RESEARCH ARTICLE

Classification technique to super market dataset for mobile devices

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

[1]H.V.P.M. COET, Amravati.
[2]Former Director, DES (Disha – DIMAT) Group of Institutes, Raipur, CG
[3]Prof. and Head, Computer Science Dep't., SGB Amravati University, Amravati.

ABSTRACT

The rapid expansion and advances of information technology permit data to be accumulated more rapidly and in larger quantities. Faced with vast new information resources, scientists, engineers, and business people need efficient systematic techniques to extract useful information and effectively valuable knowledge patterns. With the help of mobile data mining technique, user will be able to visualise classification results in mobile device any time anywhere. Classification 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 combination of classification approach and mobile programming technologies could be used for the implementation of mobile knowledge discovery applications. Classification of dataset from mobile devices would be applied which will allow remote user to carry out data mining tasks from mobile devices.

Keywords: Data mining, classification, dataset

INTRODUCTION

Classification is a supervised learning technique. It maps the data into predefined clusters. It is used to build up a model that can categorize the population of records at large level. Classification algorithm requires that the classes be defined based on the data attribute value. It describes these classes according to the characteristics of the data that is already known to belong to the classes. Classification consists of two steps, training step and the testing step. In the training step the model is constructed and in the testing step classifiers are tested to see the accuracy of the classifiers. Support Vector Machine (SVM), Neural Network, Decision tree, Naive Bayesian is the types of classification algorithm.

Examples of cases where Classification can be used are as follows:

- A bank loan officer needs to analyze the information in order to know which customers (loan applicant) are risky or which are safe.
- A marketing manager at a company needs to examine a customer with a given profile, who will buy a new computer.

In both of the above examples, a classifier or model is built to predict the label categories.

These labels are risky or safe for loan application data and yes or no for marketing data. The Data Classification process includes following two steps:

I. Building the Classifier or Model
II. Using Classifier for Classification

Using data mining techniques for mobile devices, especially Smartphone is an emerging research area. Mobile devices have recently gained significant position in several communities like governmental agencies, enterprises, social service providers (e.g., insurance, Police, fire departments), healthcare, education, and engineering organizations. However, despite of significant improvement in mobile computing capabilities, still computing requirements of mobile users, especially marketing users, is not achieved. To achieve results of data mining on mobile device, one can perform data mining techniques on separate machine and visualize results on mobile device.

Literature Review:

Most of the popular data mining algorithms are designed to work for centralized data and they often do not pay attention to the resource constraints of distributed and mobile

*Corresponding Author: Mrs. R. R. Shelke, **Email:** rajeshrishelke@rediffmail.com

environments [1-2]. In support of the third generation of data mining systems on distributed and massive data, an efficient distributed and mobile algorithm has been developed by Frank Wang and Nu Helian for global association rule mining [3], which does not need to ship all of local data to one site thereby not causing excessive network communication cost [4-5]. Classification techniques have been compared by many researchers [6-9]. According to Megha Gupta and Naveen Aggarwal, Data mining is the knowledge discovery and useful information finding from the large amounts of data stored in databases [10]. It is referred to as knowledge discovery from databases (KDD), is the automated or suitable extraction of patterns representing knowledge implicitly stored in large databases. Data mining tools forecast future trends and behaviors, allowing businesses to make practical, knowledge-driven decisions. Data mining tools can give answer to business questions that traditionally were much time consuming for resolving. Classification techniques are generally used in data mining to categorize data among various classes. Classification techniques are being used in many industries to easily identify the type and group to which a particular tuple belongs.

There are many algorithms which are used for classification in data mining shown above. Following are some algorithms:

1. Rule based classifier
2. Decision tree induction
3. Nearest neighbour classifier
4. Bayesian classifier
5. Artificial neural network
6. Support vector machine
7. Ensemble classifier
8. Regression trees

One of the most difficult tasks in the whole KDD process is to choose the right data mining technique, as the commercial software tools provide more and more possibilities together and the decision requires more and more expertise on the methodological point of view. Aditi Goel and Saurabh Kr. Srivastava found that SVM is the best classifier amongst all the classifiers. They used learning algorithms with the historical dataset to train the classifier and the test samples are used to validate the correctness of the classifier [11]. Sharon Carl, Glaston D'souza and Linet Varghese focused on Implementation of Classification Algorithms and their Comparison for Educational Dataset [12].

Data Mining Techniques to Perform Market Analysis

One of the challenges for companies that have invested heavily in customer data collection is how to extract important information from their vast customer databases and product feature databases, in order to gain competitive advantage. Several features of market basket analysis have been studied in academic literature, such as using customer interest profile and interests on particular products for one-to-one marketing, purchasing patterns in a multi-store environment to improve the sales. Market basket analysis has been intensively used in many companies as a means to find out product associations and base a retailer's promotion strategy on them. Knowledgeable decision can be made easily about product placement, pricing, promotion, profitability and also finds out, if there are any successful products that have no significant related elements. Similar products can be found so those can be placed near each other or it can be cross-sold. A retailer must know the needs of customers and adapt to them. Market basket analysis is one possible way to find out which items can be put together. Market basket analysis gives retailer good information about related sales on group of goods basis and also it is important that the retailer could know in which channel and in which region the products can be sold more and which session.

Classification Techniques Analysis

Classification algorithm requires that the classes be defined based on the data attribute value. It describes these classes according to the characteristics of the data that is already known to belong to the classes. Classification has been applied on four tables of Supermarket dataset. Classification has been applied to tables of data set. Algorithm for classification of product table is as follows:

1. Retrieve all main categories from product_category table
2. Retrieve all product class id from main categories
3. Retrieve the sub category names from the class id from product category table
4. Match each class count for sub categories with the product class count from the products table

5. If the class id product_category table is matched with class id of products table, increment the count for each class id
6. Create array of all counts of class id.
7. Grouping of category, count of items in the category and number of sub categories for each product category
8. Average out for each sub category count by the total category count.

Similarly, classification can be applied to Customer, Sales and promotions.

Experimental Results of Classification Technique

Proposed work has focused on classification of supermarket dataset. Data of supermarket has been divided into suitable classes in order to enhance business decision making process. An algorithm for classification has been applied and tested for different tables of dataset. The classification results for Product table of dataset are given as follows:
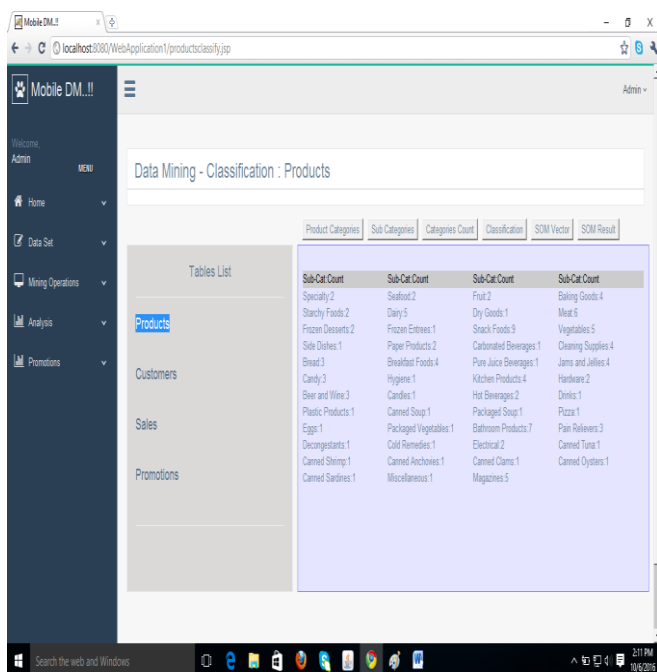

Fig 1. Classification of Products

Classification is used to find out in which group each data instance is related within a given dataset. It is used for classifying data into different classes according to some constrains. In proposed system classification has been performed in server and results are given to mobile devices with the help of web services. Results of classification on mobile device are as shown in fig 2

Fig 2. Classification of Products

Comparision of Classification technique in web and mobile applications

Sr. No	Operation type	Time required for web application in seconds	Time required for Mobile application in seconds
1	Classification of Product	2.873	0.236
2	Classification of Customer	0.266	0.071
3	Classification of Sales	66.332	0.344
4	Classification of Promotions	2.827	0.045

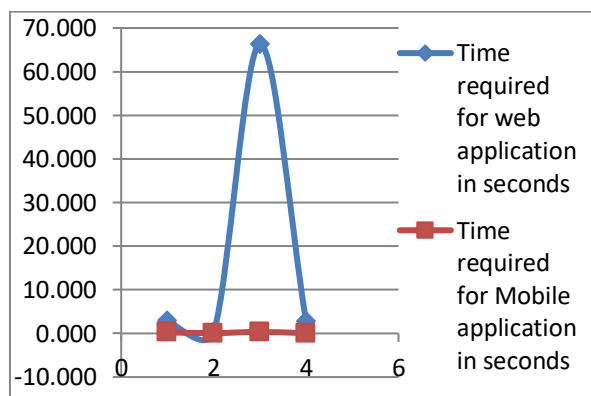**Table1. Time required for web and mobile application**

Fig 3 Comparision of time required for web and mobile application.

CONCLUSION

The internet and web services usage on mobile devices are continuously and rapidly growing. Therefore the demand is to have efficient mobile interface that can effectively display information and efficiently utilize the small size mobile screen, low bandwidth and unreliable connection etc. Also business organizations including supermarket sectors yields tones of records and sale every day, the most suitable mechanisms that can handle such vast growth of data set and information is data mining techniques. Data

41

mining is known as the process of monitoring new and innovative information from vast amount of data sets by viewing structured of dataset table and actual data of supermarket. If mobile devices are used for data mining then user can view data mining results anytime anywhere. But because of some limitations of mobile devices, it is better to perform mining on server and visualize results on mobile device. It can be called as mobile data mining. Mobile data mining is the revolution in the present era.

REFERENCES

1. Domenico Taliayand Paolo Trunfio*z*," Mobile Data Mining on Small Devices Through Web Services." http://grid.deis.unical.it/papers/pdf/MI2010.pdf Viewed on 25-03-2011

2. Dijiang Huang, Xinwen Zhang, Myong Kang, Jim Luo," MobiCloud: Building Secure Cloud Framework for Mobile Computing and Communication." www.dj.eas.edu/courses/CSE 548/2010Spring/papers/mobile % 20cloud.pdf, Viewed on 15-03-2011.

3. Frank Wang and Nu Helian, " A Distributed and Mobile Data Mining System", 2003 IEEE.

4. Sherif Rashad, Mehmed Kantardzic, and Anup Kumar, "Mobile Data Mining for Radio Resource Management in Wireless Mobile Networks",

5. Mohamed Medhat Gaber, "Foundations Of Adaptive Data Stream Mining For Mobile And Embedded Applications" ,Proceedings of the 2008 IEEE, CIBEC'08

6. DebajyotiMukhopadhyay, PradiptaBiswas, Young-Chon Kim.,"A Syntactic Classification based Web Page Ranking Algorithm",6th International Workshop on MSPT Proceedings 2006,pp 83-92.

7. Jeffrey Xu Yu, YumingOu, Chengqi Zhang and Shichao Zhang, "Identifying Interesting Visitors through Web Log Classification", 1541-1672/05/$20.00 © 2005 IEEE

8. Media Anugerah Ayua, Siti Aisyah Ismaila, Ahmad Faridi Abdul Matina, Teddy Mantorob," *A comparison study of classifier algorithms for mobile-phone's accelerometer based activity recognition*", International Symposium on Robotics and Intelligent Sensors 2012 (IRIS 2012)[ELSVIER]

9. Ben D. Fulcher and Nick S. Jones, "Highly Comparative Feature-Based Time-Series Classification", IEEE Transactions On Knowledge And Data Engineering, Vol. 26, No. 12, December 2014.

10. Megha Gupta, Naveen Aggarwal, " Classification Techniques Analysis", NCCI 2010 -National Conference on Computational Instrumentation CSIO Chandigarh, INDIA, 19-20 March 2010.

11. Aditi Goel, Saurabh Kr. Srivastava, "Study of Data Mining Algorithms in the Context of Performance Enhancement of Classification ", *International Journal of Computer Applications (0975 – 8887) Volume 134 – No.9, January 2016*

12. Sharon Carl, Glaston D'souza and Linet Varghese, "Implementation of Classification Algorithms and their Comparison for Educational Dataset", *International Journal of Innovative Science, Engineering & Technology(IJISET), Vol. 3 Issue 3, March 2016. www.ijiset.com ISSN 2348 – 7968*