APPLICATION OF CLUSTERING TECHNIQUE TO SUPER MARKET DATASET FOR MOBILE DEVICES

R. R. Shelke[1], Dr. R. V. Dharaskar[2], Dr. V. M. Thakare[3]

H.V.P.M. COET, Amravati[1], Former Director,   DES (Disha – DIMAT) Group of Institutes, Raipur, CG[2], Prof. and Head, Computer Science Deptt., GB Amravati University, Amravati[3]

rajeshrishelke@rediffmail.com

.

ABSTRACT: **The rapid development and advances of information technology allow data to be accumulated faster and in much larger quantities (i.e., data warehousing). Faced with vast new information resources, scientists, engineers, and business people need efficient analytical techniques to extract useful information and effectively valuable knowledge patterns. With the help of mobile data mining technique, user will be able to visualise clustering results in mobile device any time anywhere. Clustering of large dataset for mobile device represents a very promising area for users and professionals that need to analyze data where users, resource and applications are mobile. The combined use of a clustering approach with mobile programming technologies could be used for the implementation of mobile knowledge discovery applications. Clustering of large supermarket dataset from mobile devices would be applied which will allow remote user to perform data mining tasks from mobile devices.**
Keywords: Data mining, clustering, dataset

1. INTRODUCTION

With the rapid collection of data in wide variety of fields ranging from business transactions through medical investigations to scientific research, the demands in data analysis tools are ever growing. Today's challenges are less related to data storage and information retrieval, but can rather be found in the analysis of data on a global scale in a heterogeneous information system. The technologies such as On-Line Analytical Processing, Data Mining and Knowledge Discovery in Databases all require the integration of information and efficient query processing.

Mobile computing describes a new class of portable computing devices which are becoming general in everyday life. Mobile computing is associated with data mining. A mobile computing system always bond with dynamic network connectivity caused by heterogeneous network. Data mining is key to develop the information technology. Data mining is used to build up applications and services to mobile users. Data mining plays significant roles in exploring, analyzing, and presenting scientific data. The components of data mining can exist in different types of systems involving different technologies, databases, and data structures. As an important branch and effective tool of data mining, cluster analysis is not a new area, which has already been applied to other disciplines. A Mobile computing is a set of programs that helps Multimedia to store, modify and extract information or data from a database. The goal of mobile data mining is to provide data mining techniques like clustering for the analysis and monitoring of critical data from mobile devices.

2. LITERATURE REVIEW

According to **K. Meena, M. Durairaj and K.R. Subramanian,** Efficient Data mining Techniques are required to discover useful Information and knowledge [1]. In their work, an Algorithm for incremental mining of moving patterns is introduced. This algorithm is also useful for finding a Maximal moving sequences. Also Data Allocations schemes have been devised to utilize the knowledge of user moving patterns for proper Allocation of both personal and shared Data. Hence the efficient Location Management and questions strategy would lead to beneficial output through a Mobile computing system. A mobile computing system consists of a distributed server architecture in which a service area where the server can provide service to mobile users and it contains one or many cells where a cell refers to a

communication area covered by a base station [2-3]. Qiang Yang reviewed data mining Challenges [4]. Johannes Grabmeier, Andreas Rudolph focused on Techniques of Cluster Algorithms in Data Mining [5]. Clustering algorithms have emerged as an alternative powerful meta-learning tool to analyze the massive volume of data generated by modern applications accurately. Many researchers performed survey of Clustering Algorithms [6-11] .In particular, their main goal is to categorize data into clusters such that objects are grouped in the same cluster when they are similar according to specific metrics.

3. CLUSTERING ALGORITHMS

3.1 K-MEANS CLUSTERING

It is a partition method technique which finds mutual exclusive clusters of spherical shape. It generates a specific number of disjoint, flat (non-hierarchical) clusters. Stastical method can be used to cluster to assign rank values to the cluster categorical data. Here categorical data have been converted into numeric by assigning rank value. K-Means algorithm organizes objects into k − partitions where each partition represents a cluster.
i. K-Means Algorithm Properties
- There are always K clusters.
- There is always at least one item in each cluster.
- The clusters are non-hierarchical and they do not overlap.
- Every member of a cluster is closer to its cluster than any other cluster because closeness does not always involve the 'center' of clusters.

With the development of information technology and computer science, high-capacity data appear in our lives. In order to help people analyzing and digging out useful information, the generation and application of data mining technology seem so significance. Clustering and decision tree are the mostly used methods of data mining.

4. CLUSTERING USED IN PROPOSED SYSTEM

Clustering analysis finds clusters of data objects that are similar in some sense to one another. The members of a cluster are more like each other than they are like members of other clusters. The goal of clustering analysis is to find high-quality clusters such that the inter-cluster similarity is low and the intra-cluster similarity is high. Clustering, like classification, is used to segment the data. Unlike classification, clustering models data into groups those were not previously defined. Classification models data by assigning it to previously-defined classes, which are specified in a target. Clustering models do not use a target. Clustering is useful for exploring data. If there are many cases and no obvious groupings, clustering algorithms can be used to find natural groupings. Clustering can also serve as a useful data-preprocessing step to identify homogeneous groups on which to build supervised models. In proposed system, Clustering has been applied on supermarket dataset. Clustering algorithms has been used for different tables of dataset. Clustering algorithm for product table is as follows.

4.1. PRODUCT

1. Get each unique brand names and total count for brand names from the products table
2. Convert the brand names and product count for each brand in two arrays
3. Create an array for average brand name counts
4. Provide these arrays to the JFreechart API to generate scatter graphs for the product cluster.

Similarly, clustering algorithm can be applied to other tables of dataset also.

5. EXPERIMENTAL RESULTS OF CLUSTERING TECHNIQUE

For a successful business, identification of high-profit, low-risk customers, retaining those customers and bring the next level customers to above cluster is a key tasks for business owners and marketers. Traditionally, marketers must first identify customer cluster using a mathematical mode and then implement an efficient campaign plan to target profitable customers. Only through data mining techniques, it is possible to extract

useful pattern and association from the customer data. Data mining techniques like clustering can be used to find meaningful patterns for future predictions. Customer clustering and segmentation are two of the most important techniques used in marketing and customer-relationship management. They use customer-purchase transaction data to track buying behaviour and create strategic business initiatives.

In proposed work, Clusters are created for different tables of dataset. For Product table Clusters have been formed on the basis of brand name. Number of products of particular brand has been counted and average has been calculated. After that clusters are formed scatter plot has been plotted as shown in fig . 5.1.
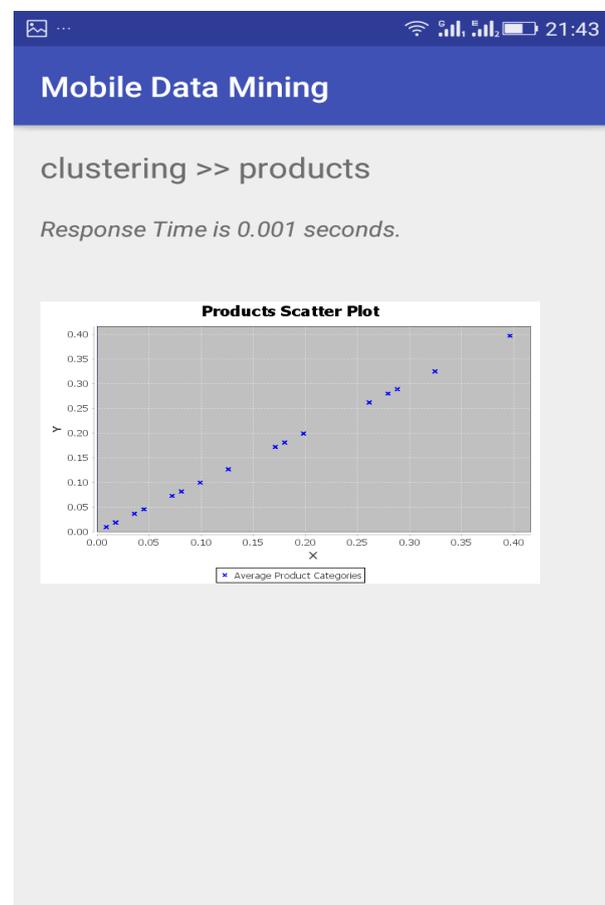

Fig 5.1 Clustering of Production

Similarly Clustering has been applied to different tables of dataset and scatter plot has been plotted.

6. CONCLUSION

Cluster analysis or clustering is the task of assigning a set of objects into groups called clusters. Main task of clustering are explorative data mining, and a common technique for statistical data analysis used in many fields, including machine learning, pattern recognition, market analysis, information retrieval, and bioinformatics. Cluster analysis itself is not one specific algorithm, but the general task to be solved. It can be achieved by various algorithms that differ significantly in their notion of what constitutes a cluster and how to efficiently find them. Cluster analysis as such is not an automatic task, but an iterative process of knowledge discovery. In this paper a clustering techniques of data mining for super market dataset has been represented. The clustering techniques are applied to different tables of database. This paper provides the major advancement in the clustering approach for data mining research using these approaches the features and categories in the surveyed work.

REFERENCES:
1. K. Meena, M. Durairaj and K.R. Subramanian, "*Data Mining Algorithm Used in a Mobile Computing Environment* "College Science in India ,ISSN 1939-2648 3 : 1 February 2009 .
2. Ashutosh K. Dubey,Ganesh Raj Kushwaha,Jay Prakash," *A Novel CSUA Based Data Mining Approach for Mobile Computing Environments.*",International Journal of Computer Applications (0975 – 8887) Volume 7, No.4, (23-26),September 2010
3. YuselSaygin, OzgurUlusoy,"*Exploiting Data Mining Techniques for Broadcasting Data in Mobile Computing Environments*." IEEE Transactions on Knowledge and Data Engineering, Volume 14, No.6, (1387-1399),November/December, 2002.
4. Qiang YANG, "*Three challenges in data mining*" , Front. Comput. Sci. China 2010, 4(3): 324–333,DOI 10.1007/s11704-010-0102-7 .
5. Johannes Grabmeier, Andreas Rudolph, "*Techniques of Cluster Algorithms in DataMining* " , Data Mining and Knowledge Discovery, 6, 303–360, 2002
6. S.R. Haskey, B.D. Blackwell, D.G. Pretty, "*Clustering of periodic multichannel time series data with application to plasma fluctuations*", Computer Physics Communications 185 (2014) 1669–1680 (ELSEVIER).
7. Aastha Joshi, Rajneet Kaur," *A Review: Comparative Study of Various Clustering Techniques in Data Mining*", International Journal of Advanced Research in Computer Science and Software Engineering ,

Volume 3, Issue 3, March 2013 ISSN: 2277 128X [Elsevier]

8. Madjid Khalilian, Norwati Mustapha,"*Data Stream Clustering: Challenges and Issues*", Proceedings of International Multiconference of Engineers and Computer Scientist, vol I,IMECS 2010.

9. Ahmed Alsayat, Hoda El-Sayed, "*Efficient Genetic K-Means Clustering for Health Care Knowledge Discovery*".

10. Amanpreet Kaur Toor, Amarpreet Singh, "*Analysis of Clustering Algorithm Based on Number of Clusters, Error Rate, Computation Time and Map Topology on Large Data Set*", International Journal of Emerging Trends & Technology in Computer Science (IJETTCS) ,Volume 2, Issue 6, November – December 2013.

11. Amritpal Singh, Prithvi Pal Singh and Sunny Sharma, "*An Inclusive Review on Data Mining Tools*", International Journal of Recent Trends in Engineering & Research (IJRTER) Volume 02, Issue 05; May - 2016 [ISSN: 2455-1457].