively, how to formulate the correlation coefficient? In this work, we present a centroid based method to solve the problem.

2.2. Centroid based comparison of fuzzy numbers

Among the extant methods for comparing fuzzy numbers, the centroid method is one simple and effective technique which has been used in various fields [10-13]. For two triangular fuzzy numbers $\tilde{A}=(a_1,a_2,a_3,1)$ and $\tilde{B}=(b_1,b_2,b_3,1)$ where a_1, a_2 and a_3 are the left threshold, the midpoint and the right threshold of $\tilde{A}$ and b_1, b_2 and b_3 are the left threshold, the midpoint and the right threshold of $\tilde{B}$, their membership functions are:

$$\mu_{\tilde{A}}(x)=\begin{cases}\dfrac{x-a_1}{a_2-a_1},\ a_1\le x\le a_2\\ \dfrac{x-a_2}{a_3-a_2},\ a_2\le x\le a_3\\ 0,\ \text{otherwise}\end{cases} \tag{2}$$

$$\mu_{\tilde{B}}(x)=\begin{cases}\dfrac{x-b_1}{b_2-b_1},\ b_1\le x\le b_2\\ \dfrac{x-b_2}{b_3-b_2},\ b_2\le x\le b_3\\ 0,\ \text{otherwise}\end{cases} \tag{3}$$

where $\mu_{\tilde{A}}(x)$ and $\mu_{\tilde{B}}(x)$ denote the membership functions of $\tilde{A}$ and $\tilde{B}$, their centroids respectively denoted by $\bar{\tilde{A}}_{cent}$ and $\bar{\tilde{B}}_{cent}$ are:

$$\bar{\tilde{A}}_{cent}=\frac{1}{3}(a_1+a_2+a_3) \tag{4}$$

$$\bar{\tilde{B}}_{cent}=\frac{1}{3}(b_1+b_2+b_3) \tag{5}$$

As we can see, fuzzy numbers can be transferred to corresponding crisp numbers by the centroid method. Based on this observation, we present a centroid based method for measuring the correlation coefficient of fuzzy numbers, as detailed in Section 2.3.

2.3. A centroid based method for measuring the correlation coefficient of fuzzy numbers

For two observed fuzzy datasets $\tilde{X}$ and $\tilde{Y}$:

$$\tilde{X}=\{\tilde{x}_1,\tilde{x}_2,...,\tilde{x}_n\}=\{(x_1^1,x_1^2,x_1^3),(x_2^1,x_2^2,x_2^3),...,(x_n^1,x_n^2,x_n^3)\} \tag{6}$$

$$\tilde{Y}=\{\tilde{y}_1,\tilde{y}_2,...,\tilde{y}_n\}=\{(y_1^1,y_1^2,y_1^3),(y_2^1,y_2^2,y_2^3),...,(y_n^1,y_n^2,y_n^3)\} \tag{7}$$

where x_i^1, x_i^2 and x_i^3 are the left threshold, the midpoint and the right threshold of the *i*th observed fuzzy value in $\tilde{X}$ (that is, $\tilde{x}_i$) and y_i^1, y_i^2 and y_i^3 are the left threshold, the midpoint and the right threshold of the *i*th observed fuzzy value in $\tilde{Y}$ (that is, $\tilde{y}_i$), their centroid datasets respectively denoted by $\bar{\tilde{X}}_{cent}$ and $\bar{\tilde{Y}}_{cent}$ can be determined by Eqs. (4) and (5):

$$\bar{\bar{X}}_{cent} = \{\bar{\bar{x}}_1^{cent}, \bar{\bar{x}}_2^{cent}, ..., \bar{\bar{x}}_n^{cent}\} \tag{8}$$

$$\bar{\bar{Y}}_{cent} = \{\bar{\bar{y}}_1^{cent}, \bar{\bar{y}}_2^{cent}, ..., \bar{\bar{y}}_n^{cent}\} \tag{9}$$

where $\bar{\bar{x}}_i^{cent}$ and $\bar{\bar{y}}_i^{cent}$ are the centroids of $\tilde{x}_i$ and $\tilde{y}_i$, $i = 1, 2, ..., n$:

$$\bar{\bar{x}}_i^{cent} = \frac{1}{3}(x_i^1 + x_i^2 + x_i^3) \ \forall i = 1, 2, ..., n \tag{10}$$

$$\bar{\bar{y}}_i^{cent} = \frac{1}{3}(y_i^1 + y_i^2 + y_i^3) \ \forall i = 1, 2, ..., n \tag{11}$$

Then, we can calculate the Pearson's correlation coefficient of the centroid datasets $\bar{\bar{X}}_{cent}$ and $\bar{\bar{Y}}_{cent}$:

$$r_{\bar{\bar{X}}_{cent}, \bar{\bar{Y}}_{cent}} = \frac{\sum_{i=1}^{n}(\bar{\bar{x}}_i^{cent} - \bar{\bar{\bar{x}}}^{cent})(\bar{\bar{y}}_i^{cent} - \bar{\bar{\bar{y}}}^{cent})}{\sqrt{\sum_{i=1}^{n}(\bar{\bar{x}}_i^{cent} - \bar{\bar{\bar{x}}}^{cent})^2 \sum_{i=1}^{n}(\bar{\bar{y}}_i^{cent} - \bar{\bar{\bar{y}}}^{cent})^2}} \tag{12}$$

where $\bar{\bar{\bar{x}}}^{cent}$ and $\bar{\bar{\bar{y}}}^{cent}$ are the means of $\bar{\bar{X}}_{cent}$ and $\bar{\bar{Y}}_{cent}$:

$$\bar{\bar{\bar{x}}}^{cent} = \frac{1}{n}\sum_{i=1}^{n} \bar{\bar{x}}_i^{cent} \tag{13}$$

$$\bar{\bar{\bar{y}}}^{cent} = \frac{1}{n}\sum_{i=1}^{n} \bar{\bar{y}}_i^{cent} \tag{14}$$

In this work, we use $r_{\bar{\bar{X}}_{cent}, \bar{\bar{Y}}_{cent}}$ to measure the correlation coefficient of the two fuzzy datasets $\tilde{X}$ and $\tilde{Y}$. The example with the comparison with extant results shows the effectiveness and advantage of our proposed measure.

3. An Application Example

In the section, we use an extant application example to verify our method. In the study of Liu and Kao [3], they collected the fuzzy data of technology and management indexes of 15 Taiwan machinery firms, as Table 1 shows. As we can see, we cannot use classic correlation coefficients to measure the correlation due to the fuzzy values of technology and management. Using Eqs. (8)-(11), we can get the centroids of these fuzzy indices, as the last columns in Table 1 show.

Table 1: The original data and their centroids.

Firms	Fuzzy technology indices			Fuzzy management indices			Centroids of fuzzy technology indices	Centroids of fuzzy management indices
	x_i^1	x_i^2	x_i^3	y_i^1	y_i^2	y_i^3		
1	0.0790	0.2120	0.3527	0.2486	0.4493	0.6616	0.2146	0.4532
2	0.6348	0.7953	0.9069	0.4902	0.7170	0.9138	0.7790	0.7070
3	0.4424	0.6287	0.7420	0.2624	0.4243	0.5752	0.6044	0.4206
4	0.1692	0.3677	0.5585	0.6258	0.8507	0.9680	0.3651	0.8148
5	0.4674	0.6029	0.7598	0.2331	0.4290	0.6233	0.6100	0.4285
6	0.6469	0.8472	0.9322	0.1395	0.3139	0.5042	0.8088	0.3192
7	0.1835	0.3179	0.4605	0.0707	0.2290	0.4003	0.3206	0.2333
8	0.3636	0.5577	0.6740	0.1076	0.2894	0.4866	0.5318	0.2945
9	0.0628	0.1992	0.4037	0.1469	0.3175	0.5056	0.2219	0.3233
10	0.1554	0.3337	0.4934	0.2555	0.4303	0.6218	0.3275	0.4359
11	0.0085	0.1123	0.3030	0.0448	0.1214	0.2188	0.1413	0.1283
12	0.2927	0.4170	0.6096	0.3371	0.5442	0.7540	0.4398	0.5451
13	0.3112	0.5098	0.6399	0.5249	0.7458	0.9495	0.4870	0.7401
14	0.2757	0.4903	0.6348	0.4957	0.7192	0.8310	0.4669	0.6820
15	0.3394	0.5011	0.6411	0.3863	0.6115	0.7715	0.4939	0.5898

Then, using our measurement method, we can get the correlation coefficient between technology and management based on the fuzzy data in Table 1. The comparison between Liu and Kao's method with our work is as Table 2 and Figure 1 show. From the comparison, we can have the following observations:

(1) Liu and Kao's correlation coefficient is a fuzzy number consisting of a set of intervals with different a-cut levels. The fuzzy correlation coefficient is difficult to be used in conventional statistical analysis.

(2) The correlation coefficient by our measure is a crisp value, that is, 0.2926, and our measure value is quite close to the a-cut value with the maximum membership by Liu and Kao's method, that is, 0.2950. Thus, our measurement method can be used to determine the correlation coefficient of fuzzy observed data, with a low computation complexity and a high confidence level.

Table 2: The correlation coefficients by Liu and Kao's and our methods.

Liu and Kao' result [3]			Our result
a-cuts	Lower	Upper	
0	-0.8281	0.9862	0.2926
0.1	-0.7748	0.9704	
0.2	-0.6933	0.9475	
0.3	-0.5947	0.9163	
0.4	-0.4861	0.8743	
0.5	-0.3696	0.8167	
0.6	-0.2426	0.7409	
0.7	-0.1091	0.6537	
0.8	0.0267	0.5425	
0.9	0.1649	0.4233	
1	0.2950	0.2950	

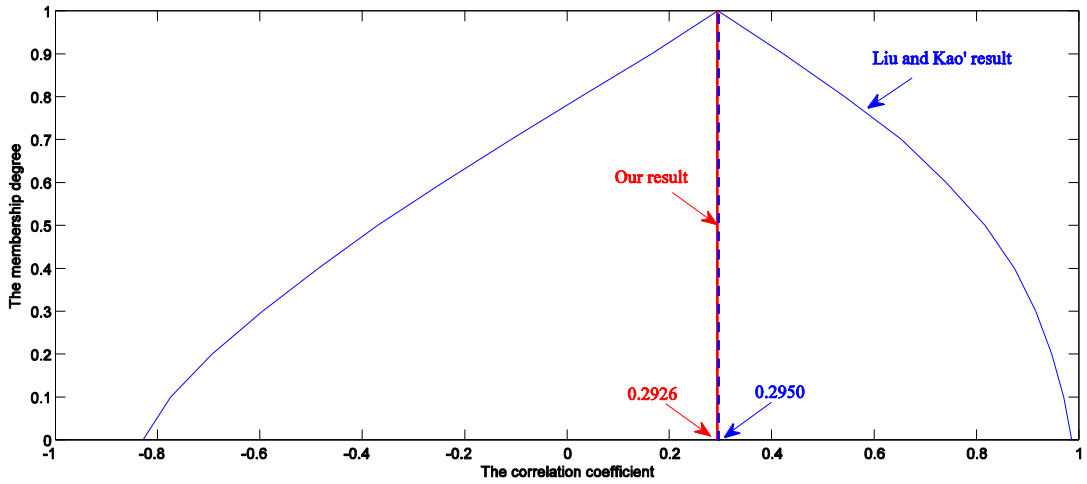

Fig. 1: The comparison with Liu and Kao's result.

4. Conclusion

In this work, we are concerned with the determination of the correlation coefficient of fuzzy observed data. The centroid technique is introduced into the classic Pearson's correlation coefficient to formulate a centroid based method. Using the proposed method, we can measure the correlation coefficient of any two fuzzy observed datasets with triangular membership functions. The measurement method is simple in computation and is verified with a high confidence level in comparison with extant results. Further works are needed to prove the mathematical properties and robustness of our proposed method. In addition, we also will try to develop more measurement methods using other fuzzy comparison techniques such as the integral method and a-cut based methods.

Acknowledgements

This work was supported by grants from the Natural Science Basic Research Project in Shaanxi Province (No. 2016JQ7005); China Ministry of Education Social Sciences and Humanities Research Youth Fund Project (No. 16YJC630102); China Postdoctoral Science Foundation (No. 2016M600209); The Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. PolyU 15201414); The Natural Science Foundation of China (Grant No. 71471158); and the Hong Kong Scholars Program Mainland-Hong Kong Joint Postdoctoral Fellows Program (No. G-YZ87 and XJ2015007).

References

[1] T. Gerstenkorn, J. Manko, "Correlation of intuitionistic fuzzy sets," *Fuzzy Set. Syst.*, vol. 44, no. 1, pp. 39-43, 1991.
[2] D.-A. Chiang, N. P. Lin, "Correlation of fuzzy sets," *Fuzzy Set. Syst.*, vol. 102, no. 2, pp. 221-226, 1999.
[3] S.-T. Liu, C. Kao, "Fuzzy measures for correlation coefficient of fuzzy numbers," *Fuzzy Set. Syst.*, vol. 128, no. 2, pp. 267-275, 2002.
[4] H.-L. Hsu, B. Wu, "An innovative approach on fuzzy correlation coefficient with interval data," *Int. J. Innov. Comput. I.*, vol. 6, no. 3(A), pp.1349-4198, 2010.
[5] J. Ye, "Correlation coefficient of dual hesitant fuzzy sets and its application to multiple attribute decision making," *Appl. Math. Model.*, vol. 38, no. 2, pp. 659-666, 2014.
[6] H. Liao, Z. Xu, X.-J. Zeng, J. M. Merigó, "Qualitative decision making with correlation coefficients of hesitant fuzzy linguistic term sets," *Knowl.-Based Sys.*, vol. 76, pp. 127-138, 2015.
[7] C.-C. Yang, "Correlation coefficient evaluation for the fuzzy interval data," *J. Bus. Res.*, vol. 69, no. 6, pp. 2138-2144, 2016.
[8] O.-I. Ban, I. G. Tara, V. Bogdan, D. Tuşe, S. G. Bologa, "Evaluation of hotel quality attribute importance through fuzzy correlation coefficient," *Technol. Econ. Dev. Econ.*, 2016, vol. 22, no. 4, pp.471-492.

[9] R. Şahin, P. Liu, "Correlation coefficient of single-valued neutrosophic hesitant fuzzy sets and its applications in decision making," *Neural Comput. Appl.*, doi:10.1007/s00521-015-2163-x, 2016.
[10] W. J. Wang, L. Luoh, "Simple computation for the defuzzifications of center of sum and center of gravity," *J. Intell. Fuzzy Syst.*, vol. 9, no. 1-2, pp. 53-59, 2000.
[11] Y.-M. Wang, J.-B. Yang, D.-L. Xu, K.-S. Chin, "On the centroids of fuzzy numbers," *Fuzzy Sets Syst.*, vol. 157, no. 7, pp. 919-926, 2006.
[12] A. Hadi-Vencheha, M. N. Mokhtarian, "A new fuzzy MCDM approach based on centroid of fuzzy numbers," *Expert Syst. Appl.*, vol. 38, no. 5, pp. 5226-5230, 2011.
[13] J. Ruan, P. Shi, C.-C. Lim, X. Wang, "Relief supplies allocation and optimization by interval and fuzzy number approaches," *Inf. Sci.*, vol. 303, pp. 15-32, 2015.
[14] H. Zhou, Z. Deng, Y. Xia, M. Fu, "A new sampling method in particle filter based on Pearson correlation coefficient," *Neurocomputing*, vol. 216, pp. 208-215, 2016.