

Pattern Discovery using Association Rule

VandanaMunde¹ PratikAhuja² ChinmayBirla³ SonuVarma⁴

^{1,2,3,4} Department of I. T. Engineering

^{1,2,3,4} Thakur College of Engineering and Technology

Abstract---Database mining is motivated by decision support problems faced by most business organizations and is described as an important area of research. One of the most challenges in database mining is developing fast and efficient algorithms that can handle large volume of data because most mining algorithms performs computation over the entire database and often the database are very large. The methodologies of data mining that are used by existing market basket analysis are apriori algorithm to find frequent item sets for a large database but it has many shortcoming which can be resolved by many other algorithms one of which is Partitioning the large database and using apriori algorithm . Existing Market basket analysis display the frequent item set when searched for a particular item but the proposed system will show the results of prediction when searched for a particular item so as to improve the business throughput.

Keywords: Retail, Market, Data, Association Rule, Account, Database, Algorithm, Admin KDD, Improved System, SDLC, Apriori, Mining, warehousing.SCM

I. INTRODUCTION

Data Mining, The Extraction of Hidden Predictive information from large databases, is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses. Data mining tools predict future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions. The automated, prospective analyses offered by data mining move beyond the analyses of past events provided by retrospective tools typical of decision support systems.

Data mining (DM), also called Knowledge-Discovery in Databases (KDD) or Knowledge-Discovery and Data Mining, is the process of automatically searching large volumes of data for patterns using tools such as classification, association rule mining, clustering, etc.. Data mining is a complex topic and has links with multiple core fields such as computer science and adds value to rich seminal computational techniques from statistics, information retrieval, machine learning and pattern recognition.

The knowledge Management strategies which we are using in this project Market Basket Analysis (MBA). Market-Basket Analysis is a process to analyze the habits of buyers to find the relationship between different items in their market basket. Market basket analysis helps to provide scientific decision support for retail market by mining association rules among items people purchased together. The algorithm which we are using is an Apriori algorithm.

Administrator can generate the Market Basket Analysis result in the form of sales report, from these decision makers can determine which products are currently

lacking or leading based in their sales. Using this data organization can improve a SCM system with better accuracy and efficiency. This project can be found useful in analyzing trends and formulating solutions to problems before hand.

II. PROBLEM DEFINITION

The basic apriori algorithm requires multiple passes over the database. For disk resident database, this requires reading the database completely for each pass resulting in a large number of disk I/Os. In these algorithms, the effort spent in performing just the I/O may be considerable for large databases. Apart from poor response times, this approach also places a huge burden on the I/O subsystem adversely affecting other users of the system. The problem can even be worse in a client-server environment.

To short come the above problems the proposed system focuses on the portioning method to improve the performance of finding the association rules in the transaction databases. The proposed system uses the Partitioning and Apriori algorithm (Improved Apriori Algorithm) to build a system, which helps directors of shops/stores, is to have detailed view about his business. The soft gains an acceptable result when runs over a quite large databases. The proposed system considers supermarket database and applies improved apriori algorithm based on partitioning.

The proposed system consists of the following steps:

1. To evaluate the importance of finding associations rules and specify the main cost of the process finding them.
2. To present, illustrate and analyze the strength and weakness of some algorithms using partitioning approach.
3. To build up a system to manage a small soft, find interesting rules related to customer routines. This system uses improved apriori algorithm that provides good efficiency.

III. DEVELOPMENT IDEA

- Integration of Knowledge Management System with Supply Chain Management System is nowadays requirement of many organizations.
- The Knowledge Management System strategies and practices will help organization to manage their supply chain in much efficient way.
- In this application we are using data mining technologies which will help organizations to predict future demand of product, inventory management and allows businesses to make Knowledge Driven Decisions.
- Knowledge management integration with SCM helps Organization to manage their business using different data mining technologies like Association Rules like Market Basket Analysis which will check customers buying product transaction details using Apriori (data mining algorithm).

– As Knowledge Management System integration will provide the organization a better way to manage their overall business and help organization to earn more revenue.

A. Functional Requirements

1) Administrator:

The administrator would log in into the system with his credentials. Then the form would be directed to his own account. The MDI form will be displayed. It manages the whole system.

2) Database:

The database with stored procedure save modification whenever the admin stores or update the customer information.

3) Development Team:

B. Non Functional Requirements:

1. 24 X 7 availability.
2. Better component design to get better performance at peak time.
3. Secure access of confidential data (Admin’s details).
4. Flexible service based architecture will be highly desirable for future extension.

IV. RESULTS

The project will help many organizations to learn customer’s behavior and manage their Supply Chains (SC) with the application. In this project we are integrating Knowledge Management System with Supply Chain. This combination of two systems will provide much efficient way to manage Overall business.

This project presents a very simple but efficient method for retail management for different items and associating them. In today’s life it very much important to have simple procedures and mining approaches for complex database.

In the proposed approach, the transaction occurs as soon as the Admin registers on the database for this account and gets his/her user id and password. Once the Admin is registered on the Database he/she can update accordingly wherever desired, search for items and their different relations, entering minimum and maximum support of each and thus finding an association between different items...

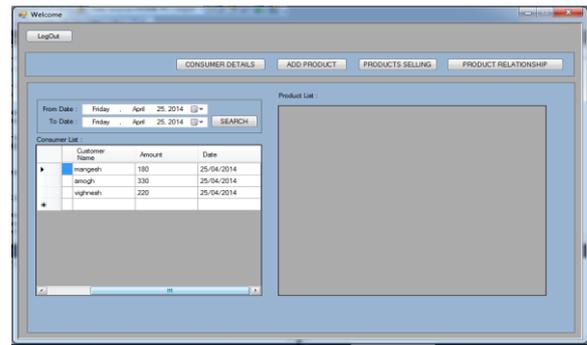


Fig. 2: Consumer Detail

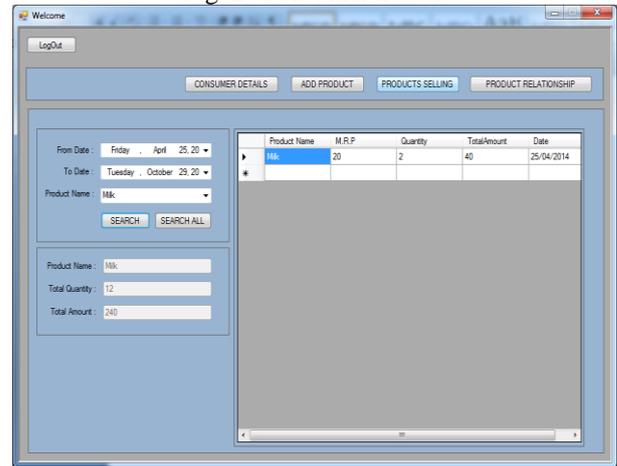


Fig. 3: ProductSelling

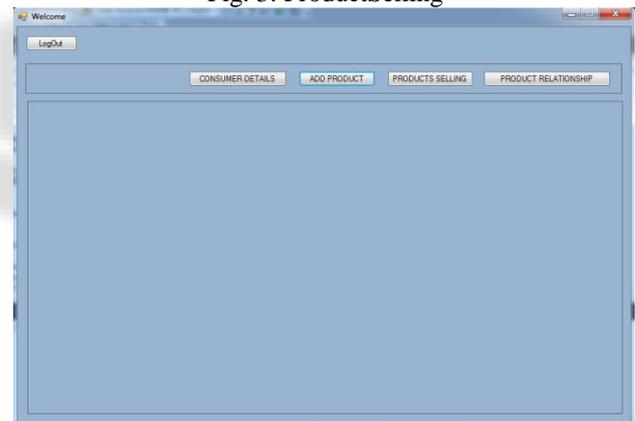


Fig. 3: Menu Form

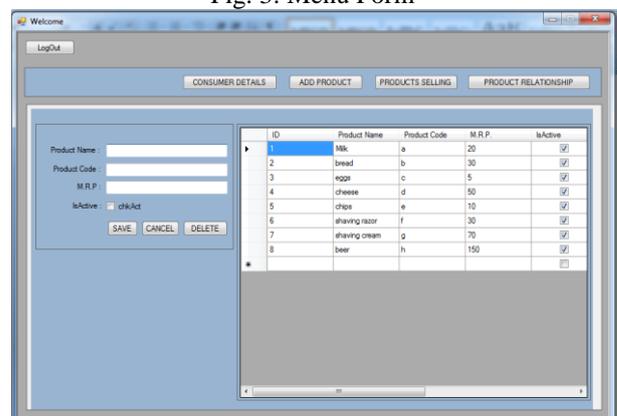


Fig. 4: Add Product



Fig. 1: Login Screen

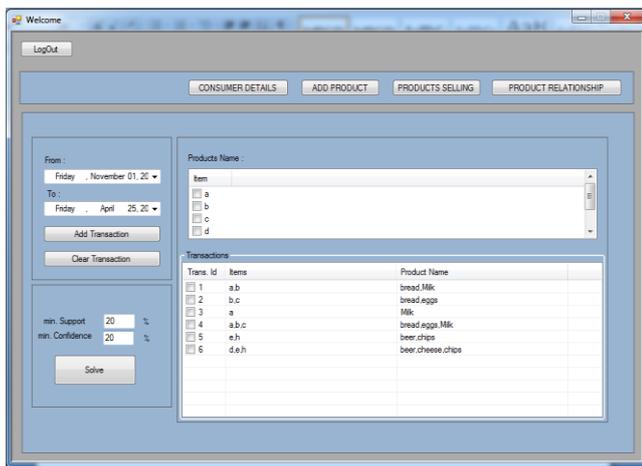


Fig. 5: Product Relationship

V. DISCUSSION

Finding large item sets – find out sets of items, which have frequency appeared together higher than give number – is a very important part in the process of finding association rule. It works with large amount of data so the problem of optimizing the process and reducing data scanning will influence the effect of this step in particular and influences all the process in general a lot. The more data that could be ignored, more the running time we saved. While database expands day after day, and become colossal we try to make the size of transaction which is needed to scan in iteration smaller and fit our work in limited resources. To evaluate the importance of finding association rules and specifies the main cost of the process finding them, also present illustrate and analyze the strength and weakness of some algorithms. The major advantage of the project is that it provides large database and mining approaches where the items are associated and market analyses in retail management where database consist of large number of customers. Thus this project give association rule of different itemset and thus gives the whole scenario of retail management and thus managers of supermarket and retail stores get ideas about the products and their purchase by customers. In the proposed approach, the transaction occurs as soon as the Admin registers on the database for this account and gets his/her user id and password. Once the Admin is registered on the Database he/she can update accordingly wherever desired, search for items and their different relations, entering minimum and maximum support of each and thus finding an association between different items...

Benefits:

- Accuracy
- User Friendly
- Promptness
- Flexible
- Secured
- Instant response
- Detailed description
- Attractive GUI

VI. FUTURE SCOPE

The proposed system can be implemented and launched at the broader scale for commercial purposes.

This software can be linked with different retail stores. This will provide attractive application in stores. Larger number of database can be stored. Also the feedback of the Customers can be added and graphically generated.

ACKNOWLEDGMENTS

We would like to express my deep gratitude to DR. B.K Mishra, Principal, Thakur College of Engineering and Technology, Mumbai for extending the opportunity for major project and providing all the necessary resources for this purpose. We express heartfelt thanks to Mr. Vandana Munde for her wonderful support for preparing the project and for giving us an opportunity to do our project on “Pattern discovery using Association Rule”. We are grateful to Mr. Vinayak Bharadi, Head of Department (Information Technology), Thakur College of Engineering and Technology, Mumbai and all the faculty members for conducting the project and their encouragement and cooperation has been a source of great inspiration. Also we would like to thank Thakur College of Engineering and Technology for providing me all the facilities for timely completion of the project.

REFERENCES

- [1] Evfimievski A (2002) “Randomization in privacy preserving data mining”. SIGKDD Explorations 4(2), Issue 2:43-48. .
- [2] Dasseni E, Verykios V, Elmagarmid A, Bertino E (2001) “Hiding association rules by using confidence and support”. In: Proceedings of 4th Information Hiding Workshop, Pittsburgh, PA, pp 369-383.