International Journal of Information and Computation Technology.
ISSN 0974-2239 Volume 4, Number 4 (2014), pp. 409-414
© International Research Publications House
http://www. irphouse.com /ijict.htm

A Survey on Association Rule Mining in Market Basket Analysis

Savi Gupta and Roopal Mamtora

Department of CSE/IT, ITM University, Gurgaon, INDIA.

Abstract

Data mining refers to extracting knowledge from large amount of data. Market basket analysis is a data mining technique to discover associations between datasets. Association rule mining identifies relationship between a large set of data items. When large quantity of data is constantly obtained and stored in databases, several industries are becoming concerned in mining association rules from their databases. For example, the detection of interesting association relationships between large quantities of business transaction data can help in catalog design, cross-marketing and various business decision making processes. A typical example of association rule mining is market basket analysis. This method examines customer buying patterns by identifying associations among various items that customers place in their shopping baskets. The identification of such associations can assist retailers expand marketing strategies by gaining insight into which items are frequently purchased by customers. It is helpful to examine the customer purchasing behavior and assists in increasing the sales This work acts as a broad area for the researchers to develop a better data mining algorithm. This paper presents a survey about the existing data mining algorithm for market basket analysis.

Keywords: Association Rule Mining, Apriori Algorithm, Market Basket Analysis.

1. Introduction

Association rule mining(ARM) is used for identification of association between a large set of data items. Due to large quantity of data stored in databases, several industries are becoming concerned in mining association rules from their databases. For example, the detection of interesting association relationships between large quantities of business transaction data can assist in catalog design, cross-marketing, and various business decision making processes. A typical example of association rule mining is market basket analysis. This method examines customer buying patterns by identifying associations among various items that customers place in their shopping baskets. The identification of such associations can help retailers to expand marketing strategies by gaining insight into which items are frequently purchased jointly by customers. This work acts as a broad area for the researchers to develop a better data mining algorithm. This paper presents a survey about the existing data mining algorithm for market basket analysis.

This review paper is organized as follows: Section I contains brief introduction of ARM, Section II depicts market basket analysis which is an application of ARM, Section III discusses the literature survey in which various data mining algorithms are discussed, section IV discusses apriori algorithm, problems and directions of data mining algorithms are depicted in section V. Then the complete paper is summarized in the section VI, which includes conclusion and future scope.

2. Market Basket Analysis: An Overview

Market basket analysis(MBA) is a data mining technique to discover associations between datasets. These associations can be represented in form of association rules. The formal statement of problem[7] can be stated as : Let I is a set of items $\{i_1,i_2,\ldots,i_m\}$.Let D is a set of transactions such that T I. Each transaction is uniquely identified with with an identifier called TID. The method can be stated as if there are two subsets of product items X and Y then an association rule is in the form of $X \rightarrow Y$ where X I and Y I. It implies that if a customer purchases X, then he or she also purchases Y. Two measures which reflect certainity of discovered association rules are support and confidence. Support measures how many times the transactional record in database contain both X and Y. Confidence measures the accuracy of rule. As an example, the information that customers who purchase computers also tend to buy printer at the same time is represented in Association Rule below.

Computer = Printer

Support = 20%, Confidence = 80%

Association rules are considered useful if they satisfy both a type equation here minimum support threshold and a minimum confidence threshold that can be set by users or domain consultants. Figure1 shows a typical Market basket analysis. This is a perfect example for illustrating association rule mining. This market basket analysis system will help the managers to understand about the sets of items are customers likely to purchase. This analysis may be carried out on all the retail stores data of customer transactions. These results will guide them to plan marketing or advertising approach. For example, market basket analysis will also help managers to propose new way of arrangement in store layouts. Based on this analysis, items that are regularly purchased together can be placed in close proximity with the purpose of further promote the sale of such items together. If consumers who purchase computers also likely to purchase anti-virus software at the same time, then placing the hardware display close to the software display will help to enhance the sales of both of these items.

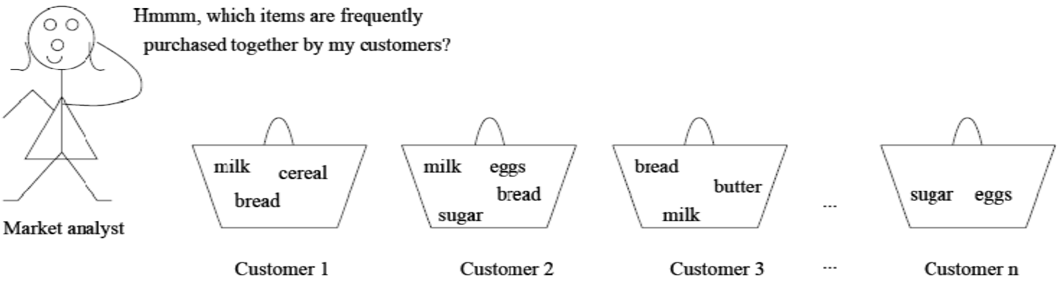


Fig. 1: Market basket analysis.

3. Literature Survey

In this section we have concentrated on presenting different areas where data mining algorithms are used. This section outlines the existing algorithms that were designed by the researchers in context of association rule mining in MBA.

3.1 Market Basket Analysis In Large Database Network

Market basket analysis is performed to take business decisions like what to place on sale, the way to place things on shelves to maximise profit etc. An analysis of past transaction data is done for this purpose. Upto now only global data about the transactions during some time period like a day or a week etc was available on computer. However the progress in bar code technology makes it possible to store data on transaction basis and as a result of this large amount of data is collected. These data sets are usually stored on tertiary storage because of limited functionality of database. So, to enhance the functionality of database and to process queries[2] such as:

1. Find all the rules that have “butter” as consequent. These rules may help to plan that what should be done to boost the sale of butter.
2. Find all rules that have “pepsi” in the antecedent. These rules may help to determine that what produce would be impacted if store discontinues selling pepsi.

3.2 Market Basket Analysis in Multiple Store Enviornment

In today’s business most of the companies have branches in different areas. To maintain economy of sales these stores. chains are growing in size. For example,Wal-Mart[3] is the largest supermarket chain in the world. The discovery of purchasing patterns in these multiple stores changes with time as well as location. In this multiple store chain basic association rules are not effective

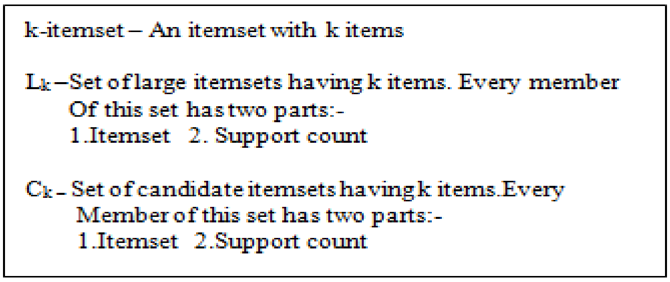
3.3 Market Basket Analysis Using Fast Algorithms

The problem of finding association rules using market basket analysis can be solved using the basic apriori algorithm[2]. But in applications like catalog design and customer segmentation the database used is very large. So, there is need of fast algorithms for this task.

4. Apriori Algorithm

4.1Working Principle[3]

1. Find all sets of items (itemsets) that have transaction support above minimum support. Itemsets with minimum support are known as large itemsets and all others as small itemsets.
2. Use the large itemsets to generate the desired rules.. For every large itemset l, find all non-empty subsets of 1. For every such subset a, find a rule which is of the form a → (1 - a) if the ratio of support(l) to support(a) is at least minconf



k-itemset – An itemset with k items

L_k –Set of large itemsets having k items. Every member
Of this set has two parts:-
1.Itemset 2. Support count

C_k - Set of candidate itemsets having k items.Every
Member of this set has two parts:-
1.Itemset 2.Support count

Fig. 2: Notations used.

4.2. Pseudo Code

Join Step: To generate C_k join L_{k-1} with itself.

Prune Step:Any (k-1) itemset which is not frequent cannot be a subset of frequent k-1 itemset.

```
$L_1$={large 1-itemsets};
for ( k = 2; $L_{k-1} \neq \emptyset$; k++) do begin
    $C_k$ = apriori-gen($L_{k-1}$);
    For all transactions t $\epsilon$ D do begin
        $C_t$ = subset($C_k$, t);
        for all candidates c $\epsilon$ $C_t$ do
            c.count++;
    end
    $L_k$ ={c $\epsilon$ $C_k$ | c.count $\geq$ minsup}
end
Answer = $U_k$ $L_k$;
```

Fig. 1: Algorithm Apriori.

5. Problems and Directions

In this paper we discussed various existing data mining algorithms. Every algorithm has its own advantage and disadvantage. This section provides some of the drawbacks of the existing algorithms and the techniques to overcome those difficulties. Among the methods discussed for data mining, apriori algorithm is found to be better for association rule mining. Still there are various difficulties faced by apriori algorithm. The various difficulties faced by apriori algorithm are-

1. It scans the database lot of times. Every time the additional choices will be created during the scan process. This creates the additional work for the database to search. Therefore database must store huge number of data services. This results in lack of memory to store those additional data.
2. Frequent item in the larger set length of the circumstances, leads to significant increase in computing time.

Those drawbacks can be overcome by modifying the apriori algorithm effectively. The time complexity for the execution of apriori algorithm can be solved by using the fast apriori algorithm. This has the possibility of leading to lack of accuracy in determining the association rule. To overcome this, the fuzzy logic can be combined with the apriori algorithm. This will help in better selection of association rules for market basket analysis.

6. Conclusions

Due to exponential growth [4]of computer hardware and system software technology there is large supply of powerful and cost effective computers. This technology provides a huge number of databases and information repositories available for transaction management information retrieval and data analysis. Physical analysis of this large amount of data is very difficult[7].This has lead to the necessity of data mining tools. Association rule mining and classification technique to find the related information in large databases is becoming very important in the current scenario.The large quantity of information collected through the set of association rules can be used not only for illustrating the relationships in the database, but also used for differentiating between different kinds of classes in a database. This paper provides some of the existing data mining algorithms for market basket analysis. The analysis of existing algorithms suggests that the usage of association rule mining algorithms for market basket analysis will help in better classification of the huge amount of data. The apriori algorithm can be modified effectively to reduce the time complexity and enhance the accuracy.

References

[1] Jiawei Han and Micheline Kamber ,Data Mining: Concepts and Techniques, 2nd ,March 2006,Chapter .

[2] R. Agrawal, R. Srikant, “Fast algorithms for mining association Rules”, Proceedings of the 20th VLDB Conference, Santiago, Chile, 1994, pp. 478–499.

[3] Yen-Liang Chen, Kwei Tang, Ren-Jie Shen, Ya-Han Hu,“Market basket analysis in a multiple store environment”, SciVerse ScienceDirect, Volume 40, Issue 2, August 2005, Pages 339-354 .

[4] Raorane A.A, Kulkarni R.V, and Jitkar B.D, “Association Rule – Extracting Knowledge Using Market Basket Analysis” ,Research Journal of Recent Sciences, Vol. 1(2), 19-27, Feb. (2012)

[5] J. Han, Y. Fu, Mining multiple-level association rules in large databases, IEEE Transactions on Knowledge and Data Engineering 11 (5) (1999) 798–805.

[6] R. Agrawal, T. Imielinski, and A. Swami. Mining association rules between sets of items in large databases. In Proc. of the ACM SIGMOD Conference on Management of Data, Washington, D.C., May 1993.

[7] Dr. M. Dhanabhakyam , Dr. M. Punithavalli Vaibhav Pandey, “A Survey on Data Mining Algorithm for Market Basket Analysis”, Global Journal of Computer Science and Technology ,Volume 11 Issue 11 Version 1.0 July 2011