A Fuzzy Rule Based Clustering Development Novel

Sachin Ashok Shinde[1], Seema Singh Solanki[2]

[1]ME 2nd Year, Computer Science & Engineering Department Dr. Seema Quadri College of Engineering & technology, Aurangabad, India

[2]Assistant Professor, Computer Science & Engineering Department, Dr. Seema Quadri College of Engineering & technology, Aurangabad, India

Abstract: *A cluster is collection of data objects that are similar to one another with the same cluster and not similar with other cluster. Also it is study on fuzzy rule based clustering development novel.So this can say about the clustering, it is nothing but grouping of set of physical objects into the classes of similar objects. The fuzzy rule based clustering is the crisp clustering when the boundaries among the cluster are vague and ambiguous. Up to yet the cluster never gets identified by the human directly but it was possible for the machines or system to identify cluster easily as per the requirements of dataset. The cluster which is fuzzy in nature is difficult to understand. The most limitations of fuzzy and crisp clustering algorithm are there sensitivity to number of potential cluster and their initial position. The clustering is not easy to understand for the human up to yet. These will be the ideas behind concept of this fuzzy clustering to make it possible understand to the human, And also to make the crisp and boundaries easy for the cluster. The idea behind this is developments of rule based algorithm for human to understand of the cluster. The accuracy of the finding cluster should be the maintain. This will be another attempt to make it possible.*

Keywords: Clustering, fuzzy, boundaries, data mining, crisp, Fuzzy clustering, K means, C means,

1. Introduction

Now a day**,** fuzzy models clustering is going to used widely because they are having the capability of to work with imprecise data for to handle nonlinear problems and acquired knowledge with these models is more interpretable than the black- box models.[1] The constructing of rule based fuzzy model is most challenging problem. It is desirable that the rule base covers all the states of the system and at the same times. There are number of approaches have been proposed to construct the rule based from numerical data. This technique includes heuristic approaches like genetic algorithm, neuro-fuzzy techniques, clustering are methods etc.[2] Many approaches have been proposed that use clustering methods for learning fuzzy classification rules from the numerical types of data, below of this mentioned some of it.Yager and filev developed and discovered the simple method based on mountain clustering for generation of fuzzy rules. The another one developer has added new concepts regarding the fuzzy neural networks based on Takagi-Sugeno model propose by Han et al.then another one zhao et al .proposed a two stage approach to extract compact takagi-sugeno fuzzy models using subtractive clustering and particle swarm optimization from numerical data. Another techniques invented by Eftekhari and katebi proposed a hybrid approach for the building optimal fuzzy model from data for nonlinear and unscented filter. Torun and Tohumonglu developed a new systematic way in order to obtain optimized fuzzy inference system for identification. Then the another new techniques developed about using the FCM to generate fuzzy rules from data to deal with the data classification problem.hossen et al .proposed a novel modified adaptive fuzzy inference system. There are number of techniques are going to be developed on fuzzy rule based algorithm related with cluster. But it wasn't able to understand to human clearly. [2] It was machine readable or system readable cluster was invented. First time these are going to develop such types of algorithm which will be understandable for human to developments of cluster and fuzzy rule based techniques.

2. Clustering

The clustering analyzes data objects without consulting a known class label. Generally class labels are not present in the training data. The objects are clustered or grouped based on the principle of the maximizing the interclass similarity and minimizing the interclass similarity. That is, cluster in the objects formed similarity with one another. Clustering techniques is about the data tuples as objects and they partition the objects into groups or cluster. Clustering is nothing but one of the important human activity, for talking about clustering takes the example of childhood, generally peoples learns to distinguish between cats and dogs, or Between animals and plants, by continuously improving the subconscious clustering schemas.[3] The cluster analysis generally used for numerous applications which includes market research, pattern recognition, data analysis and image processing. In business clustering can help marketers discover distinct group in their customer bases and characterized customer on Purchasing pattern. Clustering may also help the identification of areas of similar land use in an earth observation database and identification houses in city So the customer's data is shown with respects to customer location in city or areas, for different types of areas, which is useful to identify the location and which contains the different types of data into the three different types of groups. Each cluster is having the center .The group of cluster identifies similar types of data,With respects to its location and as per the requirements of data in to the cluster. Each group covers the nearby location with its similarity of data.[5]

A. Types of clustering

In the general life there is requirement of data in different way, for that there is requirement of clustering, according to

Paper ID: SUB14401

341

data the clustering is divided into number of parts, which are shown as below.

- **Distance Based Clustering**
  In this case data easily identify the 4 clusters into which the data can be divided by the similarity criterion is distance: two or more objects belong to the same cluster if they are "close" according to a given distance. This is called distance-based clustering, to discover cluster with arbitrary shape, this method has been developed. [6]

- **Conceptual/Exceptional Clustering**
  Another kind of clustering is conceptual clustering, which tells about two or more objects belong to the same cluster if this one defines a concept common to all that objects. In other words, objects are grouped according to their fit to descriptive concepts, not according to simple similarity measures. This can be exceptional concept so called as exceptional clustering. [6]

B. Clustering Methods

In the clustering there are lots of methods available to make the data suitable and understand to human easily. According to category of data, data have partitioned cluster in to different methods.

Partitioning Method

Partitioning method C construct a partition of a database **D** of **n** objects into a set of **k** clusters, s.t., and min sum of squared distance for the partitioning method.

$$\Sigma_{m=1}^{k}\Sigma_{t_{mi}\in Km}(C_m - t_{mi})^2$$ ………… (I)

Given D, a data set n objects, and k, the number of cluster to form a partitioning algorithm organizes the objects into k partition (k≤n).where each partitions represents the cluster. This partition will help us for the finding of partitions in cluster. [6]

Hierarchical Method

The hierarchical method Modified from Dr Seungchan Kim's slides for to make clustering suitable and easy to understand, this methods works by grouping the data objects into tree of cluster. There are two styles of hierarchical clustering algorithms to build a tree from the input set S. [8]
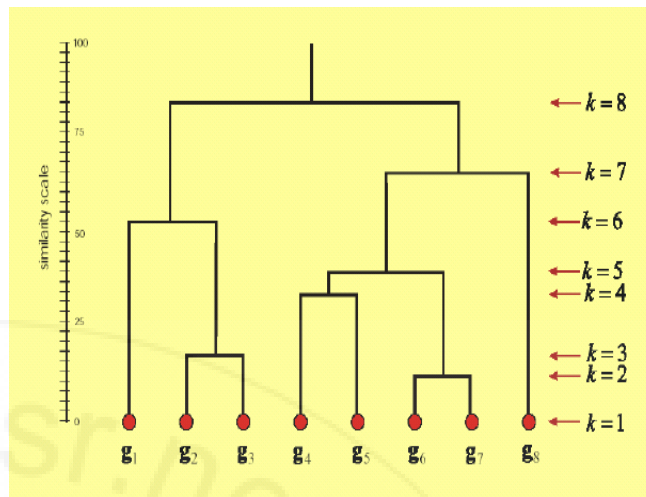

Figure 1: Hierarchical Method.

Agglomerative (bottom-up)- it is bottom up strategy starts by placing each object in its own cluster and then merges these atomic clusters into larger and larger cluster, until all of the objects are in single cluster or until certain termination conditions are satisfied. Most of the hierarchical clustering belongs to this category.

Divisive (top-down)-This top down strategy opposite process of agglomerative clustering techniques by stating with by all objects in one cluster. This techniques sub divided the cluster into small pieces until each object to be form itself or to make satisfies certain termination. In the divisive clustering condition Recursively partitioning S until singleton sets are reached. So this is the basic working of divisive techniques of cluster. [8]

3. Proposed Method
A. Fuzzy Clustering
The fuzzy clustering is nothing but set of theory and is suitable to handle problems with the vague and boundaries of the clusters. In the fuzzy clustering each object is assigned to belong to handle all of the clusters with certain degree of membership. The fuzzy clustering is superior to crisp clustering when the boundaries among the cluster are vague and ambiguous. For to maintain the limitations of both fuzzy and crisp clustering algorithms is about the number of sensitivity of cluster.[2]The fuzzy clustering is synthesis between clustering and fuzzy set theory. Our proposed approach attempts to resemble to unsupervised issues of supervised clustering. The FRBC is fuzzy rule clustering based algorithms starts to repeatedly extract the entire cluster in problems pattern. To extract each cluster the FRBC considered all the unlabeled data patterns of problems as main data have to make labeled to that data. Cluster membership is a matter of degree of the relevant cluster. The computing new invented representation of data is grouped into two parts named as part 1 and part 2. The part one shows proposed algorithms to represents the data. Each attribute of data is rescaled in interval [0 1]. [4]
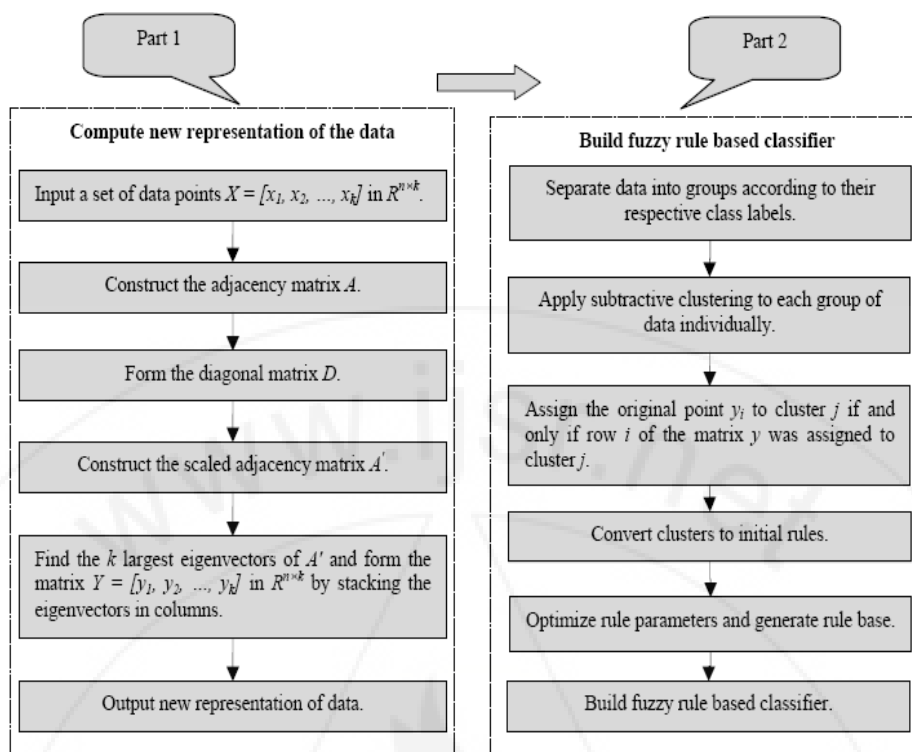
342

Figure 2: The computing Representation of data

As above figure, it is shown for the representation of data. The Data have divided into two parts, which is part 1 and part 2.the left side of data represents the part 1. This tells about the computing process of data, Which shows the adjacency matrix A and form the diagonal matrix D.[4] then it makes to representation of Eigen values of A' form of matrix. Then last stage is about the output of new representation of data. The part 2 is about the other part of data which is right hand sided

The part 2 represents the fuzzy rule based classifier. Which separate data into groups of data individually and apply for the subtractive clustering to each group of data individually. Then it assigns the original point to cluster which shows the rows and matrix was assigned to cluster. Then it converts cluster to initial rules by using the optimize rule parameter and generate rule base. Finally it builds the fuzzy rules based classifier. There is number of basic types of fuzzy clustering which is shown as below. [11]

K-means Clustering Method

The k- means clustering obtains indicator and weights of samples for the k-means process in order to develop robust k–means clustering scheme. The k means shows Input: n objects (or points) and a number k into the cluster. Randomly place K points into the space represented by the objects that are being clustered. These points represent initial group centroid. The cluster Assign each object to the group that has the closest centroid. The centroid is nothing but center point of the cluster. The k means no change in the member of all clusters. It is exclusive clustering algorithm. The figure of k-means shows as below, as assigning the value of 2 in cluster, procedure starts for assigning the value and gets some kind of outputs. As shown in figure, it assigns the object to the similar center. Then it come toward the

reassign the value and then updates the cluster means, and then it will form the mean of cluster. This will be nothing but k means clustering method. [12] So these are discussing about the first methods of clustering which is k-means which is popularly used in the clustering.
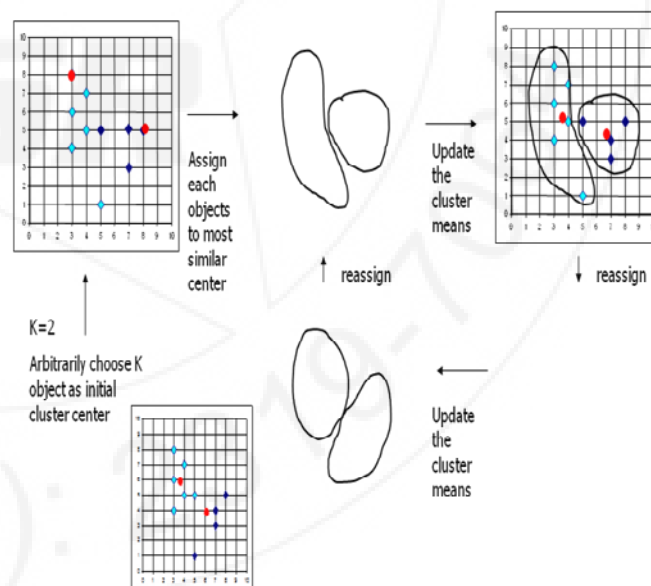

Figure 3: K-means Method Representation

C-means clustering method- Fuzzy c-means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters. This method developed by Dunn in 1973 and improved by the Bezdek in 1981 is frequently used in pattern recognition to study of the cluster. C-means is the types of clustering which shows the time requirements is high and which suffers for k means

clustering when the noise and outliers and difficulty in identifying initial portions in to clusters. It is an extension of k-means. Fuzzy c-means allows data points to be assigned into more than one cluster. Each data point has a degree of membership (or probability) of belonging to each cluster. The fuzzy c-means not a deterministic algorithm.
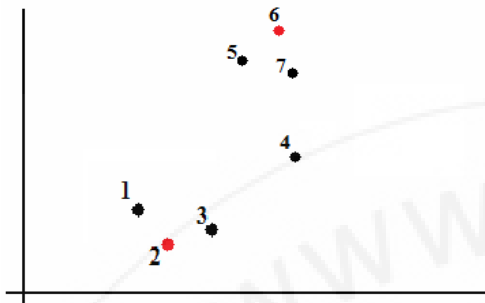

Figure 4: C-means Method Representation.

The above figure shows the c-means of data. The cluster which form to make it in connecting with each other. The figure. Exactly does this, which shows how it can be connect to points with each other are going to connect shows in the single line.[13] Point is located into the different distance and end of the clustering.
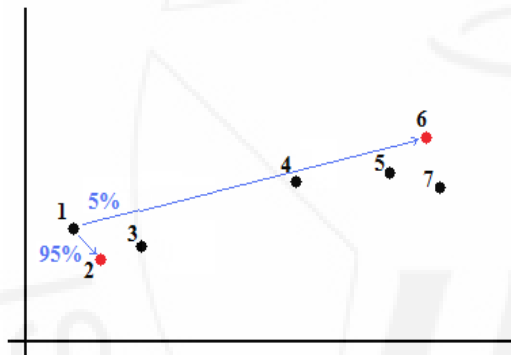

Figure 5: C-means Method Representation with connecting two points.

References

[1] E. Backer and A. Jain"A clustering performance measure based on fuzzy set decomposition'', IEEE Trans. Pattern Anal. Mach. Intell., 1981.

[2] A. Jain and R. Dubes Algorithms for Clustering Data, 1988: Prentice-Hall.

[3] B. Everitt , S. Landau and M. Leese Cluster Analysis, 2001 :Arnold .

[4] A. Rauber , J. Paralic and E. Pampalk "Empirical evaluation of clustering algorithms", J. Inf. Org. Sci., vol. 24, no. 2, pp.195 -209 2000 R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[5] A. Jain, M. Murty and P. Flynn"Data clustering: A review R. Xu and D. Wunsch"Survey of clustering algorithms", IEEE Trans. Neural Netw., vol. 16, no. 3, pp.645 -678 2005", ACM Comput. Surv., vol. 31, no. 3, pp.264 -323 1999

[6] "P. Berkhin. (2001). Survey of Clustering Data Mining Techniques. [Online]. Available:", [online] Available

[7] R. Xu and D. Wunsch"Survey of clustering algorithms", IEEE Trans. Neural Netw., vol. 16, no. 3, pp.645 -678 2005.

[8] R. Yager and D. Filev "Approximate clustering via the mountain method", IEEE Trans. Syst., Man, Cybern., vol. 24, no. 8, pp.1279 -1284 1994.

[9] R. Tibshirani , G. Walther and T. Hastie "Estimating the number of clusters in a data set via the gap statistic", J. R. Statist. Soc. *B*, vol. 63, no. 2, pp.411 -423 2001

[10] L. Wang , C. Leckie , K. Ramamohanarao and J. Bezdek "Automatically determining the number of clusters in unlabeled data sets", IEEE Trans. Knowl. Data Eng., vol. 21, no. 3, pp.335 -350 2009.

[11] R. Dubes and A. K. Jain "Validity studies in clustering methodologies", Pattern Recogniton.,vol. 11, pp.235 - 254 1979

[12] X. L. Xie and G. Beni"validity measure for fuzzy clustering", IEEE Trans. Pattern Anal. . Intell Mach., vol. 13, no. 8, pp.841 -847 1991.

[13] G. W. Milligan and M. C. Cooper "An examination of procedures for determining the number of clusters in a data set", Psychometri

Paper ID: SUB14401

344