

Market Basket Analysis

Ms. AVULA NAGABINDU

Mr.G.DILIP KUMAR

B. Tech Student, Department of CSE, Sri Mittapalli College of Engineering, Tummalapalem, NH-16, Guntur, Andhra Pradesh, India.

Assistant Professor, Department of CSE, Sri Mittapalli College of Engineering, Tummalapalem, NH-16, Guntur, Andhra Pradesh, India.

Abstract: - The project describe about the mba association rules. Although more than 20 years old, Market Basket Analysis (MBA) (or association rules mining) can still be a very useful technique to gain insights in large transactional data sets. The classical example is transactional data in a supermarket. For each customer we know what the individual products (items) are that he has put in his basket and bought. Other use cases for MBA could be web click data, log files, and even questionnaires.

With market basket analysis we can identify items that are frequently bought together. Usually the results of an MBA are presented in the form of rules. The rules can be as simple as $\{A \Rightarrow B\}$, when a customer buys item A then it is (very) likely that the customer buys item B. More complex rules are also possible $\{A, B \Rightarrow D, F\}$, when a customer buys items A and B then it is likely that he buys items D and F.

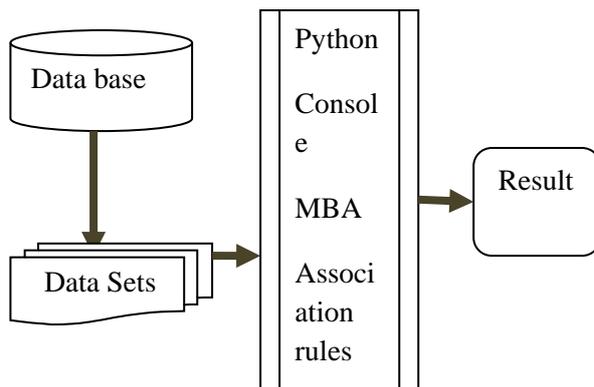
Keywords - Big social data, Social set analysis, Social business, Visual analytics, geo-spatial, GIS.

I INTRODUCTION

Association rule mining is one of the most important techniques of data mining. The

process of knowledge discovery in databases (KDD) includes a selection of data, its preprocessing, transformation, data mining and interpretation. The major goal of data mining algorithms is to extract the hidden predictive information and transform it into an understandable structure [8]. It aims at extracting interesting patterns, relations, associations among sets of items in databases. One of the major tasks of the ARM is to find the relationship among various data items in the database. An association is defined in the form of $A \Rightarrow B$, where A is the antecedent and B, is the consequent and the meaning of the rule is deduced as: A and B, both are item sets and the rule says that if a customer who purchases the A item is likely to purchase the B item as well with a conditional probability percentage factor known as %C where C is the confidence value of a rule. This helps the business managers to study the behavior and buying habits of the customers in order to increase their sales. Based on this study, items that are closely related or the items that have an attraction to each other are put under closed proximity. For instance, a customer who purchases milk is also likely to purchase bread together [5].

The interestingness measures such as support and confidence play a very important role in the association rule analysis. The support value of any transaction X with respect to T is defined as the proportion of the transactions in the dataset which contains item set X [4]. It is given as $Supp(X) = \frac{X \cup Y}{T}$. The confidence value of a rule is defined as the proportion of the transactions that contains X which also contains Y [4]. It is given by: $Conf(X \rightarrow Y) = \frac{Supp(X \cup Y)}{Supp(X)}$. The item sets that comply with minimum support and minimum confidence values are called strong association rules [5].



DATA LOCATION :

Data location means where the data is to be located.

DATA SET:

Collection of individual invoice data from the data location is known as dataset.

PYTHON CONSOLE :

The console is where you may enter, interact with and visualize data inside a command interpreter.

MBA ASSOCIATION RULES:

Association rules help to show the probability of relationships between data items between large data sets in various types of datasets.

II RELATED WORK

Trnka [5][9], in this paper, describes the implementation of market basket analysis to Six Sigma methodology. The methods of data mining provide great deal opportunities in the market sector. One of them is market basket analysis. By implementing this into Six Sigma, the results can be improved and the performance level of the process can be changed.

Yanthy et al. [5][10], in this paper, talks about the goal of data mining which is to reveal the hidden knowledge from provided data and the various algorithms that have been proposed so far. The interestingness of the rules can be determined by using various measures such as confidence, support, lift, information gain etc. since, not all the rules generated are of interest to any given user. In this paper, he studied the relationship between interesting measures and algorithms.

Cunningham et al. [5][11], he provided a model for library circulation data and applied the Apriori tool for the task of detecting subject classification categories that co-occur in

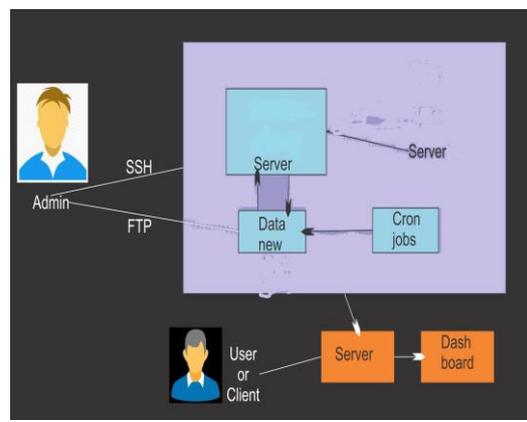
transaction records of the library borrowed books from the university. The results of the paper provide insight into the degree of “scatter” that the classification scheme foster in a particular collection of documents.

Rastogi et al. [5][12], presented in his paper the optimized association approach on association rules that contain instantiated attributes. To determine the relationship between two items such that the support and confidence of the optimized rule are maximized. He presented effective techniques for pruning the search space Neesha et al. [8] studied the various advancements in the field of data mining. In her paper, she described these advancements starting from the year 2008, a novel frequent pattern generation algorithm had been proposed in order to tackle the data imbalance problem. In 2009, an experiment was performed to compare three association rule mining algorithms: Apriori, Predictive Apriori, and Tertius, on the basis of predictions made on the status of the heart using heart disease data. The results of the experiment showed that Apriori was best suited for this type of data.

In 2010, a new algorithm SC-BF Multilevel was introduced as a better version of Apriori algorithm which was faster and efficient since it required an only a single scan of the database for mining frequent item sets. In 2012, three data mining techniques were applied upon a specific set of data consisting of students’ enrollment for the likeliness of the courses to be learned. The techniques involved clustering (k-means

algorithm), classification (Andree algorithm) and association rule (Apriori algorithm). The comparison depicted that the combined approach is better than using association rule mining alone for such kind of task.

III DESIGN OF THE WORKFLOW:



Market basket analysis is a technique that helps us in determining which products tends to be purchased together in accordance with the association rules. The primary objective is to improve the effectiveness of sales and marketing strategies with the help of previously obtained customer data. Association rules aim to identify those items which frequently occur in a database [7]. This paper presents each item is represented by a Boolean value, that is 0 and 1, where 0 represents that item is not present, whereas 1 represents that item is present. We have proposed a novel data structure, FPTree, frequent pattern mining, overcomes the main bottlenecks of Apriori. The frequent itemsets are generated only with two scans of the database. It is an extended prefix tree structure which is used for the storage of information about patterns.

The nodes of the tree are arranged in such a way that the nodes occurring more frequently will have better chances of sharing nodes than nodes occurring less frequently. Association performs better than Apriori because there is no candidate set generation, as well as the length of the frequent item, sets increases as support value decreases. Association algorithm is more efficient than the later one [2].

There exist many other algorithms for mining of frequent itemsets viz., Apriori and Eclat, Association growth algorithm preprocesses the database only twice as follows: an initial scan of the database determines the frequencies of the items. All the uncommon items- the items that do not appear in a minimum number of user-specified transactions- are discarded from it as they cannot be a part of frequent itemsets [1].

In addition to this, all the items in the transaction are sorted in descending order in reference to their frequencies. Whilst the algorithm does not depend upon the specific order of the frequencies of items, sorting in descending order may lead to much less execution time than ordered randomly. Sorting in ascending order leads to slower operations implementing even worse than in a random order [1].

This preliminary processing is demonstrated in table 1. The descending order sorted items are shown in the middle of the table. The user-specified minimal support discards all the items that are infrequent. If a user-specified support value is given as 3 then, items d and e are pruned.

IV EXPERIMENT RESULTS

A B D	B=7	B A
B D	A=6	B
B C	C=6	B C
A B D	D=3	B A
A C	E=1	A C
B C		B C
A C		A C
A B C E		B A C
A B C		B A C

Sample collection of Data

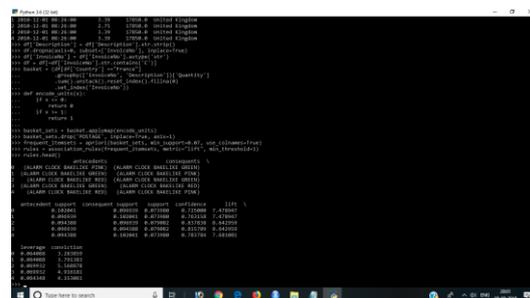
Algorithm Steps:

Asd

Fe

S. No.	Premises	Conclusion	Support	Confidence (%)
1.	A, D	A	1	10
2.	C, A, B, E	A	1	10
3.	A, B, E	B	1	10
4.	C, B, E	A	1	10
5.	B, E	B	1	10
6.	A, E	B	1	10
7.	C, E	A	1	10
8.	B, A	D	3	30
9.	B, C	D	3	30

Results generated based on the association rules



Association rules generated on the data.

V CONCLUSION

Market Basket Analysis with the help of association rules, the combination of products which makes the customer to buy more items and analyze their behavior helps to increase the business in further.

.VI FUTURE ENHANCEMENTS

This research essentially deals with effective recommendation system for the supermarket for identifying the most frequently purchased items and suitable crop respectively using association rule techniques. These proposed techniques are much useful in both supermarket . The proposed Fast Adaptive Association Rule Mining approach provides the best result among the proposed techniques. In order to improve the performance of these proposed approaches, some future enhancements have to be done. The main aim of the future enhancements would be to increase the effectiveness.

VI REFERENCES

- [1] Sotiris Kotsiantis, DimitrisKanellopoulos, "Association Rule Mining: A Recent Overview", GESTS International Transactions on Computer Science and Engineering, Vol. 32(1), 2006, pp. 71-82
- [2] https://en.wikipedia.org/wiki/Association_rule_learning
- [3] Phani Prasad, MurlidherMourya, "A Study on Market Basket Analysis Using a Data Mining Algorithm", International Journal of Emerging

Technology and Advanced Engineering, ISSN 2250-2459, ISO 9001:2008 Certified Journal, Vol. 3, Issue 6, June 2013.

[4] Akanksha Singh, K. K. Singh, Data Mining and Data Warehousing, India: Umesh Publications, 2011-2012

[5] Harpreet Kaur, Kawaljeet Singh, "Market Basket Analysis of Sports Store using Association Rules", International Journal of Recent Trends in Electrical & Electronics Engg., ISSN: 22316612, Dec. 2013

[6] Trnka., "Market Basket Analysis with Data Mining Methods", International Conference on Networking and Information Technology (ICNIT), 2010

[7] W Yanthy, T. Sekiya, K. Yamaguchi, "Mining Interesting Rules by Association and Classification Algorithms", FCST 09

[8] Cunningham, S. J. and Frank, E., "Market Basket Analysis of Library Circulation Data", International Conference on Neural Information Processing, Vol. 2, 1999

[9] Rastogi, R. and Kyuseok Shim, "Market Optimised Association Rules with Categorical and Numerical Attributes", IEEE Transactions on Knowledge and Data Engineering, Vol. 14, 2002

AUTHORS :

M.ANUSHA : B.Tech Student, Department of CSE, Sri Mittapalli College of Engineering, Tummalapalem, on NH-5, Guntur Andhra Pradesh, India.

K.V.S. RAMESH BABU : B. Tech Student, Department of CSE, Sri Mittapalli College of



Engineering, Tummalapalem, on NH-5, Guntur
Andhra Pradesh, India.

M.S.V. LAKSHMI SAHITHI : B. Tech

Student, Department of CSE, Sri Mittapalli
College of Engineering, Tummalapalem, on
NH-5, Guntur Andhra Pradesh, India.

B.SNEHA VIJAYA LAKSHMI : B.Tech

Student, Department of CSE, Sri Mittapalli
College of Engineering, Tummalapalem, on
NH-5, Guntur Andhra Pradesh, India.