International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-3 Issue-3, February 2014

Clustering Data Objects using Collective Fuzzy C-means Algorithm

A. Nellai Archana, K. Mohana Prasad

Abstract: Clustering is a collection of objects which are similar between them and dissimilar to the objects. Three steps to form cluster is to initialize data, select data randomly and use distance metrics to form optimal solution. In this paper fuzzy c-means algorithm is proposed. It is a kind of partitional clustering algorithm because it partition the data sets into several subsets. It allows one part of data in which it belong to two or more cluster. Data sets are taken as input value. There are many data sets so randomly select the variables to form the cluster group. It gives more efficient and effective clustering.

Index Terms: Clustering, Collective fuzzy c-means, K-harmonic means

I. INTRODUCTION

Data mining is used to extract data from large datasets and summarize into useful information. Clustering is a group of same elements occurring closely together. There are many types of clusters like exclusive, overlap, hierarchical and probabilistic cluster etc. clustering is an unsupervised classification. Clustering can be classified into five approaches and they are partitional algorithm, hierarchical algorithm, density -based method, grid-based method and model-based method. Clustering applications is used in various fields like marketing, insurance, landuse.

K-harmonic means is popular clustering algorithm technique. It is used to solve the minimal distance from the data point to the centres. K-harmonic mean(KHM) is not delicate to the election of initial cluster center. This will easily moves into local optima. Then improved gravitational search algorithm was used. This algorithm is based on the hybrid data clustering. Some seven data sets are correlated. It overwhelmed the convergence speed and it need more run time.

Effective fuzzy c-means algorithm is used to form the cluster group. It is done by using partitional clustering algorithm. It avoid selecting the random variable. The cluster group will be formed with less number of iterations.

II. RELATED WORK

Thangavel et al (2009) have developed centroid based clustering algorithm to overcome initialization. Harmonic average is used to calculate its performance function. Data object which is not close to centroid will have large weight and which is close to centroid will have small weight. Data sets such as Australian , pen digit, satellite image, image segmentation, mammography are taken to analyze the advantage and performance of k-harmonic means.

Zhou et al (2008) have introduced a method to minimize he integration error. Pattern recognition is done by k-means. The center is initialized and it move all points towards each other. It get efficient integration of data by surface overlapping. It integrate points in overlapping areas. Usage of memory and running time is efficient.

Cui et al(2005) has developed document clustering including text mining, information retrieval. It carry out fast document clustering. Local optima solution can be retrained. It produce good results in 50 iteration. It finds the local area of optimal solution. The documents will be grouped into a set of non-overlapping so it minimize the evaluation value of clustering.

Jyothi Bankapali et al (2011) have developed an algorithm that is cohesion based self merging algorithm. It means splitting and merging the component, some datasets are taken and it partition into several clusters. Cohesion criterion is improved when merging the sub cluster. Observation point which is not similar to the center are detected. By combining k-mean and hierarchical algorithm run time is analyzed. Execution time is less for k-harmonic means(KHM)

Huiqin chin et al (2011) simulated annealing(SA) means in certain time a solution which is good to be accepted than finding best possible solution. It handles the iteration process using the algorithm. Travelling sales man problem is taken then it is tested and compared with seven population based algorithm by applying the hybrid random key gravitational search algorithm. It results in more coefficient. Results proved that it can solve any kind of discrete optimization problem.

Esmat Rashedi e al (2009) gravity is a force of attraction based on the laws of gravity and mass interactions. Gravitational Search Algorithm(GSA) is compared with some heuristic method to analyze the performance. Population Based Search(PBS) algorithm was established. Exploration and Exploitation are the two aspects of this algorithm. Search space will be expanded by exploration and optimal solution will fond by exploitation. These are all done by gravity rules. Agents will be treated as objects and performance will be measured by masses and it results in high performance. Mohammad Dhoraghinejad et al (2012) to calculate the optimization problem, heuristic algorithm is used. Black hole operator is introduced to evaluate the unimodal bench marks. The performance of black hole operator is better than using GSA. The radius of black hole operator will become larger and it affects other object because of its strong gravity. It prevents from premature convergence. Unler et al (2008) Cellular manufacturing system(CMS) is an application of group technology. CMS helps to minimize cost of material by grouping machine cells. Effective performance is done by cell formation.

Manuscript published on 28 February 2014.
* Correspondence Author (s)
K.Mohana Prasad*, Department of Computer Science and Engineering, Sathyabama University, Chennai, India.
A.Nellai Archana, Department of Computer Science and Engineering, Sathyabama University, Chennai, India.

Reduction in setup time and throughput time are the benefits of CMS over traditional manufacturing system. Results are compared with the optimal solution. It improves grouping efficacy and gives same results for test problem.

III. EXISTING SYSTEM

K-means is the easiest method because it have small number of iteration. It find the cluster center by calculating the sum of squared distance. To overcome this problem of k-means, k-harmonic means(KHM) was used to solve the basic problem by changing the minimal distance from data point to the centers. If data point is near to any one center then it will give the performance result . Improved Gravitational Search algorithm was used(IGSA).

GSA gives the better results when comparing with the other methods. Every object get attracted by the force of gravity. If there occurs many object then it lose population diversity. KHM algorithm will easily get stuck into local optima but convergence speed will be faster than gravitation search algorithm. GSA will have better result in global optima. So these two methods are combined and have the merits of both IGSA and KHM. Artset1, Artset2, Iris, Glass, Cancer, CMC ,Wine are seven data sets. By assigning a value for p (for example p=2.5 ,3, 3.5…) , they find the result for KHM, PSO,PSOKHM. Using KHM(X,C) and F-measure , clustering quality was evaluated. It is effective for some kind of problems. But it needs more run time.

IV. PROPOSED SYSTEM

The proposed collective fuzzy c-means(FCM) is used to form the cluster with less number of iterations. Data sets are taken and cluster will be initialized. Normally variables will be selected randomly but here tried to avoid selecting the variable randomly. Every cluster will have different density of data points. The difference between two individual data points is evaluated to measure the conventional distance. The two data points should be taken if they are in the same group and if they near to each other. FCM is a kind of partitional clustering algorithm in which it partition into disjoint subsets.

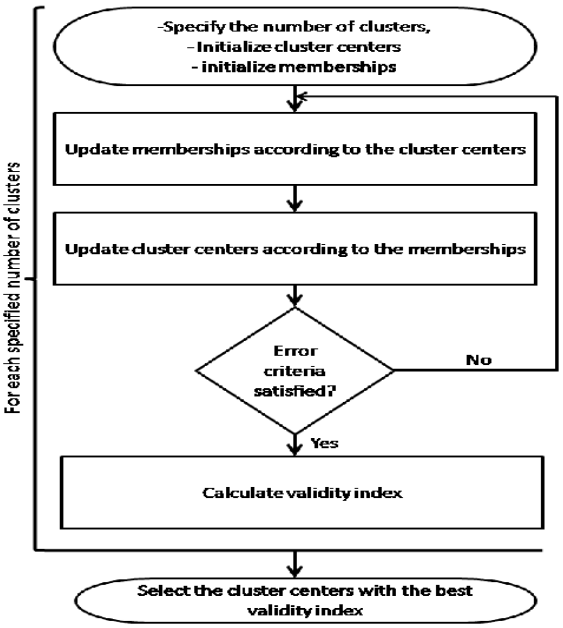

Fig 1. Flowchart diagram

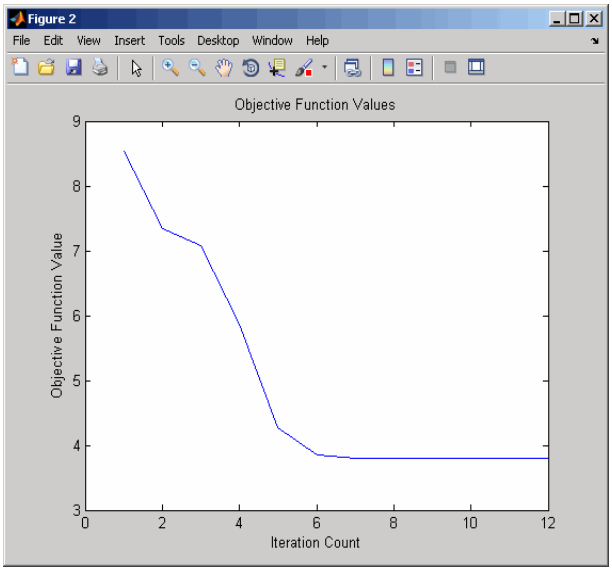

Fig 2.Random using fuzzy c-means

Distance measure will be corrected by the regulatory factor based on cluster density. If data sets are partitioned equally then it will be suitable for clustering data groups with hyperspherical shape It uses both the data set shape and the iteration results. UCI data and artificial data are operated for the two sets of experiments. The results will be proved that it shows better performance than existing methods.

V. CONCLUSION

In this paper collective fuzzy c means is used to form the cluster group by taking some data sets from UCI repository. The quality of clustering is calculated by finding the distance measures. It shows better performance when comparing with the existing algorithm and also the cluster is formed with the minimum number of iterations

REFERENCES

[1] K.Thangavel and N.Karthikeyani Visalakshi ,"Ensemble based Distributed k-Harmonic Means Clustering " International Journal of Recent Trends in Engineering, Vol 2 , No. 1 , November (2009)

[2] Zhou, H., &Lin, Y.H(2008). "Accurate integration of multi-view range images using k-means clustering" Pattern Recognition, 41(1), 152-175

[3] Cui, X.,&Potok, T.k " Document Clustering using Particle Swarm Optimization ", In IEEE swarm intelligence symposium Pasadena, California(2005)

[4] Jyothi Bankapali et al/International Journel on Computer Science and Engineering(IJCSE) , "Combining K-harmonic mean and hierarchical algorithm for robust and efficient data clustering with cohesion self-merging" ISSN:0975-3397, Vol3. No 6 June(2011)

[5] Huiqin Chen, Sheng Li and Zheng Tang,"Hybrid Gravitational Search Algorithm with Random-key Encoding Scheme Combined with Simulated Annealing"IJCSNS International Journel of Computer Science and Network Security, VOL.!! No.6,June2011

[6] Esmat Rashedi, Hossein Nezambadi-pour,Saeid Saryazdi, "A Gravitational Search Algorithm", Information Sciences 179(2009)

[7] Mohammad Doraghinejad,Mailhe maghfoori, Hossein Nezambadi-pour,"A Hybrid Algorithm Based on Gravitational Search Algorithm for Unimodal Optimization", Shahid Bahonor University of Kerman,Iran(2012)

[8] Unler, A., & Gungor, Z(2008), Applying K-harmonic means clustering to the part-machine classification problem. Expert Systems with Applications, doi:10.1016/j.eswa.2007.11.048

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