

Data Mining For Supermarket Sale Analysis Using Association Rule

Mrs. R. R. Shelke

H.V.P.M. COET, Amravati

Dr. R. V. Dharaskar

Former Director
DES (Disha – DIMAT) Group of
Institutes, Raipur

Dr. V. M. Thakare

Prof. and Head,
Computer Science Dept., SGB
Amravati University, Amravati

ABSTRACT

Data mining is the novel technology of discovering the important information from the data repository which is widely used in almost all fields. Recently, mining of databases is very essential because of growing amount of data due to its wide applicability in retail industries in improving marketing strategies. Analysis of past transaction data can provide very valuable information on customer behavior and business decisions. The amount of data stored grows twice as fast as the speed of the fastest processor available to analyze it. Its main purpose is to find the association relationship among the large number of database items. It is used to describe the patterns of customers' purchase in the supermarket. This is presented in this paper.

KEYWORDS:

Data mining, Associations, supermarket

I. INTRODUCTION

Data mining tasks can be classified into two categories: Descriptive mining and Predictive mining. Descriptive mining refers to the method in which the essential characteristics of the data in the database are described. Clustering, Association and Sequential mining are the main tasks involved in the descriptive mining techniques tasks. Predictive mining deduces patterns from the data in a similar manner as predictions. Predictive mining techniques include tasks like Classification, Regression and Deviation

detection. Mining Frequent Itemsets from transaction databases is a fundamental task for several forms of knowledge discovery such as association rules, sequential patterns, and classification. An itemset is frequent if the subsets in a collection of sets of items occur frequently. Frequent itemsets is generally adopted to generate association rules. The objective of Frequent Item set Mining is the identification of items that co-occur above a user given value of frequency, in the transaction database. Association rule mining is one of the principal problems treated in KDD and can be defined as extracting the interesting correlation and relation among huge amount of transactions.

II. LITERATURE REVIEW

Association Rule Discovery has become a core topic in Data Mining. It attracts more attention because of its wide applicability. Association rule mining is normally performed in generation of frequent itemsets and rule generation in which many researchers presented several efficient algorithms [1-5]. T. Karthikeyan and N. Ravikumar, aim at giving a theoretical survey on some of the existing algorithms [3]. The concepts behind association rules are provided at the beginning followed by an overview to some of the previous research works done on this area. The advantages and limitations are discussed and concluded with an inference. Association rule

mining discovers the frequent patterns among the itemsets. It aims to extract interesting associations, frequent patterns, and correlations among sets of items in the data repositories. For Example, In a Laptop store in India, 80% of the customers who are buying Laptop computers also buy Data card for internet and pen drive for data portability.

Kinds of Association Rule Mining :

According to *Udaiveer Singh Parmar*, Based on the number of data dimensions involved, we can distinguish association rules on basis of dimensions [6] :

A. Single-dimensional association rule: An association rule is a single-dimensional, if the items or attributes in an association rule reference only one dimension.

B. Multidimensional association rule:

If a rule references more than one dimension, such as the dimensions like study-level, income, and buys, then it is a multidimensional association rule. Let X an item set, the following rule is an example of a multidimensional rule: Study-Level(X, "20...25") \rightarrow income(X, "30K.... 40K") \rightarrow buys(X, "performant computer"):

Based on the types of values handled by the rule, we can distinguish two types of association rules:

Boolean association rule: A rule is a Boolean association rule, if it involves associations between the presence or the absence of items. For example, the following rule is a Boolean association rules obtained from market basket.

analysis: buys(X, "computer") \rightarrow buys(X, "scanner").

Quantitative association rule: a rule is called quantitative association rule, if it describes associations between quantitative items or attributes. In these rules, quantitative values for items or attributes are partitioned into intervals.

For the past decades, there are several efforts has been made to discover the scalable and efficient methods for mining frequent ARs. However, mining least ARs is still left behind. As a result, ARs that are rarely found in the database are pruned out by the minimum support-confidence threshold. As a matter of fact, the rarely ARs can also reveal the useful information for detecting the highly critical and exceptional situations. One suggested a method to mine the ARs by considering only infrequent itemset. The drawback is, Matrix-based Scheme (MBS) and Hash-based scheme (HBS) algorithms are very expensive in term of hash collision. Ding proposed Transactional Co-occurrence Matrix (TCOM) for mining association rule among rare items. However, the implementation wise is quite complex and costly. Yun, et al., [7] introduced the Relative Support Apriori Algorithm (RSAA) to generate rare itemsets. The challenge is, it takes similar time taken as performed by Apriori if the allowable minimum support is set to very low. Koh, et al., [8] suggested Apriori-Inverse algorithm to mine infrequent itemsets without generating any frequent rules. However, it suffers from candidate itemset generations and costly in generating the rare ARs. Multiple Support Apriori (MSApriori) algorithm has been used to extract the rare ARs. In actual implementation, this algorithm is facing the "rare item problem". Many of them are using the percentage-based approach to improve the performance as faces

by the single minimum support based approaches. An objective measure called lift and chi-square as correlation measure for ARs has been introduced. Lift compares the frequency of pattern against a baseline frequency computed under statistical independence assumption. Omiecinski proposed two interesting measures based on downward closure property called all confidence and bond [9]. There are two algorithms for mining all confidence and bond correlation patterns by extending the pattern-growth methodology. In term of mining algorithms, Agrawal, et al., [10] proposed the first ARs mining algorithm called Apriori. The main bottleneck of Apriori is, it requires multiple scanning of transaction database and also generates huge number of candidate itemsets. AlhasanBala et al.[11] suggested FP-Growth algorithm which amazingly can break the two limitations as faced by Apriori series algorithms. Currently, FP-Growth is one of the fastest approach and most benchmarked algorithms for frequent itemsets mining.

III. PROPOSED SYSTEM FOR ASSOCIATION RULE MINING

One of the most popular data mining methods is to discover frequent itemsets from a transaction dataset and develop association rules. Finding frequent itemsets (itemsets with frequency larger than or equal to a user defined minimum support) is significant because of its combinatorial explosion. As soon as frequent itemsets are found, it is straightforward to produce association rules with confidence greater than or equal to a user stated minimum confidence. Apriori is an important algorithm for finding frequent itemsets

using candidate generation. It is considered as a level-wise whole search algorithm using anti-monotonicity of itemsets i.e. “If an itemset is not frequent, any of its superset is never frequent”.

Algorithm used for association for proposed system is as follows:

1. Get unique product ID and product name from product and sales tables where the product ID of product and sales table are matched
2. Convert all product names and product ID into two arrays
3. Retrieve product class ID for each product from the product table
4. Retrieve unique product category for each product class ID from product_category table.
5. Create array of product Categories
6. Provide the product category array to apriori algorithm
7. Store results of Apriori in database table finaloutput

IV. RESULTS OF ASSOCIATIONS

Market basket analysis is one of the most common and beneficial technique of data analysis for marketing and retailing. The main purpose of Market basket Analysis is to decide what products are usually bought together by the customer. Market basket analysis identifies purchasing habits of customers. It offers awareness into the combination of products within a customer's 'basket'. The term 'basket' normally applies to a single order. However, the analysis can be useful to other variations. In Market Basket Analysis one can analyze combination of products to be sold together and this will be helpful

for both retailer as well as manufacturing company. A store could use this analyzed information to place products frequently sold together into the same area, so that store product selling gets increased. To find association rules Apriori algorithm has been used. Association rule from the frequent itemset has been generated. A Sales table of supermarket dataset has been used. A set of association rules has been obtained by applying Apriori algorithm.

Result generated after applying association mining is as shown in fig. 1

| ID | Name | Class ID | Category |
|-----|------------------------|----------|-------------|
| E36 | Party Nuts | 1 | Specialty |
| E37 | Lox | 24 | Seafood |
| E38 | Scallops | 2 | Seafood |
| E39 | Low Fat Cookies | 45 | Snack Foods |
| E40 | BBQ Potato Chips | 12 | Snack Foods |
| E41 | Corn Chips | 12 | Snack Foods |
| E42 | Low Fat BBQ Chips | 12 | Snack Foods |
| E43 | Sugar Cookies | 45 | Snack Foods |
| E44 | Chocolate Chip Cookies | 45 | Snack Foods |
| E45 | Fudge Cookies | 45 | Snack Foods |

Category : Fruit

Items : 1-Ebony Cantalope 2-Ebony Honey Dew 3-Ebony Fuji Apples 4-Ebony Oranges 5-Ebony Tangerines

Item Ebony Cantalope & Ebony Honey Dew are selected
 Item Ebony Cantalope & Ebony Fuji Apples are selected
 Item Ebony Cantalope & Ebony Tangerines are selected
 Item Ebony Honey Dew & Ebony Fuji Apples are selected
 Item Ebony Honey Dew & Ebony Oranges are selected
 Item Ebony Fuji Apples & Ebony Tangerines are selected
 Item Ebony Fuji Apples & Ebony Oranges are selected

Fig 1. Result of Associations

The rapid growth and advances of information technology enable data to be accumulated faster and in much larger quantities. Faced with vast new information resources, scientists, engineers, and business people need efficient analytical techniques to extract useful information and effectively uncover new, valuable knowledge patterns. In this paper, association rule mining for supermarket dataset has been presented. Mining has been applied to sales data of dataset. In proposed system, the apriori algorithm has been used on super market dataset which gives associations of two products which has maximum support.

REFERENCES:

1. Wiwik Novitasari, Arief Hermawan, Zailani Abdullah, Rahmat Widia Sembiring and Tutut Herawan, "A Method of Discovering Interesting Association Rules from Student Admission Dataset", International Journal of Software Engineering and Its Applications Vol. 9, No. 8 (2015), pp. 51-66.
2. Thabet Slimani, Amor Lazzez, "Efficient Analysis of Pattern and Association Rule Mining Approaches".
3. T. Karthikeyan and N. Ravikumar, "A Survey on Association Rule Mining", International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 1, January 2014.
4. Miss. Pooja Rajendra Harne, "Mining of Association Rules: A Review Paper", International Journal of Environmental

V. CONCLUSION

Engineering Science and Technology Research
Vol. 3, No. 1, June 2015.

5. Rajdeep Kaur Aulakh, “ Association Rules Mining Using Effective Algorithm: A Review”, International Journal of Advanced Research in Computer Science and Software Engineering Research Paper, Volume 5, Issue 3, March 2015.
6. Udaiveer Singh Parmar, Prof. Anand Motwani and Prof. Anurag Shrivastava, “Association Rule Mining Various Ways: A Comprehensive Study”, International Journal of Electrical, Electronics and Computer Engineering 4(2): 134-138(2015).
7. Yun Sing Koh, Russel Pears, “Rare Association Rule Mining via Transaction Clustering”, Seventh Australasian Data Mining Conference (AusDM 2008), Glenelg, Australia. Conferences in Research and Practice in Information Technology (CRPIT), Vol.87.
8. Yun Sing Koh and Nathan Rountree, “Finding Sporadic Rules Using Apriori-Inverse”, T.B. Ho, D. Cheung, and H. Liu. (Eds.): PAKDD 2005, LNAI 3518, pp. 97–106, 2005. Springer-Verlag Berlin Heidelberg 2005.
9. Ashok Savasere, Edward Omiecinski, Shamkant Navathe, “An Efficient Algorithm for Mining Association Rules in Large Databases”, Proceedings of the 21st VLDB Conference Zurich, Switzerland, 1995.
10. Rakesh Agrawal, Ramakrishnan Srikant, “Fast Algorithms for Mining Association Rules”, Proceedings of the 20th VLDB Conference Santiago, Chile, 1994
11. Alhassan Bala, Mansur Zakariyya Shuaibu, Zaharaddeen Karami Lawal and Rufa'i Yusuf

Zakari, “Performance Analysis of Apriori and FP-Growth Algorithms (Association Rule Mining)”, Int. J. Computer Technology & Applications, Vol 7(2), 279-293 IJCTA | March-April 2016 .