



INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

IMPROVED ROUGH FUZZY POSSIBILISTIC C-MEANS (RFPCM) CLUSTERING ALGORITHM FOR MARKET DATA

T.Buvana*, Dr.P.krishnakumari

*Research scholar, Department of Computer Science, Karpagam University, Coimbatore, India
Director, Department of Computer Applications (MCA), Rathinavel Subramaniam College of Arts and
Science, Coimbatore, India

ABSTRACT

Despite the wide variety of techniques available for grouping individuals into Market segments, K-means clustering algorithm is the popular and widely used method. The k means clustering algorithm aims to partitioning the given 'n' number of observations into k clusters to minimize an objective function. But the main shortcoming of k means clustering algorithm is its computational complexity and its performance in terms of execution time. In order to get the better performance Rough Fuzzy Possibilistic C-Means (RFPCM) algorithm is proposed. RFPCM algorithm increases the speed of grouping the clusters. This algorithm uses Rough C-Means (RCM), Fuzzy C-Means (FCM), and Possibilistic C-Means (PCM) algorithms. Further to improve the performance of RFPCM, the proposed method uses two different approaches: Ranking method and Query Redirection Method. Ranking Method facilitate to estimate the likelihood of the occurrence of data items or objects. It helps to create cluster that are having similar properties between all data point with in that cluster. The another approach is Query Redirection method which provides a mechanism to retrieve the data in the form of tables applicable to the request can be satisfied by more than one Logical table sources. RFPCM with Ranking Approach (RFPCMRA) and RFPCM with Query Redirection Approach (RFPCMQRA) are implemented in Matlab and the result shows that RFPCMQRA gives better performance compared with RFPCMRA in terms of execution time and computational complexity.

KEYWORDS: Clustering, RFPCM Algorithm, Ranking Method, Query Redirection Method.

INTRODUCTION

Clustering is a data mining technique that separates your data into groups whose members belong together. This is similar to assigning animals and plants into families where the members are alike. [1]. Clustering does not require a prior knowledge of the groups that are formed and the members who must belong to it. Clustering is the process of partitioning a given set of objects into disjoint clusters.

K- Means Clustering is a method for making groups of the data set or the objects that are having similar properties. But the process complexion of the first k-means rule is incredibly high, particularly for big information sets. Moreover, this rule leads to differing types of clusters counting on the random alternative of initial centroids [1]. Several attempts were made by researchers for improving the performance, accuracy and efficiency of the k-means algorithm. So this paper mainly focused to improving K-means algorithm efficiency.

[http:// www.ijesrt.com](http://www.ijesrt.com)

RELATED WORK

Gurjit Singh, Navjot Kaur (2013) "Hybrid Clustering Algorithm with Modifications Enhanced K-Means and Hierarchal Clustering" In this paper an improvement in K-means clustering is shown. Centroids are determined systematically so as to produce clusters with better accuracy using the Ranking Method. [1]. the second phase makes use of an efficient way for assigning data points to clusters.

Manpreet kaur, Usvir Kaur (2013) "Comparison Between K-Mean and Hierarchical Algorithm Using Query Redirection" In this paper query redirection method that improved K-means clustering algorithm performance and accuracy in distributed environment [2]. By using this approach the elapsed time is reduced and the cluster is of better quality.

Madhuri A. Dalal, Nareshkumar D. Harale, Umesh L. Kulkarni (2011) "An Iterative Improved k-means

Clustering” discuss an iterative approach which is beneficial in reducing the number of iterations from k-mean algorithm , so as to improve the execution time or by reducing the total number of distance calculations [3]. So Iterative improved K-means clustering produces good starting point for the K-means algorithm instead of selecting them randomly. And it will lead to a better cluster at the last result.

LITERATURE REVIEW

Data mining is the process of extracting previously unknown and actionable information as of large, difficult data bases. Segmentation is a key data mining technique. Segmentation is a collection of consumers that react in a related way to a particular marketing approach. So the key in segmentation is to choose how to split the database into data. There are two ways to doing this segmentation. A production can identify the personality of the segments in advance and then allocate its customers to these groups (this is known as priori segmentation) or it can use software to analyze the data and identify naturally occurring clusters of behaviors, which then form the basis of segments. This is the process of data mining known, unsurprisingly as clustering. Segmentation is made more multifarious because customer may belong to multiple segments. Over time they may also budge from one segment to another. Simply the clustering technique is more powerful than a priori segmentation as it offers more flexibility and can develop with these changes [1].

Clustering is one of the most significant methods used in market segmentation. It gives the high profit and low risk of marketing data. Selection of best segments from the data set and then classify or clustering the data into similar groups, it is difficult to conquer the best cluster. Many methods are implemented to increase the speed of grouping in clustering, proposed RFPCM algorithm is worn[2]. It motivates how to elevate the recital of the data and effective result for the high dimensional data. Its purpose is to focus a method based clustering selection technique on those aspects of the data most useful analysis and future prediction. It can reduce the execution time and computational complexity of data and may improve the accuracy and performance.

RFPCM Algorithm

The hybrid algorithm Rough Fuzzy Possibilistic C means creates by the use of K-means variations algorithms. This algorithm specifies all the categories of Hard C means, Rough C means, Fuzzy C means and Possibilistic C means. So we can say

RFPCM for numerical data performs better over other c-mean variants. Among K-means algorithms RFPCM gives improved results over other variations of k-means algorithm generated by the system. The major issues concerning data mining in large databases are efficiency and scalability. The objective function of RFPCM algorithm is very low compared with other algorithms. The objective function is

$$V_i^{RFP} = \begin{cases} w_{low} \times A1 + w_{up} \times B1 & \text{if } \underline{B}U_i \neq \emptyset \text{ and } B(U_i) \neq \emptyset \\ A1 & \text{if } \underline{B}U_i \neq \emptyset \text{ and } B(U_i) = \emptyset \\ B1 & \text{otherwise} \end{cases}$$

Improve the performance of Rough Fuzzy Possibilistic C Means algorithm used two approaches. That is RFPCM with Ranking Approach (RFPCMRA) and RFPCM with Query Redirection Approach (RFPCMQRA). Market segmentation process data set used. Create the cluster using Ranking method then execute the algorithm for checking performance and accuracy. This process same as for Query redirection method also then to compare two approaches using the parameter execution time. Calculate the time in milliseconds, the market segmentation data set to reduce the empty clusters.

RFPCM with Ranking Approach (RFPCMRA)

Ranking Method, all the jobs are arranged or ranked in the order of importance from the simplest to the hardest, or in the reverse order each successive job being higher or lower than the previous one in the sequence. It is not necessary to have job descriptions, although they are useful.[3] A common practice is to arrange all the jobs according to their requirements by rating them and then to establish the group or classification. In this method, jobs are not split up into their component parts; however comparison is made on the basis of whole jobs. So the ranking function introduces new opportunities to optimize the results of K-means clustering algorithm [1].

Search of relevant records or similar data search is a most popular function of database to obtain knowledge. There are certain similar records that we want to fall in one category or form one cluster. That's why, we need to rank the more relevance student marks by a ranking method and to improve search effectiveness. In last, related answers will be returned for a given keyword query by the created index and better ranking strategy. So I have applied this Ranking method with K-means clustering method because this method is also having the

property to find relevant records. So it is also helpful in creating clusters that are having similar properties between all data points within that cluster.

RFPCM with Query Redirection Approach (RFPCMQRA)

To have an efficient query management system most of the predefined and ad hoc queries should access summary data instead of detailed data [4]. One needs to employ query navigators to redirect base table queries to the aggregate and summary table level and examine which tables and columns were accessed and the number of rows retrieved. Check response time for these queries and any effects these have for e.g. paging or locking. Break down the predefined queries into smaller queries for processing. Consider doing most of your resource intensive processing away from the current level of detail and try to run large queries during off-peak hours. This will give more processor time to smaller queries and will aid in getting quick results [2]. Try to push query processing up from the client to the application server level. As part of end user workstation design, consider the employment of thin clients, forcing Query processing and scheduling up to the server level.

RESULTS

1. RFPCM Algorithm Cluster Creation

In this case, clusters are created in K-means clustering algorithm, using the thought of threshold value. Graph that is given below shows the amount of clusters that are made on the basis of the threshold value. On the basis of the centroid the clusters are formed. This graph is made on the basis of the values x and y, which values are taken on the both axis of the graph. The Euclidean distance is calculated between both the centroid and the data points. Each cluster is shown with different color in order to distinguish between them.

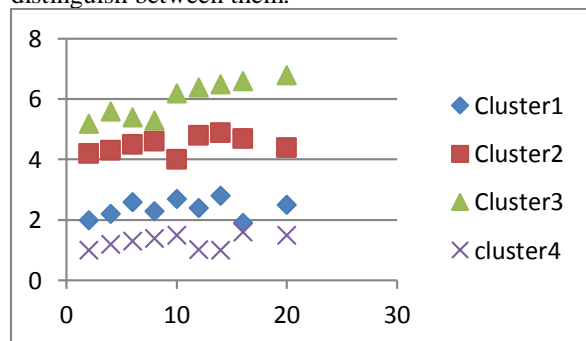


Fig 2: Graph Shows Clusters in RFPCM Algorithm

2. Performance of RFPCMRA

In this graph x-axis represents number of records and y-axis represents times that are included during ranking method when applied on clustering approach.

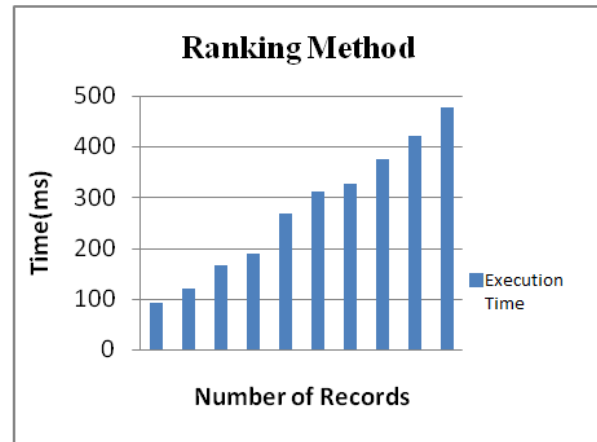


Fig 3: Performance of RFPCMRA using execution time.

The execution time for ranking method is less. So this is an appropriate approach applied for clustering method. As in case of only RFPCM for 50 records take the execution time that is 98ms, but in this case of ranking method, for the purpose of executing same number of records, it takes 91ms. And the main table that shows the execution time for the Ranking method for each particular records.

Table 1 Execution time for RFPCMRA

Number of Records	RFPCMRA execution time(ms)
50	90
100	120
150	157
200	175
250	247
300	290
350	310
400	344
450	398
500	410

3. Performance of RFPCMQRA

Execution time examination [2] for K-means clustering algorithm is done on the basis of the number of records that are measured for clustering and how much time is taken by this whole procedure. As if the number of records are 100 that are considered for clustering, then it takes execution time 62ms and so on for all records. So in this way using

different number of records, the execution time differentiation is shown.

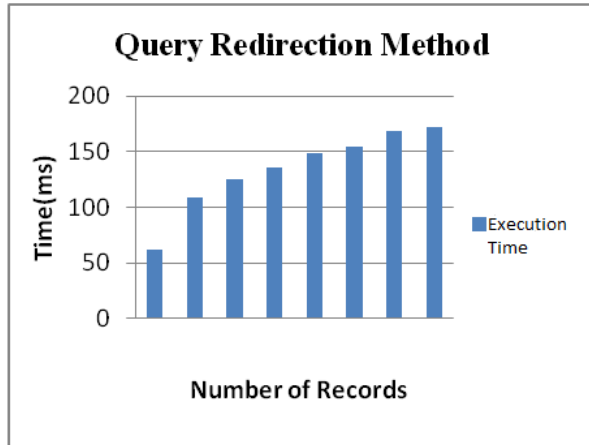


Fig4. Performance of RFPCMORA

In the table that also shows the number of records and the clustering execution time taken by K-means clustering algorithm using Query Redirection method is shown. As if the number of records are 50, the execution time will be 62ms and so on. With the help of this type of tables we can easily calculate the performance.

Table 2.Execution time for RFPCMORA

Number of Records	RFPCMORA Execution time(ms)
50	69
100	92
150	109
200	125
250	130
300	145
350	159
400	162
450	189
500	190

4. Performance Analysis for RFPCMRA with RFPCMORA

Execution time analysis for K-means clustering algorithm using Ranking and Query Redirection Method is done on the basis of the number of records that are considered for clustering and how much time is taken by this whole process. As if the number of records are 100 that are considered for clustering, then it takes execution time 72ms and so on for all records. So in this way using different number of records, the execution time differentiation is shown

Table 2.Comparison of RFPCMRA with RFPCMORA

Number of records	RFPCMRA execution time(ms)	RFPCMORA Execution Time(ms)
50	90	69
100	120	92
150	157	109
200	175	125
250	247	130
300	290	145
350	310	159
400	344	162
450	398	189
500	410	190

In the below graph, x-axis shows the no. of records and y-axis time in millisecond. It shows that time taken by RFPCM algorithm via Ranking Method for searching records is less than RFPCM algorithm via Query Redirection Method. RFPCMORA gives better performance when compared with RFPCMRA algorithm.

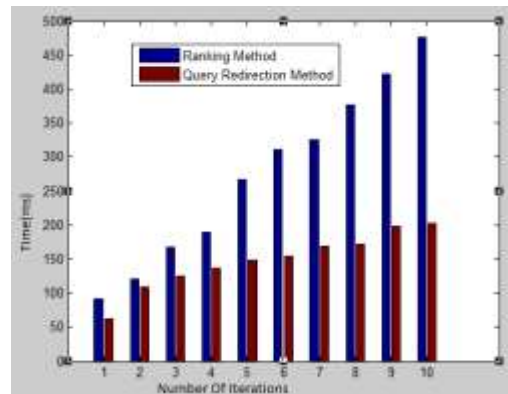


Fig 5: Comparison of RFPCMRA with RFPCMORA

CONCLUSION

Existing Improved k-means algorithm RFPCM gives reduced value of objective function for Market Segmentation process data clustering. So the RFPCM clustering algorithm for Market segmentation process data set performs better over other variations of k-means algorithm. But this algorithm does not always guarantee good results as the accuracy and efficiency of large number of market data set. Proposed applied methodologies increasing efficiency of RFPCM algorithm and Users find better results corresponding to queries and Execution time also decreased. RFPCMRA based method that improved RFPCM

clustering algorithm performance and accuracy. By using query redirection approach users get correct segments related to their search in less execution time also in distributed environment. The experimental results show that Rough Fuzzy Possibilistic algorithm with Query redirection Approach performs better than Ranking Approach. So conclude as Query Redirection Gives better performance over RFPCM algorithm.

For the future work, comparison between these algorithms can be implemented on the basis of normalization, by taking normalized and un-normalized data will give different results. So if this approach is considered, then the performance of K-means clustering algorithm is improved for large samples of data set that are also normalized in nature. Invent New Segmentations techniques for high speed clustering with minimal cost and high accuracy. Use this technique in various application areas.

REFERENCES

1. Navjot Kaur, Jaspreet Kaur Sahiwal, & Navneet Kaur "Hybrid Clustering Algorithm with Modifications Enhanced K-Means and Hierarchical Clustering", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 5, May 2013.
2. Manpreet kaur, Usvir Kaur " Comparison Between K-Mean and Hierarchical Algorithm Using Query Redirection" International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 7, July 2013.
3. Madhuri A. Dalal & Nareshkumar D. Harale "An Iterative Improved k-means Clustering" Proc. of Int. Conf. on Advances in Computer Engineering, 2011.
4. Tapas Kanungo, Senior Member, IEEE, David M. Mount Member, IEEE "An Efficient k-Means Clustering Algorithm: Analysis and Implementation" IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 24, NO. 7, JULY 2002.
5. Jaehui Park, Sang-goo Lee "Probabilistic Ranking for Relational Databases based on Correlations" ACM 2010.
6. Shi Na & et.al, "Research on k-means Clustering Algorithm", IEEE Third International Symposium on Intelligent Information Technology and Security Informatics, 2010
7. Bradley P.S., Fayyad U.M., "Refining Initial Points for KMeans Clustering", Proc. of the 15th International Conference on Machine Learning (ICML98), J. Shavlik (ed.), Morgan Kaufmann, San Francisco, 1998, pp. 91-99.
8. Chen Zhang, Shixiong Xia, "K-means Clustering Algorithm with Improved Initial center", in Second International Workshop on Knowledge Discovery and Data Mining (WKDD), pp. 790-792, 2009.
9. "Data Mining: Concepts and Techniques", Jiawei Han University of Illinois at Urbana-Champaign Micheline Kamber, Text book.
10. Tajunisha N., Saravanan V., "An increased performance of clustering high dimensional data using Principal Component Analysis", Proceedings of the IEEE first international conference on integrated intelligent computing pp 17-21, (2010).
11. Tajunisha N., Saravanan V., "An increased performance of clustering high dimensional data using Principal Component Analysis", Proceedings of the IEEE first international conference on integrated intelligent computing pp 17-21, (2010).