e-ISSN : 2349-9745
p-ISSN : 2393-8161
Scientific Journal Impact Factor (SJIF): 1.711
International Journal of Modern Trends in Engineering and Research
www.ijmter.com

A Review on Fuzzy C-Mean Clustering Algorithm

Raulji Jitendrasinh G
Department of Computer and Science Engineering, Parul Institute of technology

Abstract—Data mining is a process which finds useful patterns from large size of data. And Data mining commonly involves four classes of tasks or techniques Classification, Clustering, Association Rule Mining, and Regression. Among these all the Clustering is a task of assigning a set of objects into groups called clusters. The clustering algorithms can be having two categories; hard clustering and soft (fuzzy) clustering. In the Hard clustering data's are divided into different clusters, where each data element belongs to only one cluster. In soft clustering, data elements belong to more than one cluster, and associated with each element by the levelt of membership. Here, in this paper represent a survey on fuzzy c means clustering algorithm.
Keywords- Soft Clustering,Hard Clustering,FCM.

I. INTRODUCTION

Clustering algorithms play an important role in discovering useful knowledge from large databases. The goal is that the objects in a group will be similar (or related) to one other and different from (or unrelated to) the objects in other groups. Clustering is a mathematical tool that attempts to discover structures or certain patterns in a dataset, where the objects have a certain degree of similarity inside each cluster. It can be achieved by various algorithms. Cluster analysis is a repetitive process of knowledge discovery. It will require modifying parameter and preprocessing until the result achieves the desired properties.

In Clustering, one of the most widely used algorithms is fuzzy clustering algorithms. Fuzzy set theory was first proposed by Zadeh in 1965 & it gave an idea of uncertainty of belonging which was described by a membership function. For each of the clusters,the data points are assigned membership values and fuzzy clustering algorithm allow the clusters to grow into their natural shapes. The fuzzy clustering algorithms can be divided into two types one is Classical fuzzy clustering algorithms and other is Shape based fuzzy clustering algorithms [1]. Classical fuzzy clustering algorithms can be divided in to three Categories.I) The Fuzzy C-Means algorithm II) The Gustafson-Kessel algorithm III) The Gath-Geva algorithm. Shape based fuzzy clustering algorithm can be divided into I) Circular shape based clustering algorithm II) Elliptical shape based clustering algorithm III) Generic shape based clustering algorithm.This paper, represent a review on fuzzy c means, and enhanced version of fcm such as pcm, fpcm and their advantages and disadvantages of real time applications.

II. FUZZY C MEAN ALGORITHM

Fuzzy clustering is a powerful unsupervised method for the analysis of data and construction of models. In many situations, fuzzy clustering is more natural than hard clustering. Objects on the boundaries between several classes are not forced to fully belong to one of the classes, but rather are assigned membership degrees between 0 and 1 indicating their partial membership. Fuzzy c-means algorithm is most widely used. Fuzzy c-means clustering was first reported in the literature for a special case (m=2) by Joe Dunn in 1974. It can be improved by Bezdek in 1981[3]. The FCM employs fuzzy partitioning such that a data point can belong to all groups with different membership grades between 0 and 1[2].

1. Initialize $U=[u_{ij}]$ matrix, $U^{(0)}$
2. At k-step: calculate the centers vectors $C^{(k)}=[c_j]$ with $U^{(k)}$

$$C_i = \frac{\sum_{j=1}^{n} u_{ij}^{m} x_j}{\sum_{j=1}^{n} u_{ij}^{m}}$$

3. Update$U^{(k)}$,$U^{(k+1)}$
4. $$d_{ij} = \sqrt{\sum_{i=1}^{n} (x_i - c_i)}$$

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{d_{ij}}{d_{kj}}\right)^{2/(m-1)}}$$

5. $if \parallel U(k+1) - U(k) \parallel < \epsilon$ $then$ STOP; otherwise return to step 2.

Here m is any real number greater than 1,
uij is the degree of membership of *xi* in the cluster *j*,
xi is the *i*th of d-dimensional measured data,
cj is the d-dimension center of the cluster,

This algorithm works by assigning membership to each data point corresponding to each cluster center on the basis of distance between the cluster center and the data point. More the data is near to the cluster center more is its membership towards the particular cluster center. Clearly, summation of membership of each data point should be equal to one. After each iteration membership and cluster centers are updated according to the formula.

Advantages
1) Converges
2) Unsupervised

Limitations:
1) Long computational time
2) Sensitivity to the initial guess (speed, local minima)
3) Sensitivity to noise and One expects low (or even no) membership degree for outliers (noisy points).

III. POSSIBILISTIC C-MEANS(PCM)

To overcome difficulties of the fcm, Krishnapuram and keller roposed a new clustering model named Possibilistic c-Means (PCM)[7].
Algorithm
Fix the number of clusters C; fix m,1<m<∞;\
Set iteration counter l=1;
Intialize the possiblistic C-parttion U(0);
Estimate ηi
Repeat
Update the prototypes using U(l), as indicated bel ow;
Compute U(l+1)
Increment l;
Until (|| U(l-1)-U(l)||<ε);
Set iteration counter l=1;
Reestimate ηi
Repeat protypes using U(l), as indicated below;

Compute U(l+1)
Increment l;
Until (|| U(l-1)-U(l)||<ε);
ηi –determines distance at which the membership value of a point in a cluster becomes 0.5.

$$\eta_i = K \frac{\sum_{j=1}^{N} u_{ij}^{m} d_{ij}^{2}}{\sum_{j=1}^{N} u_{ij}^{m}}$$

Advantage
Clustering noisy data samples
Disadvantages
1) Very sensitive to good initialization
2) Coincident clusters may result
Because the columns and rows of the typicality matrix are independent of each other Sometimes this could be advantageous (start with a large value of c and get less distinct clusters)

IV. FUZZY POSSIBILISTIC C MEANS ALGORITHM(FPCM)

To overcome problems occurred in the pcm, Pal defines a clustering technique that integrates the features of both Fuzzy a Possibilistic c-means called Fuzzy Possibilistic c-Means (FPCM). Membership and Typicality's are very significant for the accurate characteristic of data substructure in clustering difficulty. An objective function in the fpcm depending on both membership and typicality's are represented as:

$$J_{FPCM}(U,T,V) = \sum_{i=1}^{c}\sum_{j=1}^{n}(u_{ij}^{m} + t_{ij}^{\eta})\, d^{2}(x_j, v_i)$$

Which of the following constraints

$$\sum_{i=1}^{c} u_{ij} = 1, \forall j \in \{1, \dots, n\}$$
$$\sum_{i=1}^{c} t_{ij} = 1, \forall i \in \{1, \dots, c\}$$

FPCM generates Memberships and possibilities at the same time, together with the usual point prototypes or cluster center for each cluster.
Advantage
1) Ignores the noise sensitivity deficiency of FCM
2) Overcomes the coincident clusters problem of PCM.
Disadvantages
The row sum constraints must be equal to one

V. CONCLUSION

The initially developed FCM makes use to determine the similarity between prototypes and data points, and it performs well only in the case of clustering encircling clusters. Furthermore, there are many algorithms are developed by based on the FCM for clustering more general dataset. According to this review, there are some points that can be further improvement in the future using advanced clustering technique to achieve more efficient accuracy in the result and reduce the time taken for data and/or information extract from large dataset.

VI. ACKNOWLEDGMENTS

With the cooperation of my guide, I am highly indebted to Asst. Prof. ANKUR N SHAH, for his valuable guidance and supervision regarding my topic as well as for providing necessary information regarding review paper.

REFERENCES

[1] Bassam M. El-Zaghmouri, Marwan A. Abu-Zanona" Fuzzy C-Mean Clustering Algorithm Modification and Adaptation for Applications ",World of Computer Science and Information Technology Journal (WCSIT),year 2012.
[2] R.Suganya, R.Shanthi"Fuzzy C- Means Algorithm- A Review", International Journal of Scientific and Research Publications, Volume 2, Issue 11, November 2012 1 ISSN 2250-3153
[3] Deepak Kumar Niware Dr. Setu Kumar Chaturvedi"Web Usage Mining through Efficient Genetic FuzzyC-Means",(IJCSIS) International Journal of Computer Science and Information Security,Vol. 11, No. 9,year 2013.
[4] Ming-Chuan Hung and Don-Lin Yang " An Efficient Fuzzy C-Means Clustering Algorithm", IEEE,2001.
[5] R. Krishnapuram amd J.M. Keller "A possibilistic approach to clustering", IEEE Trans. Fuzzy Systems, Vol. 1, Pp. 98-110,1993.
[6] Rui Xu, Student Member, IEEE and Donald Wunsch II, Fellow, IEEE" Survey of Clustering Algorithms" IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 16, NO. 3,2005.
[7] Timothy C. Havens, Senior Member, IEEE, James C. Bezdek, Life Fellow,IEEE, Christopher Leckie," Fuzzy c-Means Algorithms for Very Large Data", IEEE TRANSACTIONS ON FUZZY SYSTEMS, VOL. 20, NO. 6,2012.
[8] Tejwant Singh, Mr. Manish Mahajan," Performance Comparison of Fuzzy C Means with Respect to Other Clustering Algorithm", IJARCSSE ,VOLUME 4,2014.

