Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 2, Issue. 11, November 2013, pg.161 – 165

RESEARCH ARTICLE

PREDICTING MISSING ITEMS IN SHOPPING CART USING ASSOCIATIVE CLASSIFICATION MINING

K. Jagadish Kumar[1]
Sahini Sairam[2]
Associate Professor[2]
M.Tech[1]
Assistant Professor[1]

Abstract: ***The primary task of association rule mining is to detect frequently co-occurring groups of items in transactional databases. The intention is to use this knowledge for prediction purposes. So many researches has focused mainly on how to expedite the search for frequently co-occurring groups of items in "shopping cart" and less attention has been paid to the methods that exploit these "frequent itemsets" for prediction purposes. This paper contributes to the latter task by proposing a technique that uses the partial information about the contents of a shopping cart for the prediction of what else the customer is likely to buy, for example, If bread, butter, and milk often appear in the same item, then the presence of butter and milk in a shopping cart suggests that the customer may also buy bread. More generally knowing which items a shopping cart contains, we want to predict often items that the customer is likely to add before proceeding to the checkouts. So this paper presents a technique called the "Combo Matrix" whose principal diagonal elements represent the association among items and looking to the principal diagonal elements, the customer can select what else the other items can be purchased with the currently contents of the shopping cart and also reduces the rule mining cost. The association among items is shown through Graph. The frequent itemsets are generated from the Combo Matrix. Then association rules are to be generated from the already generated frequent itemsets. The association rules generated form the basis for prediction. The incoming itemsets i. e. the contents of the shopping cart will be represented by set of unique indexed numbers and the association among items is generated through the Combo Matrix. Finally the predicted items are suggested to the Customer.***

I INTRODUCTION

In recent years, the Web has become the primary means or information dissemination. In past, the site developed is based on the concept of Online shopping with shopping carts that it can fulfill some user's requirements that, they can sit at home and purchase product with reduced time and cost and the user can purchase their products as quickly as possible. The user can purchase the product from a specified branch and the site is user friendly as the user found that the online shopping is easier to work on it. Apart from user purchasing the product, administrator has to add product details according to branch requirements. Actually, at the same time user's are in need of more fast progress in shopping cart area. If the user is purchasing some products like bread, they definitely will go for one more needed

product jam. Here comes the problem, the user have to type the related product or have to search for the required product to add that product to the shopping cart. It will consume some amount of user's time. To avoid those things, we are in need of predicting the products related to the user's products. In this System we propose a predicting missing items in online shopping using association rule mining. The primary task of association mining is to detect related items according to user's purchasing products. A large number of user's are interested in purchasing products and the related products. Here the association rule associate (union) the product with predicted and related products that the user wants to buy. More generally, knowing which items a shopping cart contains, we want to predict the other items that the customer is likely to add before proceeding to check out counter .Suppose the shopping cart of customer at the check out counter bread ,butter milk, and after that a predicted items should be added to the shopping cart and we used Classification rule mining is applied to database with tables that has related and predicted products. Apart from the predicting items, we go for more advanced data security checks are made to prevent unauthorized access by other user.

1.1 DATA MINING

Data Mining refers to extracting or mining information from large amounts of data. Data mining has attracted a great deal of attention in the information industry and in society as a whole in recent years, due to the wide availability of huge amounts of data and the imminent need for turning such data into useful information and knowledge. 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 analysis offered by data mining move beyond the analysis of past events provided by retrospective tools typical of decision support systems. Data mining tools can answer business questions that traditionally were too time consuming to resolve. They scour databases for hidden patterns, finding predictive information that experts may miss because it lies outside their expectations. Most companies collect and refine massive quantities of data. Data mining techniques can be implemented rapidly on existing software and hardware platforms to enhance the value of existing information resources and can be integrated with new products and systems as they are brought on-line. When implemented on high performance client/server or parallel processing computers, data mining tools can analyze massive databases to deliver answers to many questions. The information and knowledge gained can be used for application ranging from market analysis, fraud detection, and customer retention, to production control and science exploration. Data Mining plays an important role in online shopping for analyzing the subscribers" data and understanding their behaviors and making good decisions such that customer acquisition and customer retention are increased which gives high revenue.

1.2 ASSOCIATION RULE MINING

Association Rule Mining is a popular and well researched method for discovering interesting relations between variables in large databases. Association rules are statements of the form {X1, X2, …, Xn} => Y meaning that if all of X1, X2,… Xn is found in the market basket, and then we have good chance of finding Y. the probability of finding Y for us to accept this rule is called the confidence of the rule. Normally rules that have a confidence above a certain threshold only will be searched. In many situations, association rules involves sets of items that appear frequently. For example, a good marketing strategy cannot be run involving items that no one buys. Thus, much data mining starts with the assumption that sets of items with support are only considered. The discovery of such associations can help retailers develop marketing strategies by gaining insight into which items are frequently purchased together by customer and which items bring them better profits when placed with in close proximity. The two types of finding association between products existing in a large database are Boolean [7] and Quantitative. Boolean association rule mining finds association for the entire dataset. Quantitative association rule mining finds association for the clusters formed from the dataset.

1.3 PREDICTION

Data mining automates the process of finding predictive information in large databases. Questions that traditionally required extensive hands-on analysis can now be answered directly from the data quickly. The primary task of association mining is to detect frequently co-occurring groups of items in transactional databases. The intention is to use this knowledge for prediction purposes. Early attempts for prediction used classification and performance was favorable. In this project, any item is allowed to be treated as a class label its value is to be predicted based on the presence of other items. Put another way, knowing a subset of the shopping cart‟s contents, we want to "guess" (predict) the rest. Suppose the shopping cart of a customer at the checkout counter contains bread, butter, milk, cheese, and pudding. Could someone who met the same customer when the cart contained only bread, butter, and milk, have predicted that the person would add cheese and pudding? It is important to understand that allowing any item to be treated as a class label presents serious challenges as compared with the case of just a single class label. The number of different items can be very high, perhaps hundreds, or thousand, or even more. To generate association rules for each of them separately would give rise to great many rules with two obvious consequences: first, the memory space occupied by these rules can be many times larger than the original database (because of the task‟s combinatorial nature); second, identifying the most relevant rules and combining their sometimes conflicting predictions may easily incur prohibitive computational costs. In this work, both of these problems are solved by developing a technique that answers user‟s queries (for shopping cart completion) in a way that is acceptable not only in terms of accuracy, but also in terms of time and space complexity. This paradigm can be exploited in diverse applications. For example, in the each "shopping cart" contained a set of hyperlinks pointing to a Web page; in medical applications, the shopping cart may contain a patient‟s symptoms, results of lab tests, and diagnoses; in a financial domain, the cart may contain companies held in the same portfolio. In all these databases, prediction of unknown items can play a very important role. For instance, a patient‟s symptoms are rarely due to a single cause; two or more diseases usually conspire to make the person sick. Having identified one, the physician tends to focus on how to treat this single disorder, ignoring others that can meanwhile deteriorate the patient‟s condition. Such unintentional neglect can be prevented by subjecting the patient to all possible lab tests. However, the number of tests one can undergo is limited by such practical factors as time, costs, and the patient‟s discomfort. A decision support system advising a medical doctor about which other diseases may accompany the ones already diagnosed can help in the selection of the most relevant additional tests.

II SYSTEM DESIGN

Any project developed today is said to be good only if it has some basic characteristics such as modularity, loose coupling and high cohesion. A component is classified as good quality only if it is modular, loosely coupled and has high cohesion i.e., each component should be independent of the other and each component must be focused only on its particular purpose. Finally the component should be modular so that the development of the components is understandable, can easily be enhanced in the future and also easy to locate and correct errors without affecting the other components involved in the project. The following sections deals with how this system is designed, the modules involved and overall architecture diagram of the system which shows the modules present in the system.

2.1 SHOPPING CART PREDICTION ARCHITECTURE

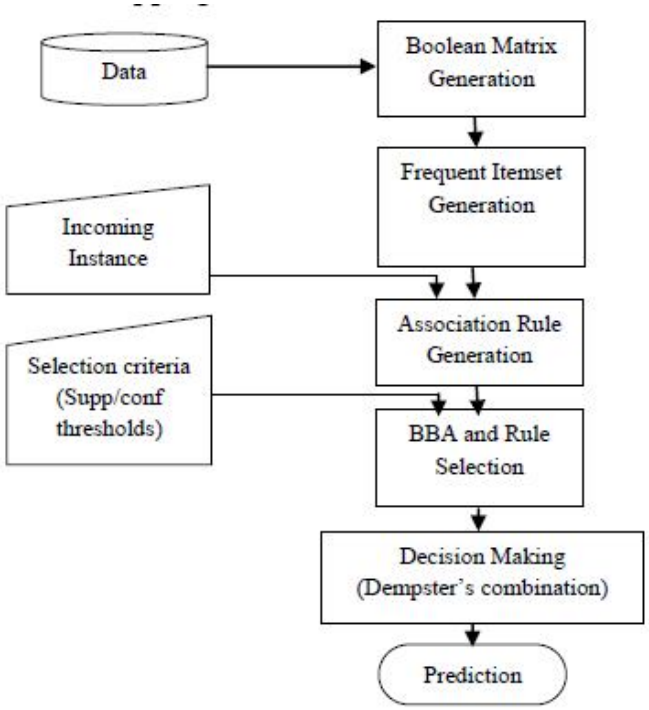

Fig 2.1 Architecture

Fig.2.1 shows the shopping cart prediction architecture in which the Boolean Matrix is generated by transforming the database into Boolean values. The frequent item sets are generated from the Boolean matrix. At this stage we need the Support value. Then association rules are to generated from the already generated frequent item ets. It takes minimum confidence from the user and discovers all rules with a fixed antecedent and with different consequent. The association rules generated form the basis for prediction. We assign BBA [2] value to each association rule generated. This gives more weight to rules with higher support masses are assigned based on both their confidence and support values. The incoming item set i.e the content of incoming shopping cart will also be represented by a Boolean vector and AND operation is performed with each transaction vector to generate the association rules. Finally the rules are combined to get the predictions. Dempster‟s rule of combination (DRC) [6] is used to combine the evidences. When searching for a way to predict the presence or absence of an item in a partially observed shopping cart s, we wanted to use association rules. However, many rules with equal antecedents differ in their consequents—some of these consequents contain the desired item to be predicted, others do not. The question is how to combine (and how to quantify) the potentially conflicting evidences. DRC [6] is used for this purpose. Finally the predicted items are suggested to the user.

III IMPLEMENTATION

This topic consists of detailed description of each and every module with its advantages and data and execution flow of each module with algorithm. It helps to understand each and every module of the project more deeply and clearly. Each description consists of the basic concept of the module, input and also the excepted output.

3.1 MODULES

The project has been divided into various modules and each module has been completed within a scheduled time line. The following are the modules of the project are boolean matrix generation,Frequent itemset generation, Association rule generation, BBA and decision making.

IV CONCLUSION

The fast algorithm generating frequent item sets without generating candidate item sets proposed performs well compared to existing approaches. The execution time is much improved as shown in performance testing. The algorithm used Boolean vector with relational AND operation to discover frequen item sets and generate the association rule. Association rules formed the basis of prediction. The algorithm is applied in a demo shopping cart application. When user adds each item to the cart the algorithm is executed and the prediction is displayed.

REFERENCES

- Agrawal. R, Imielinski. T and Swami. A, "Mining Association Rules between Sets of Items in Large Databases," Proc. ACM Special Interest Group on Management of Data (ACM SIGMOD), pp. 207-216, 1993.
- M. Kubat, Hafez. A, Raghavan V. V, J. R. Lekkala and hen W. K. : "Itemset Trees for Targeted Association Querying," IEEE Trans . Knowledge and Data Eng. , vol. 15, no. 6, pp. 1522-1534, Nov. /Dec. 2003.
- Kasun Wickramaratna ," Predicting missing Items in Shopping Carts", IEEE Transaction on Knowledge and data engineering, Vol. 21, No. 7, July 2009
- C. C Aggarwal, C. Procopius, and P. S. Yu, "Finding Localized Associations in Market Basket Data," IEEE Transaction on Knowledge and Data": Eng. , vol. 14, no. 1, pp. 51-62, Jan. /Feb. 2002.
- R. Bayardo and R. Agrawal, "Mining the Most Interesting Rules," Proc. ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining, pp. 145-154, 1999.
- Liu. B, Hsu, Y. M. WAND Ma,: "Integrating Classification and Association Rule Mining," Proc. ACM SIGKDD Int'l Conf. Know. Disc. Data. Mining (KDD '98), pp. 80-86, Aug. 1998.
- Li . W, Han. J, and Pei. J, "CMAR: Accurate and Efficient Classification Based on Multiple Class-Association Rules," Proc. IEEE Int'l Conf. Data Mining (ICDM '01), pp. 369-376, Nov. /Dec. 2001
- Yen S. J. and Chen A " An efficient approach to discovering knowledge from large database". In proc. Of the IEEE/ACM International Conference on parallel distributed Information system, Pages 8-18, 1996.
- Lee K. L, G Lee and Chen A. L. P. Efficient Graph based Algorithm for discovering and maintaining association rules in large database.

