RESEARCH PAPER | Management | Volume : 5 | Issue : 8 | August 2015 | ISSN - 2249-555X

Market Basket Analysis by Using Apriori Algorithm in Terms of Their Effectiveness Against Various Food Product

KEYWORDS: Market Basket,Data Mining, Frequentitemset,Apriori algorithm, Association rules, cross-selling,up-selling.

Teena vats

Assistant Professor, RCIT,Najafgarh, New Delhi-110043

ABSTRACT *Many established business enterprises accumulate large quantities of data from their day-to-dayoperations. For example, huge amounts of consumer purchase data are collected daily at thecheckout sales counter of the grocery stores. Market Basket Analysis is search for meaningfulassociation rules in the form of statements such as "People who buy milk are likely to buybread" in customer purchase data. Such valuable information can be used for the purpose ofcross-selling and up-selling , in addition to influencing sales promotions, store design and discount plans.*

In this thesis we present a methodology known as Association Rules Mining which is one of themain application areas in Data Mining and is useful for discovering interestingrelationshipswithin items hidden in large data sets. We give an overview of the problem and explainapproach that have been used to attack this problem. We then define Apriori algorithm forfinding frequent itemsets and then using these to determine association rules which highlightgeneral trends in the supermarket database.

INTRODUCTION

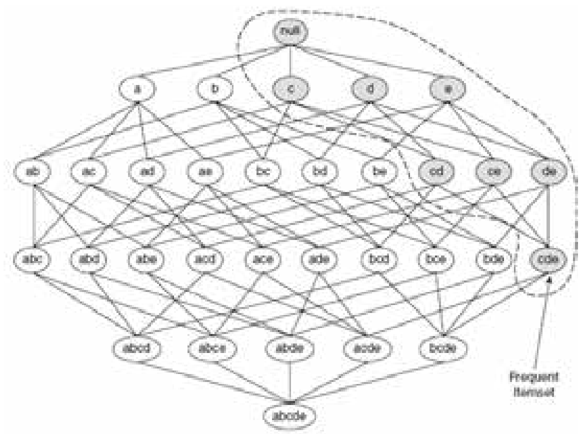
Market Basket Analysis is one of the most regularly used data mining technique used for data analysis in Marketing and Retailing. The purpose of Market Basket Analysis is to understand buying prototype of the customers and to determine what products customer purchase together. It takes its name from the idea of customers throwing all their purchases into a shopping cart (a ―Market Basket‖) for the duration of grocery shopping. Knowing what commodities people purchase as a group can be very helpful to a vendor or to any other company. A store could use this information to place products frequently sold together into the same area. Many practioners think that promoting two products together with high association value is profitable. But if both products sold independently at different times will result in higher returns. A catalog or world wide web merchant could also use association rules to determine the layout of their catalog and order form. Direct Marketers could use the basket analysis results to determine what new products to offer their prior customers.

Mining Association Rules also called Market Basket Analysis is one of the application areas of Data Mining. Consider a market with a collection of massive customer transactions. An association rule is A=>B where A is called as antecedent and B is the resultant. A and B are sets of items and the rule means that customers who buy A are likely to buy B with probability %c where c is called the confidence. Such a rule may be ―Ninety percent of people who buy Milk also buy Bread―. Data preprocessing operations can be performed using WEKA 3.6.9

THE APRIORI ALGORITHM

This section describes how the support measure helps to reduce the amount of candidateitem sets explored during frequent item set generation. The use of support for pruning candidate item sets is guided by the following principle.

(Apriori Principle). *If an item set is frequent, then all of its subsets must also be frequent.* To illustrate the idea behind the Apriori principle, consider the item set lattice shown in Figure



Suppose {c, d, e} is a frequent itemset. Clearly, any transaction that contains {c, d, e} must also contain its subsets, {c, d}, {c, e}, {d, e}, {c}, {d}, and {e}. As a result, if {c, d, e} is frequent, then all subsets of {c, d, e} (i.e., the shaded itemsets in this figure) must also be frequent.

Frequent Itemset generation in the Apriori algorithm

Apriori is the first association rule mining algorithm that pioneered the use of support based Pruning to systematically control the exponential growth of candidate itemsets. Consider the following Market Basket transactions shown in Table 2.1. Each row in this table corresponds to a transaction which contains a unique identifier labelled TID and a set of items bought by a given customer.

TID	ITEMS
1	{ Bread, Milk }
2	{ Bread, Butter, Cold Drink, Biscuit}
3	{ Milk, Butter, Cold Drink, Cake }
4	{ Bread, Milk, Butter, Cold Drink }
5	{ Bread, Milk, Cold Drink, Cake }

Volume : 5 | Issue : 8 | August 2015 | ISSN - 2249-555X

Provides a high-level illustration of the frequent itemset generation part of the Apriori

Algorithm for the transactions shown in Table 1.1. We assume that the support threshold is 60%, which is equivalent to a minimum support count equal to 3.

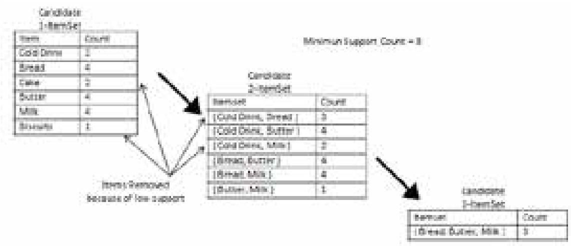


Illustration of frequent itemset generation using the *Apriori* algorithm.

Initially, every item is considered as a candidate 1-itemset. After with their supports, the Candidate itemsets {Cake} and {Biscuit} are discarded because they appear in fewer than three transactions. In the next iteration, candidate 2-itemsets are generated using only the frequent 1-itemsets because the Apriori principle ensures that all supersets of the infrequent 1-itemsets must be infrequent. Because there are only four frequent 1-itemsets, the number of candidate 2-itemsets generated by the algorithm is 6. Two of these six candidates, {Cold drink, Bread} and{Cold drink, Milk}, are subsequently found to be infrequent after computing their shore up values. The remaining four candidates are regular, and thus will be used to generate candidate 3-itemsets. With the Apriori principle, we only need to keep candidate 3-itemsets whose subsets are frequent. The only candidate that has this property is {Bread, butter, Milk}.

Algorithm 1.1 Frequent itemset generation in the Apriori algorithm.

1: k = 1.

2: Fk = $\{ i \mid i \in I \wedge \sigma(\{i\}) \geq N \times$ *minsupp*}. {Find all frequent 1-itemsets}

3: **repeat**

4: k = k + 1.

5: Ck = apriori-gen(Fk−1). {Generate candidate itemsets}

6: **for** each transaction $t \in T$ **do**

7: Ct = subset(C k, t). {Identify all candidates that belong to t}

8: **for** each candidate itemset $c \in$ Ct **do**

9: $\sigma(c) = \sigma(c) + 1$. {Increment support count}

10: **end for**

11: **end for**

12: Fk = { c | c ∈ Ck ∧ $\sigma(c) \geq N \times$ *minsupp*}. {Extract the frequent k-itemsets}

13: **until** Fk = ∅

14: □ Fk

CONCLUSIONS AND FUTURE WORK

Apriori algorithm is one of the fastest and earliest tool for Association Mining. We used the Apriori algorithm for mining association rules in large database of Reliance Fresh. As seen in the result in the previous section , what we knew about the data is the ID's of the items. The Association rules were mined from the dataset. Let me give example of rules derived from large database set: *"The people who bought milk and cake bought biscuits with confidence*

80%". These types of rules can help the store owner to place these products together in a store to achieve maximum profits. We blindly mined the data and found the association rule. The data consisted of items and the amount of items. Also, Some of the amount were in fractions i.e. in kilograms and some of them were in numbers. We did not use amount of data in mining of the rules. We had to take the data as binary i.e. it exist in the transaction or not.

A future work could be to use the information i.e. the amount of items and their prices for deriving more meaningful rules. Also this study was only based on data got from one of the

markets of the chain Reliance Fresh. A future work could be to analyze the data from many of the branches of the chain to cover more customers and determine their buying behavior. This geographical based analyzation would help to understand the customer better and achieve more useful and meaningful results.

REFERENCE

1. Brin, S., R. Motwani and J. D. Ullman and S. Tsur, \Dynamic itemset counting andimplicationrules for market basket data", Proceedings of the ACM SIGMOD In-ternational Conference on Management of Data, Vol. 26, No. 2, pp. 255{264,New York, May 1997. || 2. Hipp, J., U. GÄuntzer and G. Nakhaeizadeh, \Mining Association Rules: Deriving a Superior | Algorithm by Analysing Today's Approaches", Proceedings of the 4thEuropean Conference on Principles and Practice of Knowledge Discovery, Lyon,France, September 2000. | 3. Trnka., "Market Basket Analysis with Data Mining Methods", International Conference on Networking and Information Technology (ICNIT) ,2010. || 4. 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) || 5. Srinivasan G, "Services Sector and its Contribution to the Indian Economy", Yojana, Volume 55, pp. 5-7, September 2011 696 IJRITCC | September 2013, Available @ http://www.ijritcc.org || 6. Anand H.S. and Vinodchandra S.S.," Applying Correlation Threshold on Apriori Algorithm," 2013 IEEE International Conference on Emerging Trends in Computing, Communication and Nanotechnology (ICECCN 2013). || 7. Lekha, Dr. C V Srikrishna and Dr. Viji Vinod," Utility of Association Rule Mining: a Case Study using Weka Tool,"2013 IEEE. || 8. Agrawal, R. , Imielinski, T. and Swami, A. 1993. Mining Association Rules between Sets of Items in Large Databases, Proceedings of the ACM SIGMOD International Conference on Management of Data. Association Rule Learning 2011. Retrieved January 12, 2014 from http://en. wikipedia. org/wiki/ Association_ rule_learning Jason Brownlee on March 17, | 2014 in Uncategorized A scalable algorithm for the market basket analysis - ResearchGate. Availablefrom: