

Mining Frequent Itemsets from Super Bazaar Data Repositories using Apriori Algorithm

Dr. Shivaji D. Mundhe¹, Mr. D.R. Vidhate²

¹Director, ¹Sinhgad Institute of Management and Computer Application (SIMCA),
Narhe (Ambegaon), Pune -411 041, Maharashtra, India.

²Research Student, College of Computer Application for Women, Satara, Maharashtra, India.

Abstract - Super bazaar is a self service shop. It provides everything under one roof. Consumer is important for super bazaar business. Consumers take many buying decisions every day. It is important to study purchase patterns of consumer visiting super bazaars. Knowledge Mining is a way of obtaining purchase patterns of consumers from super bazaar data repositories. Association rule mining is a major technique in the area of knowledge mining. Association rule mining finds frequent itemsets from a set of transactional databases. Apriori algorithm is used in the association rule mining. This paper gives overview of finding frequent itemsets from super bazaar databases using Apriori algorithm which will help further to generate strong association rules. This paper will provide valuable insight into buying behavior of consumer from various super bazaars.

Key Words: Super Bazaar, Consumer, Knowledge Mining, Apriori Algorithm, Association Rules, Frequent Itemsets.

1. INTRODUCTION

In recent years, there is explosive growth of transactional data in super bazaar databases. This has led to the development of techniques capable in the automatic extraction of knowledge from databases. Knowledge mining is a new technology which helps super bazaars to find hidden knowledge in data repositories. Knowledge mining is used to identifying valid, potentially useful and unknown patterns from a large amount of data [4]. On the basis of type of knowledge to be mined, there are different tasks involved in knowledge mining. One of the tasks is association rule mining.

2. ASSOCIATION RULE MINING

Association rules mining is an important branch of knowledge mining research. They are used for finding frequent patterns and associations among sets of items in transactional databases, relational databases, and other information repositories. [2] An association rule is the relationship between two disjoint itemsets, X and Y.

An association rule is of the form:- $X \Rightarrow Y$

$X \Rightarrow Y$: - When X occurs, Y also occurs.

Given a set of items $I = \{I_1, I_2, \dots, I_m\}$ and a database of transactions $D = \{t_1, t_2, \dots, t_n\}$ where

$t_i = \{I_{i1}, I_{i2}, \dots, I_{ik}\}$ and $I_{ij} \in I$, an association rule is an implication of the form $X \Rightarrow Y$ where $X, Y \subseteq I$ are sets of items called itemsets and $X \cap Y = \emptyset$.

Association rule mining has been used in a retailing where discovering of purchase patterns between products is very useful for decision making.

3. FREQUENT ITEMSETS

Frequent pattern analysis allows a researcher to systematically identify patterns that emerge from database. Frequent pattern mining comprises frequent itemset mining and association rule induction. The frequent itemset mining is basis for of association rule mining. It is method of market basket analysis. Frequent itemset plays very important role in many knowledge mining tasks that try to find interesting patterns from data repositories. Finding frequent itemsets are those with frequency larger than or equal to a user specified minimum support. The identification of sets of items, products and characteristics which often occur together in the given database can be seen as one of the most basic tasks in frequent itemset mining. The association rule mining can be reduced to mining frequent itemset. Once frequent itemsets are obtained, it is straightforward to generate association rules with confidence larger than or equal to a user specified minimum confidence [5].

4. MARKET BASKET ANALYSIS

Market Basket Analysis is a knowledge mining technique that is widely used to identify consumer patterns such that if customer buys certain group of items then customers

are likely to buy another group of items. Market basket analysis is an important component in retail organizations. It is a very useful technique for finding out co-occurrence of items in consumer shopping baskets. Such information can be used to provide the super bazaars with information to understand purchasing behavior of consumer in super bazaar. Market basket analysis is an important component of analytical system in retail organizations to determine the placement of goods, designing sales promotions for different segments of customers to improve customer satisfaction and hence the profit of the supermarket. [3]

5. SUPPORT

It is the measure of how often the collections of items in an association occur together as percentage of all transactions. Support(s) for an association rule $X \Rightarrow Y$ is the percentage of transactions in the database that contains $X \cup Y$. Every association rule has support. The rule that has very low support may occur simply by chance. A low support rule is not profitable to promote items that customers seldom buy together. So, support is often used to eliminate uninteresting rules. Association rule find all set of items that has support greater than minimum support. Support could be absolute or relative.

6. CONFIDENCE

Confidence for an association rule $X \Rightarrow Y$ is the ratio of the number of transaction that contain both antecedent and consequent to the number of transaction that contain only antecedent. A rule with low confidence is not meaningful. Confidence (α) for an association rule $x \Rightarrow y$ is the ratio of number of transactions that contains $X \cup Y$ to the number of transactions that contains X .

7. MINIMUM THRESHOLD VALUES

The strength of an association rule can be measured in terms of its support and confidence. The rules derived from itemsets with high support and high confidence. The number of association rules that can be derived from a dataset are large. Interesting association rules are those whose support and confidence are greater than minimum support and minimum confidence. The number of association rules discovered is affected by a user's decision concerning the minimum support threshold and minimum confidence threshold. Threshold values can be set by user or domain expert. It may be decided on the basis of number of transactions in database. Association rule need to satisfy a user specified minimum support and user specified minimum confidence at the same time. support and confidence values occur between 0% and 100%.

8. APRIORI ALGORITHM

The Apriori Algorithm is an influential algorithm for mining frequent itemsets for boolean association rules.[1]

• **Basic Conceptuations:**

1. Apriori Property:

- a. Any subset of frequent itemset must be frequent.
- b. An itemset is called a candidate itemset if all of its subsets are known to be frequent.

2. Join Operation:

To find L_k , a set of candidate k-itemsets is generated by joining L_{k-1} with itself.

3. Prune step:

Remove those candidates in C_k that cannot be frequent.

• **Apriori Algorithm Pseudo-code:**

C_k : Candidate itemset of size k

L_k : frequent itemset of size k

$L_1 = \{\text{frequent items}\};$

for ($k = 1; L_k \neq \emptyset; k++$) do begin

$C_{k+1} =$ candidates generated from L_k ;

for each transaction t in database do

increment the count of all candidates in C_{k+1} that are contained in t

$L_{k+1} =$ candidates in C_{k+1} with min_support

end

return $\cup_k L_k$;

• **Algorithm Steps:**

1. Find all frequent itemsets:

This step finds all frequent itemsets using minimum support count.

2. Generating Association Rules from Frequent Itemsets:

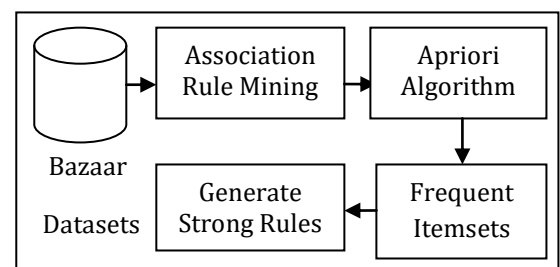
The frequent itemsets found in step (1) are used to generate association rules as:

For each frequent itemset "I", generate all nonempty subsets of I.

For every nonempty subset s of I, output the rule " $s \rightarrow (I-s)$ "

if $\text{support_count}(I) / \text{support_count}(s) \geq$ minimum confidence threshold.

9. FRAMEWORK TO GENERATE ASSOCIATION RULES



This framework applies association rule mining on transactional data from super bazaar data repositories. Apriori is widely used algorithm for performing association rule mining. Apriori initially finds frequent itemsets by using minimum support count, join and pruning. The frequent itemsets are then used to generate strong association rules by using minimum confidence threshold.

10. CONCLUSION

Knowledge mining in super bazaar sector can be used for market campaigns and to study buying behavior of consumers for profitability of the business. Association rule mining is a useful technique in the area of knowledge mining. Apriori algorithm is one of the important algorithm of association rule mining. Apriori could be used to find frequent items in a given transaction of database. Apriori algorithm find the tendency of a customer on the basis of frequently purchased itemsets. This paper has given the overview of Apriori algorithm as a tool used to find the hidden purchase pattern of the frequently used itemsets. The Super Bazaars industry will be more successful in this competitive market if adopted knowledge mining technology for market campaigns.

11. FUTURE RESEARCH

The research framework presented in this paper will be implemented by collecting data from selected super bazaar databases. Apriori algorithm will be applied on selected transactions in the databases. It will generate important purchase trends for the benefit of super bazaar. It is also possible find more specific patterns by selecting transactions based on demographic factors.

REFERENCES

- [1] http://www3.cs.stonybrook.edu/~cse634/lecture_notes/07apriori.pdf
- [2] Charanjeet Kaur,"Association Rule Mining using Apriori Algorithm: A Survey", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 2, Issue 6, June 2013.
- [3] Loraine Charlet, Annie M.C.; Kumar, D. Ashok," Market Basket Analysis for a Supermarket based on Frequent Itemset Mining" International Journal of Computer Science Issues (IJCSI);, Vol. 9 Issue 5, p257, Sep2012.
- [4] Margaret H. Dunham, S. Sridhar,"Data Mining: Introductory and Advanced Topics" Pearson Education, Inc., 2006.
- [5] Agrawal R, Srikant R," Fast algorithms for mining association rules", Proceedings of the 20thVLDB conference, pp 487-499, 1994.
- [6] Abhang Swati Ashok, JoreSandeep S.," The Apriori algorithm: Data Mining Approaches Is To Find Frequent Item Sets From A Transaction Dataset", International Journal of Innovative Research in Science, Engineering and Technology, Volume 3, Special Issue 4, April 2014.
- [7] Sotiris Kotsiantis, Dimitris Kanellopoulos," Association Rules Mining: A Recent Overview", GESTS International Transactions on Computer Science and Engineering, Vol.32 (1), pp. 71-82, 2006.
- [8] Ms Shweta,Dr. Kanwal Garg," Mining Efficient Association Rules Through Apriori Algorithm Using Attributes and Comparative Analysis of Various Association Rule Algorithms", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 6, June 2013.
- [9] Krutika. K.Jain, Anjali . B. Raut," Review paper on finding Association rule using Apriori Algorithm in Data mining for finding frequent pattern", International Journal of Engineering Research and General Science Volume 3, Issue 1, January-February, 2015.
- [10] Shilpi Singla, Arun Malik," Survey on various improved Apriori Algorithms", International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 11, November 2014.