International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 05 Issue: 02 | Feb-2018 www.irjet.net p-ISSN: 2395-0072

Market Basket Analysis based on frequent Itemset Mining

Mahesh Behera[1], Ankush Fartale[2], Aniket Bhagat[3], Prof. Nidhi Sharma[4]

[1, 2, 3]Department of Computer Engineering, Bharati Vidyapeeth College of Engineering, Navi Mumbai
[4]Professor, Department of Computer Engineering, Bharati Vidyapeeth College of Engineering, Navi Mumbai

---

Abstract – *Market Basket Analysis is an effective tool in retail industry which will help the market owner to increase the business and improve sales distribution techniques. This is totally done by association rule mining in which it analyses the customer behavior against the purchasing item from market. It analyses the customer purchasing pattern and generate frequent itemsets. After generation of frequent itemset it is easy to find most popular itemset and worst item combination from large transactional data instead of reading it manually. Generation of frequent itemset will enhance the market strategy, placement of goods and many more. Which results in increase in sales of goods and any one can do profitable business.*

Key Words: Association Rules, Frequent Itemsets, Apriori, Market Basket Analysis.

1. INTRODUCTION

One of the challenges for companies that have invested a lot in consumer data collection is how to mine important information from their vast customer databases and product feature databases, in order to gain economical advantage. Several aspects of market basket analysis have been studied in academic literature, such as using customer interest profile and interests on particular products for one to one marketing, purchasing patterns in a multi-store environment to improve the sales. Market basket analysis has been intensively used in many companies as a means to discover product associations and base a retailer's promotion strategy on them.

A retailer must know the needs of customers and adapt to them. Market basket analysis is one achievable way to find out which items can be place together. Market basket analyses gives retailer good information about related sales on group of goods basis Customers who buys bread often also buy several products related to bread like milk, butter or jam. It makes sense that these groups are placed side by side in a retail center so that consumers can access them promptly. Such related groups of goods also must be located side-by-side in order to remind customers of related items and to lead them through the center in a logical manner. Market basket analysis determines which products are bought together and to design the supermarket arrangement, and also to design promotional campaigns. Therefore the Market consumer behaviors need to be analyzed which can be done during dissimilar data mining techniques.

Well-versed decision can be made easily about product placement, pricing, endorsement, profitability and also finds out if there are any successful products that have no significant related elements. Similar products can be found so those can be placed near each other or it can be cross-sold.

2. Problem Statement

Nowadays people buy daily goods from super market nearby. There are many supermarkets that provide goods to their customer. The problem many retailers face is the placement of the items. They are unaware of the purchasing habits of the customer so they don't know which items should be placed together in their store. With the help of this application shop managers can determine the strong relationships between the items which ultimately helps them to put products that co-occur together close to one another. Also decisions like which item to stock more, cross selling, up selling, store shelf arrangement are determined.

3. Objectives

The main objective of Market Basket Analysis is to get better efficiency of market and sales strategy using consumer transactional data collected during the sales transaction.

To spot the frequent items on or after the transaction on the basis of support and confidence.

To generate the association rules from the frequent item sets.

4. Proposed System

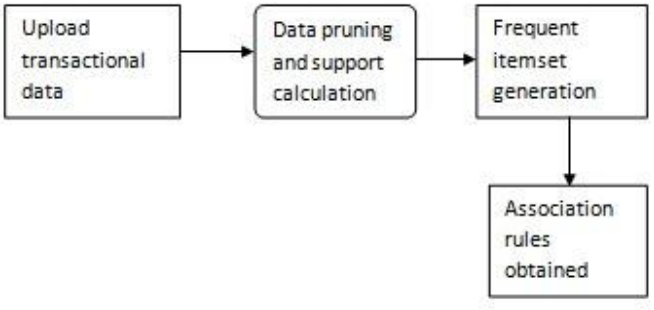


Fig – 1: Flow of Proposed System

Our Basket Analysis system itself gathers Frequent Itemsets from set of Transaction & other resources, which are then classified according to their semantic orientation and intensity. The proposed system is software which will collect frequent items from multiple transactional databases and then obtained data will be analyzed for association mining. There are multiple algorithms available for association rule mining out of those Apriori Algorithm is used in our Basket Analysis system.

In our Basket Analysis system the software will preset support and confidence for association rule mining which will use Apriori algorithm by default for frequent itemset mining.

5. Methodology

1. Data Collection: The data was collected from http://www.salemmarafi.com/wpcontent/uploads/2014/03/groceries.csv due to the unavailability of data from the supermarkets.

2. Data Pre-processing: The data collected was mapped manually as integer values. For example "Fruit" was labeled as 1, "Bread" as 2, "Soups" as 3 and so on. The mapped integer's values were then saved in a text file and given as the input to the system.

3. Apriori Algorithm: The Apriori algorithm was used for processing the input data and result was produced as the list of rules that are strongly associated with each other.

6. Association Rule Mining

Association rule mining is generally used to extract the interesting correlation, frequent pattern, association among sets of items in database. It consists of two main measures: support and confidence. Support can be defined based on minimum value or more than minimum value. Confidence can be defined how many times the number of statement has become true. Association rule mining finds the rules which satisfy the minimum support and confidence.
Association rule mining is a two-step process: First step is to find frequent item that is less than minimum support.
Second step is to combine all frequent items. We will look at some of these useful measures such as support and confidence.

Support: -It is the percentage of number of customer who purchased item A and item B,

Support = Probability (A and B)

Support = (# of transactions involving A and B) / (total number of transactions).

Confidence: - is the strength of implication of a rule; it is the percentage of number of customer who purchased item A and item B together,

i.e. Confidence = Probability (B if A) = P(B/A) Confidence = (# of transactions involving A and B) / (total number of transactions that have A).

7. Apriori Algorithm

Apriori is a one of the famous, most important, and scalable algorithm for mining frequent Itemsets and association rule mining. Apriori was introduced by Agrawal and Srikant in 1993. Apriori algorithm is used to find all frequent itemset in a given database. Apriori algorithm is to make multiple scan over database. It uses breath first strategy to search over items in the database. Apriori algorithm is being used by so many industry for transactional operation and also it can be used in real time applications like (shopping mall, general store, grocery shop etc.) by collecting the item bought by customer over the time so frequent item can be generated. Apriori algorithm requires two important things: minimum support and minimum confidence. First we can checks the item whether they are greater than or equal to minimum support and after than we can find the frequent item set. Second thing is minimum confidence constraint is used to form association rules.

General Process of the Apriori algorithm

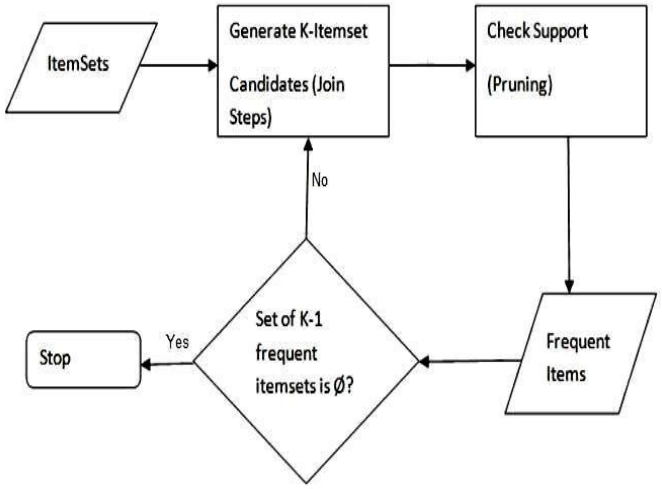
The entire algorithm can be divided into two steps:

Step 1: Apply minimum support to find all the frequent sets with k items in a database.

Step 2: Use the self-join rule to find the frequent sets with N+ 1 item with the help of frequent N-Itemsets. Repeat this process from N=1 to the point when we are unable to apply the self-join rule.

This approach of extending a frequent itemset one at a time is called the "bottom up" approach.

Cdn: Candidate itemset of size n
Ln: frequent itemset of size n
L1 = {frequent items};
For (n=1; Ln != Ø ; n++)
Do begin
Cdn+1 = candidates generated from Ln;
For each transaction T in database do
Increment the count of all candidates in
Cdn+1 that are contained in T
Ln+1= candidates in Cdn+1 with
min_support
End
Return Un Ln



Consider 10 transactions given below of minimum support is 20% and minimum confidence is 70%, determine the frequent item sets and association rules using the APRIORI algorithm.

TRANSACTION	ITEMS
T1	I1,I2,I3,I4
T2	I1,I4
T3	I1,I5
T4	I4,I5
T5	I7,I6,I5,I4
T6	I3,I2,I1
T7	I4,I3,I2,I1
T8	I3,I4,I5,I1
T9	I1,I2
T10	I2,I3,I1

Solution:

ITEM	SUB COUNT
I1	8
I2	5
I3	5
I4	6
I5	4
I6	1
I7	1

Min. Support Count= 20% of 10= 2
Omit less than this value in the above table

L1

ITEM	SUB COUNT
I1	8
I2	5
I3	5
I4	6
I5	4

C2

ITEM	SUB COUNT
{I1, I2}	5
{I1, I3}	5
{I1, I4}	4
{I1, I5}	2
{I2, I3}	4
{I2, I4}	2
{I2, I5}	-Nil-
{I3, I4}	3
{I3, I5}	1
{I4, I5}	3

Because min support is 2, omit the values less than 2 in the above table to produce L2

L2

ITEM	SUB COUNT
{I1, I2}	5
{I1, I3}	5
{I1, I4}	4
{I1, I5}	2
{I2, I3}	4
{I2, I4}	2
{I3, I4}	3
{I4, I5}	3

C3

ITEM	SUB COUNT
{I1, I2, I3}	4
{I1, I2, I4}	2
{I1, I2, I5}	-Nil-
{I2, I3, I4}	2
{I2, I3, I5}	-Nil-
{I3, I4, I5}	1

L3

ITEM	SUB COUNT
{I1, I2, I3}	4
{I1, I2, I4}	2
{I2, I3, I4}	2

Min Confidence=70%
Association Rules

a. {I1,I2} => {I3} = 4/5 = 80%
b. {I1,I2} => {I4} = 2/5 = 40%

{I2,I3} => {I4} = 2/4 = 50%
Max Combination is {I1, I2 and I3}.

We have Comparative Analysis of Association rule mining algorithms:

	Apriori	FP-Growth	Eclat
Approach	Breadth First Search	Divide & Conquer less than Apriori	Depth First Search less than both
Execution Time	Fast	Slow	Less than both
Advantage	Suitable for large Database and generate Frequent Itemset	Works on small Dataset	Not good for frequent item generation
Disadvantage	Multiple Scans	Consumes more memory	Consumes more memory

In recent researches, many algorithms were developed likeApriori, FP-Growth, Eclat etc.

We have comparatively analyzed various association rulesmining algorithm like Apriori, FP- Growth, Eclat etc. We also have compared these algorithm using same examples to understand their working. Major issues of all algorithms are not suitable for large dataset and we found this Apriori algorithm is able to generate frequent itemsets from large dataset in well manner as compared to other algorithm.

8. Requirement

Hardware Requirement:-

Hardware	Specification
Processor	Intel Core 2 Duo or above
RAM	2GB or more

Software Requirement:-

Software	Specification
Operating System	Microsoft Windows 7/8/10 or Unix
Platform	Python Django

9. Conclusion

Market basket analysis generates the frequent itemset i.e. association rules can easily tell the customer buying behavior and the retailer with the help of these concepts can easily setup his retail shop and can develop the business in future. The main algorithm used in market basket analysis is the Apriori algorithm. It can be a very powerful tool for analyzing the purchasing patterns of consumers. The three statistical measures in market basket analysis are support, confidence. Support measures the frequency an item appears in a given transactional data set, confidence measures the algorithm's predictive power or accuracy. In our example, we examined the transactional patterns of grocery purchases and discovered both obvious and not-so-obvious patterns in certain transactions.

10. References

1. Data Mining for Business Intelligence by Galit Shmueli, et. al. First edition published by Wiley.

2. Groceries data set is an open sources dataset referred from www.salemmarafi.com/code/market-basket analysis-with-r

3. Data Mining and Business Analytics with R by Johannes Ledolter. Published by John Wiley & Sons, year 2013.

4. Jiawei Han, Micheline Kamber, Jain Pei, "Data Mining Concept and Technique 3rd Edition". Jugendra Dongre, GendLal

5. Prajapati, S. V. Tokekar, "The Role of Apriori Algorithm for Finding the Association Rules in Data Mining", IEEE 2014.

6. https://towardsdatascience.com/a-gentle introduction-on-market-basket-analysis-association rules-fa4b986a40ce