International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-9 Issue-1S3, December 2019

On Computing Cluster Centers of Trapezoidal Fuzzy Numbers

S. Sreenivasan, B. J. Balamurugan

***Abstract:** In this paper we compute cluster centers of trapezoidal fuzzy numbers through fuzzy c means clustering algorithm and kernel based fuzzy c means clustering algorithm. A new complete metric distance between the trapezoidal fuzzy numbers is used to compute the cluster centers on the set of trapezoidal fuzzy numbers.*

***Keywords:** Fuzzy clustering, Kernel function, Trapezoidal fuzzy numbers, Fuzzy C number clustering algorithms.*

I. INTRODUCTION

In the history of fuzzy clustering, Dunn [3] fuzzy c means (FCM) clustering method proposed by and Bezdek [1] improved the algorithm, Dubois, D and Prade, H.[2]. Based on the similar idea of FCM construction, Yang and Ko [11] proposed new type of fuzzy clustering procedure called fuzzy c numbers (FCN) algorithm. However, Hung. W.L [6] explained the drawbacks of FCN over FCM and provided the alternate fuzzy c numbers algorithm (AFCN). Based on the concept of robust statistics, the AFCN algorithm improves on the weakness found in FCN. In [10] we have computed cluster centers of trapezoidal fuzzy numbers using two different metric distances.

In this paper, we use a new metric distance d_f to discuss the robustness of the fuzzy c means type algorithms.

The section II describes the definition of trapezoidal fuzzy number and the distance d_f, introduced by Hadi Sadoghi et.al [4], between two trapezoidal intervals. Two clustering algorithms [10, 13, 14] on the space of all trapezoidal fuzzy intervals are described in the section III based on the distance d_f. Numerical example is given to compare the effectiveness of the two algorithms FCM and KFCM in section IV. Conclusions are given in section V.

II. PRELIMINARIES

Definition 1: $X=(m_1,m_2,\alpha,\beta)_{LR}$ with $\alpha>0,\beta>0$ is called a LR-type fuzzy number if

$$X(x)=\begin{cases} L\left(\dfrac{m_1-x}{\alpha}\right), & for\ \ x\le m_1 \\ 1, & for\ \ m_1\le x\le m_2 \\ L\left(\dfrac{x-m_2}{\beta}\right), & for\ \ x\ge m_2 \end{cases}$$

We note that the trapezoidal type of fuzzy number is the simplest and most used one among LR -type fuzzy intervals.

Let $X=(m_1,m_2,\alpha,\beta)_{LR}$ be a LR -type fuzzy interval. If L and R of the form

$$T(x)=\begin{cases} 1-x, & 0\le x\le 1 \\ 0, & Otherwise \end{cases}$$

Then

$$X(x)=\begin{cases} 1-\dfrac{m_1-x}{\alpha}, & for\ \ x\le m_1 \\ 1, & for\ \ m_1\le x\le m_2 \\ 1-\dfrac{x-m_2}{\beta}, & for\ \ x\ge m_2 \end{cases}$$

is said to be a trapezoidal fuzzy interval

Let the set $F_{LR}(R)$ of all LR -type fuzzy intervals. Hadi Sadoghi et.al [4] gives a metric distance $d_f^2(X,Y)$ based on Yang[11] between $X=(m_{1x},m_{2x},\alpha_x,\beta_x)_{LR}$ and $Y=(m_{1y},m_{2y},\alpha_y,\beta_y)_{LR}$ which are in $F_{LR}(R)$ as follows:

$$d_f^2(X,Y)=\frac{1}{4}\left(g_-^2+g_+^2+\left(g_- -(\alpha_x-\alpha_y)\right)^2+\left(g_+ -(\beta_x-\beta_y)\right)^2\right)$$

Where $g_- =2(m_{1x}-m_{1y})-(m_{2x}-m_{2y})$ and $g_- =2(m_{1x}-m_{1y})+(m_{2x}-m_{2y})$

Theorem 1: $(F_{LR}(R),d_f)$ is metric.

Theorem 2: $(F_{LR}(R),d_f)$ is complete.

III. FUZZY CLUSTERING ALGORITHMS

In this section we recall FCM clustering algorithm[1] and KFCM clustering algorithm[12,13].

Let the set $X=\{\tilde{X}_1,\tilde{X}_2,\tilde{X}_3,\ldots,\tilde{X}_n\}$ of fuzzy numbers in $F_{LR}(R)$ with $\tilde{X}_k=(m_{1x_k},m_{2x_k},\alpha_{x_k},\beta_{x_k})$, $1\le k\le n$. Let c be the number of clusters.

Manuscript published on 30 December 2019.
***** Correspondence Author (s)
S. Sreenivasan*, School of Advanced Sciences, VIT University, Chennai Campus, Chennai, India. Email: sreenivasan.s2017@vitstudent.ac.in
B. J. Balamurugan, School of Advanced Sciences, VIT University, Chennai Campus, Chennai, India. Email: balamurugan.bj@vit.ac.in

Let $V=\{\tilde{V}_i \mid 1<i\le c\}$ be the set of centers, where $\tilde{V}_i=(m_{1v_i},m_{2v_i},\alpha_{v_i},\beta_{v_i})$ and d_{ik} the distance between $\tilde{X}_k$ and $\tilde{V}_i$.

A. Fuzzy c Means Clustering

The set X can be divided into c fuzzy subsets using FCM by minimizing the objective function

$$J_{FCM}(U,V)=\sum_{i=1}^{c}\sum_{k=1}^{n}u_{ik}^{m}d_{ik}^{2}$$

where $d_{ik}=d(\tilde{V}_i,\tilde{X}_k)$ and u_{ik} is the membership value of the fuzzy number $\tilde{X}_k$ in cluster i with $\sum_{i=1}^{c}u_{ik}=1$, the fuzziness index $m\in[1,\infty)$ and $U=(u_{ik})_{n\times c}$ is a fuzzy c partition matrix.

The Parameters of FCM are calculated by improving the function min J step by step according to the formulas below:

$$m_{1v_i}=\frac{\sum_{k=1}^{n}u_{ik}^{m}m_{1x_k}}{\sum_{k=1}^{n}u_{ik}^{m}} \tag{1}$$

$$m_{2v_i}=\frac{\sum_{k=1}^{n}u_{ik}^{m}m_{2x_k}}{\sum_{k=1}^{n}u_{ik}^{m}} \tag{2}$$

$$\alpha_{v_i}=\frac{\sum_{k=1}^{n}u_{ik}^{m}\alpha_{x_k}}{\sum_{k=1}^{n}u_{ik}^{m}} \tag{3}$$

$$\beta_{v_i}=\frac{\sum_{k=1}^{n}u_{ik}^{m}\beta_{x_k}}{\sum_{k=1}^{n}u_{ik}^{m}} \tag{4}$$

$$u_{ik}=\frac{d_{ik}^{\frac{-2}{m-1}}}{\sum_{j=1}^{c}d_{jk}^{\frac{-2}{m-1}}} \tag{5}$$

Based on these formulas, on trapezoidal fuzzy numbers we use the following fuzzy c means clustering algorithm.

Step1: Let the fuzziness index be $m>1$, let the number of partitions $c=\{2,3,4,\ldots,(n-1)\}$ and let any $\varepsilon>0$. Choose $U^{(0)}$ be the fuzzy c partition matrix initially and take $t=0$.

Step2: Calculate cluster centers $V^{(t)}=\{\tilde{V}_i^{(t)} \mid 1<i\le c\}$ using $U^{(t)}$ and equations $(1),(2),(3)$ and (4).

Step3: Improve $U^{(t)}$ as $U^{(t+1)}$ using $V^{(t)}$ and equation (5).

Step4: Compute $E^k=Max_{i,k}\left\{\left|u_{ik}^{(t+1)}-u_{ik}^{(t)}\right|\right\}$, if $E^k\le\varepsilon$, stop. Otherwise set $U^{(t+1)}=U^{(t)}$ and move to step 2.

B. Kernel Fuzzy c Means Clustering

A kernel function [5, 7, 8, 9] maps nonlinearly the input data space into high dimensional feature space H. In the M-dimensional space R^M, let the unlabeled set $X=\{\tilde{X}_1,\tilde{X}_2,\tilde{X}_3,\ldots,\tilde{X}_n\}$ and let

$$\Phi:R^M\to H,\tilde{X}\to\Phi(\tilde{X})$$

To find the dot product in the high dimensional feature space, we can use the kernel function $K(\tilde{X}_i,\tilde{X}_j)=\Phi(\tilde{X}_i).\Phi(\tilde{X}_j)$

Examples of kernel function

- Polynomial:
  $$K(\tilde{X}_i,\tilde{X}_j)=\left(\gamma\tilde{X}_i^T.\tilde{X}_j+c\right)^d,\gamma>0,d\in N$$
- Sigmoid: $K(\tilde{X}_i,\tilde{X}_j)=\tanh\left(\gamma\tilde{X}_i^T.\tilde{X}_j+c\right)^d,\gamma>0$
- RBF: $K(\tilde{X}_i,\tilde{X}_j)=\exp\left(-\gamma\left\|\tilde{X}_i-\tilde{X}_j\right\|^2\right),\gamma>0$

where γ, c, d are kernel parameters. Since,

$$\begin{aligned}\left\|\Phi(\tilde{X}_k)-\Phi(\tilde{V}_i)\right\|^2&=\left(\Phi(\tilde{X}_k)-\Phi(\tilde{V}_i)\right)^T\left(\Phi(\tilde{X}_k)-\Phi(\tilde{V}_i)\right)\\&=\Phi(\tilde{X}_k)^T\Phi(\tilde{X}_k)-\Phi(\tilde{X}_k)^T\Phi(\tilde{V}_i)\\&\quad-\Phi(\tilde{V}_i)^T\Phi(\tilde{X}_k)+\Phi(\tilde{V}_i)^T\Phi(\tilde{V}_i)\\&=K(\tilde{X}_k,\tilde{X}_k)+K(\tilde{V}_i,\tilde{V}_i)-2K(\tilde{X}_k,\tilde{V}_i)\end{aligned}$$

when the kernel function is chosen as RBF, $K(\tilde{X}_k,\tilde{X}_k)=1$, $K(\tilde{V}_i,\tilde{V}_i)=1$, then

$$\left\|\Phi(\tilde{X}_k)-\Phi(\tilde{V}_i)\right\|^2=2\left(1-K(\tilde{X}_k,\tilde{V}_i)\right)$$

The set X can be divided into c fuzzy subsets using KFCM by minimizing the objective function

$$J_{KFCM}(U,V)=\sum_{i=1}^{c}\sum_{k=1}^{n}u_{ik}^{m}\left(1-K(\tilde{X}_k,\tilde{V}_i)\right)$$

The Parameters of kernel fuzzy c means are calculated by improving the function min J step by step according to the formulas below:

$$m_{1v_i}=\frac{\sum_{k=1}^{n}u_{ik}^{m}K(\tilde{X}_k,\tilde{V}_i)m_{1x_k}}{\sum_{k=1}^{n}u_{ik}^{m}} \tag{6}$$

Retrieval Number: A10671291S319/2019©BEIESP
DOI:10.35940/ijeat.A1067.1291S319
Journal Website: www.ijeat.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-9 Issue-1S3, December 2019

$$m_{2v_i} = \frac{\sum_{k=1}^{n} u_{ik}^m K\left(\tilde{X}_k, \tilde{V}_i\right) m_{2x_k}}{\sum_{k=1}^{n} u_{ik}^m} \quad (7)$$

$$\alpha_{v_i} = \frac{\sum_{k=1}^{n} u_{ik}^m K\left(\tilde{X}_k, \tilde{V}_i\right) \alpha_{x_k}}{\sum_{k=1}^{n} u_{ik}^m} \quad (8)$$

$$\beta_{v_i} = \frac{\sum_{k=1}^{n} u_{ik}^m K\left(\tilde{X}_k, \tilde{V}_i\right) \beta_{x_k}}{\sum_{k=1}^{n} u_{ik}^m} \quad (9)$$

$$u_{ik} = \frac{\left[1 - K\left(\tilde{X}_k, \tilde{V}_i\right)\right]^{\frac{-1}{m-1}}}{\sum_{j=1}^{c} \left[1 - K\left(\tilde{X}_k, \tilde{V}_j\right)\right]^{\frac{-1}{m-1}}} \quad (10)$$

Based on these formulas, on trapezoidal fuzzy numbers we use the following kernel fuzzy means clustering algorithm.

Step1: Let the fuzziness index be $m > 1$, let the number of partitions $c = \{2, 3, 4, \ldots, (n-1)\}$ and let any $\varepsilon > 0$. Choose $U^{(0)}$ be the fuzzy c partition matrix initially and take $t = 0$.

Step2: Calculate cluster centers $V^{(t)} = \left\{\tilde{V}_i^{(t)} \middle| 1 < i \le c\right\}$ using $U^{(t)}$ and equations (6), (7), (8) and (9).

Step3: Improve $U^{(t)}$ as $U^{(t+1)}$ using $V^{(t)}$ and equation (10).

Step4: Compute $E^k = Max_{i,k}\left\{\left|u_{ik}^{(t+1)} - u_{ik}^{(t)}\right|\right\}$, if $E^k \le \varepsilon$, stop. Otherwise set $U^{(t+1)} = U^{(t)}$ and move to step 2.

In the experiment, we used RBF kernel is used with the parameter γ defined by

$$\gamma = \left(\frac{\sum_{k=1}^{n} d\left(\tilde{X}_k, \overline{W}\right)}{n}\right)^{-1}$$

with $\overline{W} = \left(\overline{m}_{1w}, \overline{m}_{2w}, \overline{\alpha}_w, \overline{\beta}_w\right)$ is the arithmetic mean, where

$$\overline{m}_{1w} = \frac{\sum_{k=1}^{n} m_{1x_k}}{n}$$

$$\overline{m}_{2w} = \frac{\sum_{k=1}^{n} m_{2x_k}}{n}$$

$$\overline{\alpha}_w = \frac{\sum_{k=1}^{n} \alpha_{x_k}}{n}$$

$$\overline{\beta}_w = \frac{\sum_{k=1}^{n} \beta_{x_k}}{n}$$

IV. NUMERICAL EXAMPLES

In this section, we implement the algorithms with $m = 2$ and $\varepsilon = 0.001$. Consider the data set G_1 from Yang, M.S and Ko, C.H.[11] consisting of 30 trapezoidal fuzzy numbers given in "Fig. 1".

We run both algorithms on the data set G_1 with the complete metric distances d_f and $c = 3$. The corresponding results are shown in "Table. I" and the cluster centers are shown in "Fig. 3".

Consider the set G_2 consisting G_1 and one more point (98.020, 100.000, 1.630, 0.710) called outlier point shown in "Fig. 2". Now on the set G_2 with the number of clusters $c = 2$, we run both algorithms. The corresponding results are shown in "Fig. 4" and "Table. II".

Table- I: Cluster centers in set G_1

		$(m_{1v}, m_{2v}, \alpha_v, \beta_v)$
		With distance $d_f^2(X, Y)$
Algorithm-1	V_1	(39.516,41.438,1.283,0.870)
	V_2	(24.077,26.081,0.909,1.043)
	V_3	(12.858,14.652,1.130,1.202)
Algorithm-2	V_1	(39.021,40.917,1.228,0.841)
	V_2	(24.047,25.996,0.814,1.063)
	V_3	(14.133,15.795,1.146,1.166)

Table- II: Cluster centers in set G_2

		$(m_{1v}, m_{2v}, \alpha_v, \beta_v)$
		With distance $d_f^2(X, Y)$
Algorithm-1	V_1	(98.900,100.101,0.672,1.024)
	V_2	(37.561,39.540,1.283,0.867)
	V_3	(16.555,18.413,1.095,1.180)
Algorithm-2	V_1	(39.359,41.267,1.357,0.860)
	V_2	(24.138,26.116,0.865,1.050)
	V_3	(13.626,15.359,1.134,1.187)

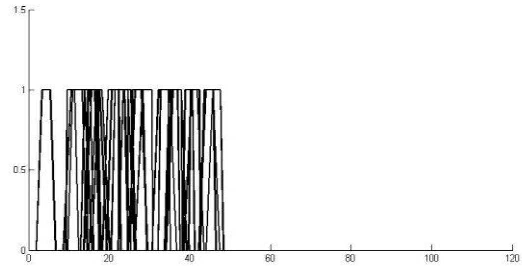
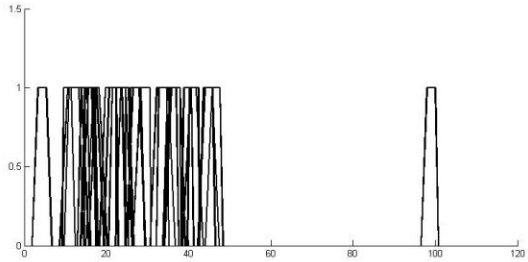
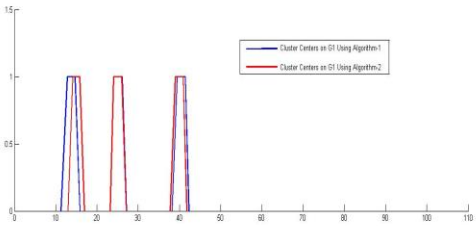


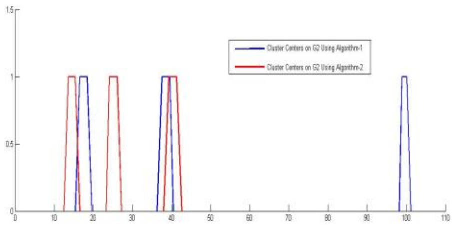
Fig. 1. Set G_1



Set G_2



Cluster centers in set G_1



Cluster centers in set G_2

When we run Algorithm-1 and Algorithm-2 on data set G_1 with the distance d_f, the cluster centers are almost same. While running Algorithm-1 on the data set G_2 (With outlier point) gives poor result, but the Algorithm-2 on G_2 provides almost the same result of G_1. That is the centers obtained by algorithm-1 on G_2 are away from the clusters whereas the centers obtained by Algorithm-2 on G_2 are within the cluster and coincides with the centers of G_1 data.

V. CONCLUSION

We used a different metric distance d_f to calculate centers of clusters through the FCM clustering algorithm and kernel based KFCM algorithm on trapezoidal fuzzy numbers. We have written MATLAB programming to compute the centers. The results show that the KFCM clustering algorithm performs better than the FCM clustering algorithm.

REFERENCES

1. J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, (Plenum Press, New York, 1981).
2. D Dubois and H Prade. *Fuzzy sets and systems: Theory and applications*, (Academic Press, New York, 1980).
3. J. C. Dunn, "A fuzzy relative of the ISODATA process and Its Use in Detecting Compact", well-Separated Cluster, J. Cybernet, 1974, 3, pp. 32 - 57.
4. Hadi Sadoghi Yazdi, Mehri Sadoghi Yazdi , Abedin Vahedian, "Fuzzy Bayesian Classification of LR Fuzzy Numbers", *IACSIT International Journal of Computer Theory and Engineering*, Vol. 1, No. 5, December 2009.
5. R Herbrich, *Learning Kernel Classifier*, (MIT Press, Cambridge, MA 2002).
6. W. L. Hung and M. S. Yang, "Fuzzy Clustering on LR-Type fuzzy numbers With an Application in Taiwanese Tea Evaluation", Fuzzy sets and systems, 2005, 150(3), pp. 561 - 577.
7. Meena Tushir, Smriti Srivastava, "A New Kernel based Hybrid c Means Clustering model", *IEEE international Fuzzy systems Conference*, 2007.
8. K R Muller, S Mika, G Ratsch, K Tsuda, B Scholkopf, "An introduction to kernel-based learning algorithms", *IEEE Transaction on Neural Networks* 12(2), 2001, pp. 181-201.
9. Rong Lan, Jiu-Iun Fan, "A Fuzzy C Means Type Clustring Algorithm on triangular Fuzzy Numbers", *Sixth international conference on Fuzzy Systems and Knowledge Discovery*, 2009.
10. S Sreenivasan and B J Balamurugan, "Computing Cluster Centers of Trapezoidal Fuzzy Numbers Through Fuzzy C Means and Kernel Based Fuzzy C Means Clustering Algorithms with Two Metric Distances Using Matlab", *International Journal of Civil Engineering and Technology (IJCIET)* 9(10), 2018, pp. 1322–1330.
11. M S Yang, C H Ko, "On a Class of fuzzy c Numbers Clustering Procedure for Fuzzy Data", *Fuzzy sets and systems*, 1996, 84(1), 49 - 60.
12. D Zhang, S Chen, "Fuzzy clustering kernel method", *in: Proc of the Internat. Conf. on Control and Automation*, 2002, pp. 123-127.
13. S Zhou, J Gan, "Mercer kernel fuzzy c means algorithm and prototypes of clusters", *in: Proc. of Conf. on Internat. Data Engineering and Automated Learning*, Vol 3177, 2004, pp. 613-618.

AUTHORS PROFILE

S. Sreenivasan, received his M.Sc. and M.Phil degrees in Mathematics from the University of Madras, Chennai, India. He has more than 20 years of teaching experience at Undergraduate and Postgraduate level courses. He is currently pursuing Ph.D. in Fuzzy Graph. He has published 5 research papers in various journals and conference proceedings. His research interest includes graph theory, Fuzzy Clustering and fuzzy logic.

Dr. B. J. Balamurugan, received his Ph.D degree in Mathematics from the University of Madras, Chennai, India. Currently, he is an Assistant Professor (Senior) of Mathematics in the School of Advanced Sciences at VIT University, Chennai Campus, Chennai, India. He has more than 22 years of teaching experience at Undergraduate and Postgraduate level courses. Dr. B. J. Balamurugan has published more than 30 research papers in various journals and conference proceedings. His research interest includes graph theory, graph grammars, fuzzy logic and Petri nets.

Retrieval Number: A10671291S319/2019©BEIESP
DOI:10.35940/ijeat.A1067.1291S319
Journal Website: www.ijeat.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)