Studies on Clustering and Fuzzy Clustering

K.Velusamy[#], A. Sakthivel[*], M. Jayakeerthi[#]

[#]*DIM3 Technology and Solutions-Nehru College of Arts and science*
Coimbatore, Tamilnadu, India
[*]*Periyar University College of Arts and Science*
Pennagaram, Tamilnadu, India

Abstract— **In recent years, data mining is widely preferred area by researcher for discovering new knowledge and it has various concepts in that. The clustering is most popular concept in data mining and getting new knowledge from hidden data structure. Databases having lot of meaningful information and stored in unordered as table structure. Clustering can give useful reports based on requirements from unlabeled data. In this paper studies about information of clustering, fuzzy clustering and describing their merits and demerits.**

Keywords— **Data Mining, Clustering, Soft Computing, Fuzzy Logic, Fuzzy Clustering,**

I. INTRODUCTION

Data Mining is a tool that is mining the knowledge from the large amount of database and [1] referred to as KDD. There is massive amount of data available in real time environment and that data is not useful information without converted into make useful information.

Data Cleaning, Data Integration, Data Transformation, Data Mining, Pattern Evaluation and Data Presentation are steps for takeout information from database for data mining..

II. CLUSTERING

Clustering is a foremost responsibility notion of data mining and it is task of separating piece of information points into homogeneous module so that items is similar means those points put into same class and different means those points put into different classes. The term "similarity" should be understood as mathematical similarity, calculated in a number of well-defined intelligence. In metric spaces, similarity is often defined by means of a distance.

Clustering is an unsupervised [2] learning which is finding groups from unlabeled data and the class-prediction is done on unlabeled facts after a supervised learning on pre-labelled data. Clustering can find different types of similarity measures may be used to identify classes depending on the data and its application, in clustering, information is divided into crisp clusters and each data belongs to exactly one cluster.

Hard clustering methods are based on classical set theory, and require that an object either does or does not belong to a cluster. It is a separation of data into a specified number of mutually exclusive subsets.

The four main classes of clustering algorithms available in the literature are partitioning methods, hierarchical methods, density-based clustering and grid-based clustering. Clustering adds to the value of living databases by helpful

secreted relationships in the data. Clustering also called as the hard clustering, data is divided into distinct clusters.

A. Characteristic of Clustering

- When clustering high dimensional objects, the accuracy and efficiency of hard clustering have been very poor. Because objects may belong to different cluster in different subspaces comprised of different combinations of dimensions. Sensitive to initial centers
- Use other heuristics to find good initial centers
- Specifying the number of centers very subjective
- It is sensitive to outliers and not gives accurate result
- Initial centroid value has been calculated.
- Inability to recover from database corruption
- Complexity

III. FUZZY LOGIC

Soft computing differs from conventional computing in that, unlike hard computing, it is tolerant of imprecision, uncertainty, partial truth, and approximation. In effect, the role model for soft computing is the human mind.

Fuzzy systems are appropriate for vague or fairly accurate reasoning, particularly for the system with a mathematical model that is difficult to derive. Fuzzy logic allows decision making with predictable values under vague information.

IV. FUZZY CLUSTERING

Fuzzy Logic is a compilation of notebooks and packages and that are planned to bring in fuzzy set theory with fuzzy logic in the mathematical area and which is provides a powerful tool for studying fuzzy logic and developing fuzzy applications in many area.

In fuzzy clustering, the [3] data points can belong to multiple clusters, and associated with each of the elements are membership marks which point out the degree to which the data points fit in to the unlike clusters.

Fuzzy clustering is now a mature and vibrant area of research with highly innovative advanced applications. Encapsulating this through presenting a careful selection of research contributions, this book addresses timely and relevant concepts and methods, whilst identifying major challenges and recent developments in the area.

Fuzzy clustering methods [4] allow the 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. The discrete nature of the hard partitioning also causes difficulties with algorithms based

on analytic functional, since these functional are not differentiable.

Fuzzy clustering is a progression of conveying these membership levels, and then [5] using them to assign data elements to one or more clusters. One of the most widely used fuzzy clustering algorithms is the Fuzzy max-min cluster.

A. Characteristic of Fuzzy Clustering

- The accuracy and efficiency of soft clustering have been very efficient when clustering high dimensional objects
- Always converges and [6] Long computational time
- Sensitivity to the initial guess (speed, local minima)
- Sensitivity to noise
- Low membership degree means outliers
- Initial centroid value will calculated randomly.

V. DISCUSSIONS

The clustering methods are used in many real time application and fuzzy clustering also used. Clustering is incorporated with fuzzy logic is called fuzzy clustering. The above both scheme easily go down into local optima, which are solved by optimization techniques. The very recognizable method for clustering is K-Means algorithm and fuzzy clustering algorithm is improved k means such as Fuzzy C Means.
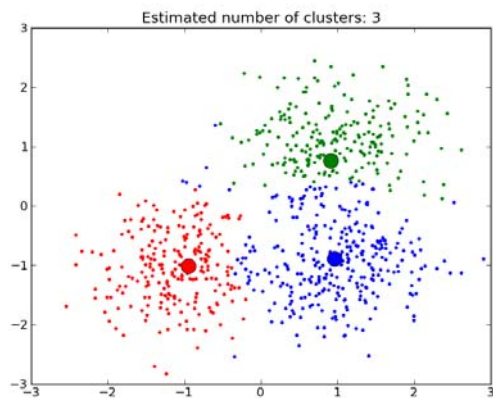

Fig. 1 Clustering

Fig. 2 Fuzzy Clustering

VI. CONCLUSIONS

This paper presents the studies characteristic about fuzzy clustering and hard clustering. This paper shows the both merit and demerit of fuzzy and hard clustering method. It is also described basic notation of both fuzzy clustering and clustering and their prominent algorithms.

REFERENCES

[1] K. Velusamy and R.Manavalan , "Performance Analysis of Unsupervised Classification based on Optimization," *International Journal of Computer Applications (0975 – 8887, Volume 42– No.19, March 2012.*

[2] Raju. G,Thomas.B, Tobgay. S and Kumar T.S, "Fuzzy Clustering Methods in Data Mining: A Comparative Case Anaylisis," International Conference on Advanced Computer Theory and Engineering, 2008. ICACTE '08 (489-493),ISBN : 978-0-7695-3489-3.

[3] Dilip kumar praihar and sanjib Chandra de sharker , "A comparative study of fuzzy c means algorithm and entropy based fuzzy clustering algorithms," computing and informatics, vol 30,2011, 701-720.

[4] S.P Priyadharshini and Ramachandra V. Pujeri, "Performance of Fuzzy Clustering based Optimization ," International Journal of Advances in Engineering & Technology, May 2014, ISSN 22311963.

[5] Jianxiong Yang and Junzo Watada, "fuzzy clustering analysis of datamining : application of an accident mining system ," International journal of innovative computing, information and control icic international 2012-ISSN 1349-4198.

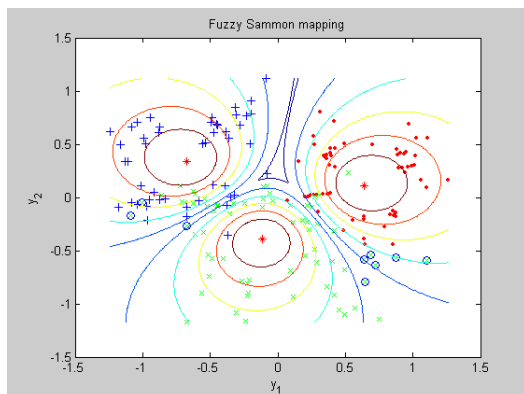[6] J.Valente de Oliveria and W.Pedrycz, "Advances in fuzzy clustering and its applications", John Wiley @ sons, ISBN 978-0-470-02760-8