International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-11, September 2019

Computing Cluster Centers of Triangular Fuzzy Numbers using Innovative Metric Distance

S. Sreenivasan, B. J. Balamurugan

***Abstract*: In this paper we compute cluster centers of triangular fuzzy numbers through fuzzy c means clustering algorithm and kernel based fuzzy c means clustering algorithm. An innovative distance between the triangular fuzzy numbers is used and the distance is complete metric on triangular fuzzy numbers. The set of triangular fuzzy numbers and an another set with the same triangular fuzzy numbers by including an outlier or noisy point as an additional triangular fuzzy number are taken to find the cluster centers using MATLAB programming. An example is given to show the effectiveness between the algorithms.**

***Keywords* : Fuzzy c means clustering algorithms, Fuzzy Clustering, Kernel function, Triangular fuzzy numbers.**

I. INTRODUCTION

In the history of fuzzy clustering, fuzzy c means (FCM) clustering procedure introduced by Dunn [3] and improved by Bezdek [1] are most used and discussed. However, FCM is not good if the cluster of points containing outlier or noisy points. A clustering algorithm is good if it is strong to deal with the real time problems.

In this paper we use KFCM on triangular fuzzy numbers and discuss the robustness of the fuzzy c means type algorithms with the new vertex distance. And we used the complete metric distance to find the distance between triangular fuzzy numbers.

The distance which is complete and metric on triangular fuzzy numbers is described in section II. Based on the complete metric distance two clustering algorithms on triangular fuzzy numbers are given in section III. 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. A COMPLETE METRIC DISTANCE FOR TRIANGULAR FUZZY NUMBERS

We consider the triangular fuzzy numbers on $R=(-\infty,\infty)$.

Definition 1: A fuzzy set $\tilde{A}=(A^L,A^C,A^R)$ is said to be triangular fuzzy number, if the membership function of the triangular fuzzy number is

$$\tilde{A}(x)=\begin{cases}0, & x<A^L\\ \dfrac{x-A^L}{A^C-A^L}, & A^L\le x\le A^C\\ \dfrac{x-A^R}{A^C-A^R}, & A^C\le x\le A^R\\ 0, & x>A^R\end{cases}$$

Let the set of all fuzzy numbers be $F(R)$. To run the clustering algorithms on the set $F(R)$ we apply the distance d [2] defined as follows:

Definition 2: The distance between two triangular fuzzy numbers $\tilde{A}=(A^L,A^C,A^R)$ and $\tilde{B}=(B^L,B^C,B^R)$ which is complete and metric is defined as $d^2(\tilde{A},\tilde{B})=\frac{1}{3}[(A^L-B^L)^2+(A^C-B^C)^2+(A^R-B^R)^2]$

This is an efficient method to calculate the distance between two triangular fuzzy numbers is used in this paper. Yang and Ko [9] proved that $(F(R),d)$ is a complete and metric.

III. FUZZY CLUSTERING ALGORITHMS

In this section we recall FCM clustering algorithm[1] and KFCM clustering algorithm[10], [11].

Let $X=\{\tilde{X}_1,\tilde{X}_2,\tilde{X}_3,...,\tilde{X}_n\}$ be the set of fuzzy numbers in $F(R)$ with $\tilde{X}_k=(X_k^L,X_k^C,X_k^R), 1\le k\le n$. Let the number of clusters be n. Let $V=\{\tilde{V}_i \,|\, 1<i\le c\}$ is the set of centers, where $\tilde{V}_i=(V_i^L,V_i^C,V_i^R)$ and d_{ik} be the distance between $\tilde{X}_k$ and $\tilde{V}_i$.

A. Fuzzy c Means Clustering

The fuzzy c means clustering algorithm divides X into c fuzzy subsets by minimizing the function

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

where $d_{ik}=d(\tilde{V}_i,\tilde{X}_i)$ 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.

Manuscript published on 30 September 2019.
*Correspondence Author(s)
S. Sreenivasan*, School of Advanced Sciences, VIT University, Chennai Campus, Chennai, India.
B. J. Balamurugan, School of Advanced Sciences, VIT University, Chennai Campus, Chennai, India.

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

$$V_i^L = \frac{\sum_{k=1}^{n} u_{ik}^m X_k^L}{\sum_{k=1}^{n} u_{ik}^m} \tag{1}$$

$$V_i^C = \frac{\sum_{k=1}^{n} u_{ik}^m X_k^C}{\sum_{k=1}^{n} u_{ik}^m} \tag{2}$$

$$V_i^R = \frac{\sum_{k=1}^{n} u_{ik}^m X_k^R}{\sum_{k=1}^{n} u_{ik}^m} \tag{3}$$

$$u_{ik} = \frac{d_{ik}^{\frac{-2}{m-1}}}{\sum_{j=1}^{c} d_{jk}^{\frac{-2}{m-1}}} \tag{4}$$

Based on these formulas, on triangular 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 let $t=0$.

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

Step3: Improve $U^{(t)}$ by $U^{(t+1)}$ using $V^{(t)}$ and equation (4).

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

Let the unlabeled set $X=\{\tilde{X}_1,\tilde{X}_2,\tilde{X}_3,\ldots,\tilde{X}_n\}$ in the M-dimensional space R^M, let

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

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

Examples of kernel function

- Linear: $K(\tilde{X}_i,\tilde{X}_j)=\tilde{X}_i^T\tilde{X}_j$
- Polynomial:
  $K(\tilde{X}_i,\tilde{X}_j)=(\gamma\tilde{X}_i^T.\tilde{X}_j+c)^d, \gamma>0, d\in N$
- Sigmoid: $K(\tilde{X}_i,\tilde{X}_j)=\tanh(\gamma\tilde{X}_i^T.\tilde{X}_j+c)^d, \gamma>0$
- 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 kernel fuzzy c means clustering algorithm divides X into c fuzzy subsets by minimizing the 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:

$$V_i^L = \frac{\sum_{k=1}^{n} u_{ik}^m K(\tilde{X}_k,\tilde{V}_i) X_k^L}{\sum_{k=1}^{n} u_{ik}^m K(\tilde{X}_k,\tilde{V}_i)} \tag{5}$$

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

$$V_i^R = \frac{\sum_{k=1}^{n} u_{ik}^m K(\tilde{X}_k,\tilde{V}_i) X_k^R}{\sum_{k=1}^{n} u_{ik}^m K(\tilde{X}_k,\tilde{V}_i)} \tag{7}$$

$$u_{ik} = \frac{\left(1-K(\tilde{X}_k,\tilde{V}_i)\right)^{\frac{-1}{m-1}}}{\sum_{j=1}^{c}\left(1-K(\tilde{X}_k,\tilde{V}_j)\right)^{\frac{-1}{m-1}}} \tag{8}$$

Based on these formulas, on triangular 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 let $t=0$.

Step2: Calculate cluster centers $V^{(t)}=\{\tilde{V}_i^{(t)} | 1<i\le c\}$ using $U^{(t)}$ and equations (5), (6) and (7).

Step3: Improve $U^{(t)}$ by $U^{(t+1)}$ using $V^{(t)}$ and equation (8).

Retrieval Number: J11400881019/19©BEIESP
DOI: 10.35940/ijitee.J1140.0981119
Journal Website: www.ijitee.org

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

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

$$\gamma = \left(\frac{\sum_{k=1}^{n} d\left(\tilde{X}_k, \bar{W}\right)}{n} \right)^{-1} \text{ with } \bar{W} = \left(\bar{W}^L, \bar{W}^C, \bar{W}^R\right) \text{ is}$$

the arithmetic mean, where

$$\bar{W}^L = \frac{\sum_{k=1}^{n} X_k^L}{n}$$

$$\bar{W}^L = \frac{\sum_{k=1}^{n} X_k^C}{n}$$

$$\bar{W}^L = \frac{\sum_{k=1}^{n} X_k^R}{n}$$

IV. NUMERICAL EXAMPLE

We run both the algorithms FCM and KFCM using the metric distance d to compare the effectiveness with a numerical example. We implement the algorithms with $m = 2$ and $\varepsilon = 0.00001$. Consider the data set D_1 given by Hung and Yang [5] consisting of 20 triangular fuzzy numbers given in "Fig. 1".

For the set D_1 given in "Fig. 1", the suitable number of clusters $c = 2$. Therefore, on the set D_1 with the number of clusters $c = 2$ we run both the FCM and KFCM algorithms. The corresponding results are given in "Fig. 3" and "Table. I".

Consider the set D_2 consisting D_1 and one more point (99.29, 100, 101.79) called outlier point shown in "Fig. 2". Now on the set D_2 with the number of clusters $c = 2$ we run both the FCM and KFCM algorithms. The corresponding results are shown in "Fig. 4" and "Table. II".

Table- I: Cluster centers in set D_1

$\left(V^L, V^C, V^R\right)$		
	FCM Algorithm	***KFCM Algorithm***
V_1	(10.901,12.068,13.221)	(11.220,12.376,13.473)
V_2	(23.628,24.891,25.712)	(23.611,24.853,24.641)

Table- II: Cluster centers in set D_2

$\left(V^L, V^C, V^R\right)$		
	FCM Algorithm	***KFCM Algorithm***
V_1	(10.901,12.068,13.221)	(11.220,12.376,13.473)
V_2	(23.628,24.891,25.712)	(23.611,24.853,24.641)

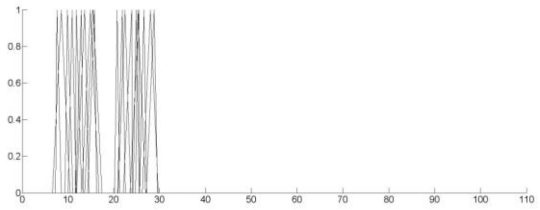

Fig. 1.Set D_1

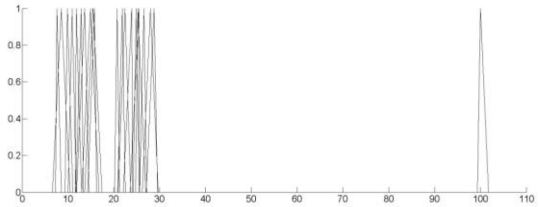

Fig. 2.Set D_2

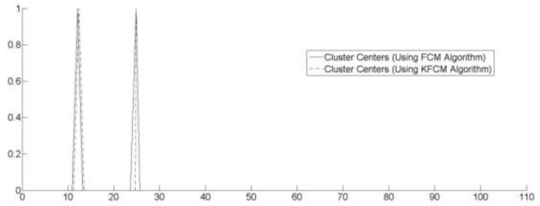

Fig. 3.Cluster centers in set D_1

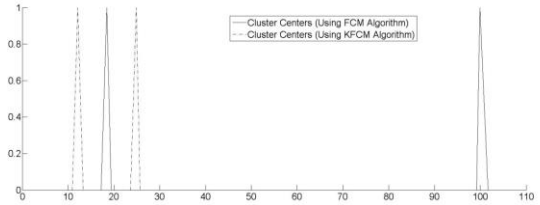

Fig. 4.Cluster centers in set D_2

When we run FCM and KFCM algorithms on data set D_1 with the distance d, the cluster centers are almost same. While running FCM algorithm on the data set D_2 (With outlier point) gives poor result, but the KFCM algorithm on D_2 provides almost the same result of D_1. That is the centers obtained by FCM algorithm on D_2 are away from the clusters whereas the centers obtained by KFCM algorithm on D_2 are within the cluster and coincides with the centers of D_1 data.

V. CONCLUSION

We use the innovative metric distance $d^2(\tilde{A}, \tilde{B})$ to find the cluster centers. We run FCM algorithm and KFCM algorithm using MATLAB. We have theoretically verified that KFCM algorithm gives better result than FCM with noisy point and outliers. KFCM performs well for the sets D_1 and D_2 with the distance $d^2(\tilde{A}, \tilde{B})$ examined in this paper.

REFERENCES

1. J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, (Plenum Press, New York, 1981).
2. C. T. Chen, "Extensions of the TOPSIS for group decision making under fuzzy environment", Fuzzy sets and systems, Vol. 114, 2000, pp. 1-9.
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. R. Herbrich, *Learning Kernel Classifier*, MIT Press, Cambridge, MA 2002.
5. 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.
6. K. R. Muller, S. Mika, G. Ratsch, K. Tsuda and B. Scholkopf, "An introduction to kernel-based learning algorithms", IEEE Transaction on Neural Networks 12(2), 2001, pp. 181-201.
7. B. Scholkopf, A. Smola, K. R. Muller, "Nonlinear Component analysis as a kernel eigenvalue problem", Neural Computation 10, 1998, pp. 1299-1319.
8. LAN, Rong and FAN Jiu-lun, "A Fuzzy C-means Type Clustering Algorithm on Triangular Fuzzy Numbers", Sixth International Conference on Fuzzy Systems and Knowledge Discovery, 2009.
9. M. S. Yang and C. H. Ko, "On a Class of fuzzy c Numbers Clustering Procedure for Fuzzy Data", Fuzzy sets and systems, 1996, 84(1), pp. 49 - 60.
10. D. Zhang and S. Chen, "Fuzzy clustering kernel method", in: Proc of the Internat. Conf. on Control and Automation, 2002, pp. 123-127.
11. S. Zhou and 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: J11400881019/19©BEIESP
DOI: 10.35940/ijitee.J1140.0981119
Journal Website: www.ijitee.org

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