ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 2, February 2018

Fuzzy C- Means Algorithm for Clustering

[1] Fasi Ahmed Parvez, [2] Asiya
[1] Associate Professor and HOD in BITS, Telangana, India.
[2] Assistant Professor, Department of Computer Science and Engineering, BITS, Telangana, India.

***Abstract* - Clustering is a method of grouping the objects into clusters. In general, the clustering algorithms can be classified into two categories namely hard clustering and soft (fuzzy) clustering. In hard clustering, each data point either belongs to a cluster completely or not. In case of soft clustering techniques, fuzzy sets are used to cluster data, so that each point may belong to two or more clusters with different degrees of membership. Fuzzy C Means (FCM) is a very popular soft clustering technique, and similarly, K-means is an important hard clustering technique In this paper we represent a survey on fuzzy c means clustering algorithm. These algorithms have recently been shown to produce good results in a wide variety of real-world applications.**

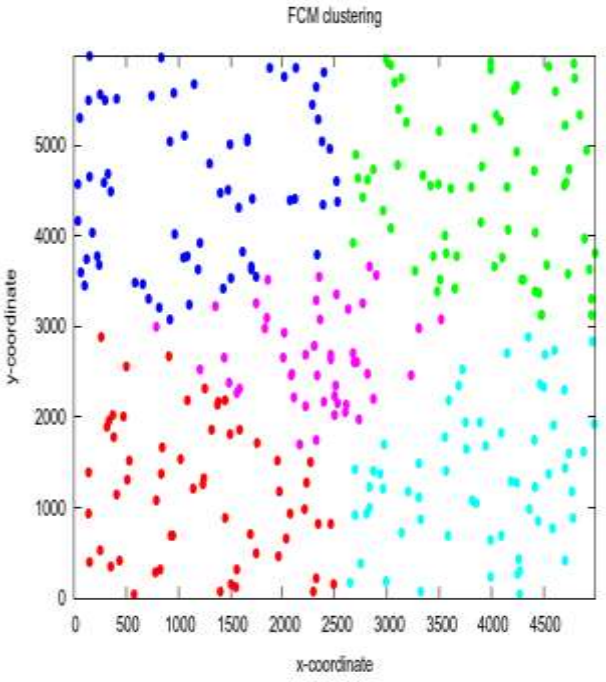
1. INTRODUCTION

Data clustering is identified as a vital area of data mining. It is the process of classify data elements into different groups (known as clusters) where the elements within a group possess high similarity while they differ from the elements in a different group. Clustering process follow the following two properties: 1) High Intra cluster property and 2) Low inter cluster property.

2. FUZZY C MEANS CLUSTERING:

The fuzzy clustering algorithms can be classified into two types 1) Classical fuzzy clustering algorithms 2) Shape based fuzzy clustering algorithms. Classical fuzzy clustering algorithms can further be classified into three types.1) The Fuzzy C Means algorithm 2) The Gustafson Kessel algorithm 3) The Gath Geva algorithm. Similarly, Shape based fuzzy clustering algorithm can be divided into 1) Circular shape based clustering algorithm 2) Elliptical shape based clustering algorithm 3) Generic shape based clustering algorithm. Fuzzy C-means (FCM) is a data clustering technique where each data point belongs to a cluster to some degree that is specified by a membership grade. This technique was originally introduced by Jim Bezdek in 1981 [4] as an improvement on earlier clustering methods. It provides a method of how to group data points that populate some multidimensional space into a specific number of different clusters. The main advantage of fuzzy c – means clustering is thatit allows gradual memberships of data points to clusters measured as degrees in [0,1]. Since the absolute membership is not calculated, FCM can be exceptionally quick because the no of repetition required to achieve a targeted clustering exercise corresponds to the accuracy required.

In this algorithm, data are bound to each cluster by means of a Membership Function, which represents the fuzzy behavior of the algorithm. Based on the distance between two data points, the clusters are formed in this research work.



In each repetition of the FCM algorithm, the following function p is minimized:

$$p=\sum_{i=1}^{n}\sum_{j=1}^{k} u_{ij}^{m}\| x_i - k_j \|^2$$

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 2, February 2018

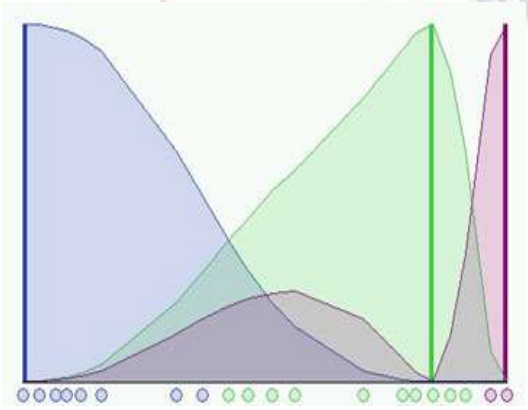
Here, *n* is the no of data points, k is the no of clusters required; the centre vector for cluster *j is* k_j, and u_{ij} membership degree for the i^{th} data point *xi* in cluster *j*. The closeness of the data point *xi* to the centre vector *kj* of cluster j is measured by ‖*xi*−*kj*‖ partitioning of the fuzzy is done through an effective iteration of the objective function based on above, membership u_{ij} that is updated and the cluster centers c_j by:

$$u_{ij} = \frac{1}{\sum_{k=1}^{C}\left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|}\right)^{\frac{2}{m-1}}}$$

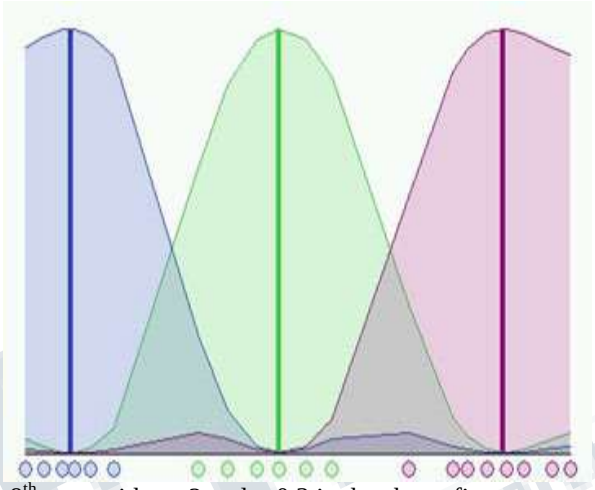
$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^{m} \cdot x_i}{\sum_{i=1}^{N} u_{ij}^{m}}$$

This repetition will stop when $\max_{ij}\left\{\left|u_{ij}^{(k+1)} - u_{ij}^{(k)}\right|\right\} < \varepsilon$, where ε is a termination criteria range from 0 to 1, with k repetitive steps.

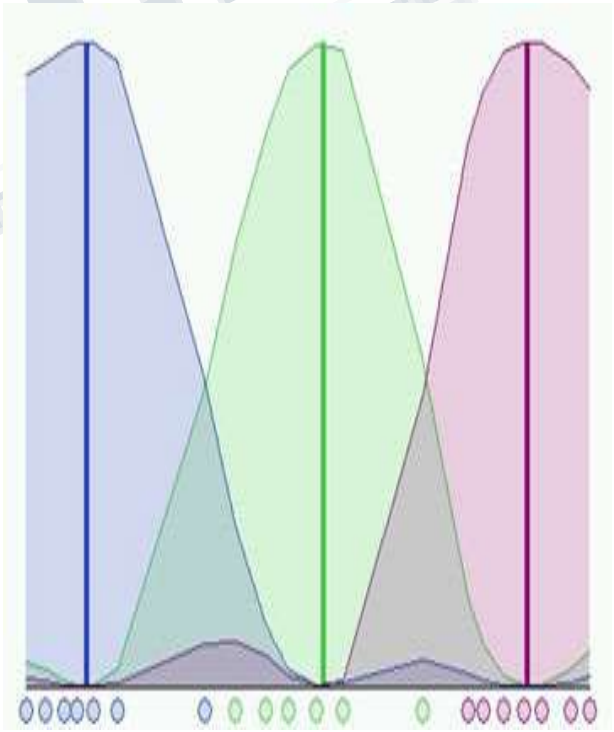
Let us consider the simple example to how the application of the FCM. Initializing the algorithm is done with 20 objects and 3 clusters to compute the U matrix. Figures below show the membership value for each data and for every cluster. The data color has the nearest cluster based on the membership function.



in the simulation shown in the figure above we have used a fuzziness coefficient m = 2 and we have also imposed to terminate the algorithm when $\max_{ij}\left\{\left|u_{ij}^{(k+1)} - u_{ij}^{(k)}\right|\right\} < 0.3$ .. The picture shows that based on specific position of clusters fuzzy distribution depends. No step is performed yet so that clusters are not identified very well. Algorithm can be run until it encounters stop condition. The final condition reached at the



8^{th} step with m=2 and =0.3 in the above figure.



Bigger computational effort is required to get higher accuracy. The next figure shows better result with same

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 2, February 2018

and =0.01, but we needed 37 steps It is also important to notice that different initializations cause different evolutions of the algorithm. In fact it could converge to the same result but probably with a different number of iteration steps.

Advantages

1) Gives best result for overlapped data set and comparatively better then k-means algorithm.

2) Unlike k-means where data point must exclusively belong to one cluster center here data point is assigned membership to each clcluster center as a result of which data point may belong to more than one cluster center.

CONCLUSION

The experimental results show that the FCM algorithm is efficient for smaller data set and for smaller number of clusters. It can be noticed that the time taken for 5 clusters is less than the time taken for 10 clusters.

FCM clustering which constitute the oldest component of software computing, are really best for handling the problems related to understandability of patterns, incomplete or noisy data, , human interaction and it can provide approximate solutions quickly. They have been mainly used for discovering association rules and functional dependencies as well as image retrieval. The increase in the number of data points also increases the execution time. The execution time of FCM clustering algorithm for arbitrary data points depends only on the number of clusters and not on the data points and vice versa. The distance between data points and some shape of the distribution, has the effect on the performance and behavior

REFERENCES:

1. A. K. Jain, M. N. Murty and P. J. Flynn, "Data Clustering: A review",ACM Computing Surveys, vol. 31, no. 3, 1999.

2. V. S. Rao and Dr. S. Vidyavathi, "Comparative Investigations and Performance Analysis of FCM and MFPCM Algorithms on Iris data", Indian Journal of Computer Science and engineering, vol.1, no.2, 2010pp. 145-151

3. Y. Yong, Z. Chongxun and L. Pan, "A Novel Fuzzy C means Clustering Algorithm for Image Thresholding", Measurement Science Review, vol. 4, no.1, 2004

4. J. C. Bezdek, "Pattern Recognition with Fuzzy Objective Function Algorithms", New York: Plenum Press, 1981

5. Hopner , K, R., Runkler, 1999 "Fuzzy Cluster Analysis", John Wily & sons.

6. Berkhin P, "Survey of Clustering Data Mining Techniques", Technical Report,

Authors profile:

Mr. FASI AHMED PARVEZ is Associate Professor and HOD in BITS, Telangana India. He is pursuing his Ph.D. in data mining. He has more than 12 years of experience in the field of teaching engineering students. He has published more than 20 papers in International Journals. He has presented 6 research papers at various National/International conferences. He is member of ISTE. His research areas include mining, data base, programming languages

Ms. Asiya is Assistant Professor in the department of computer science and Engineering, BITS, Telangana, India. She has over 2+ years experience and her field of interest include mining, IOT, Cloud