Study on the Customer targeting using Association Rule Mining

Surendiran.R[(1)] , Rajan.K.P[(2)] , Sathish Kumar.M[(3)]

[(1)]Lecturer MCA dept, Madurai Kamaraj University

[(2)]Thiagarajar School of Management, Madurai, India

[(3)] Student of MCA dept, Madurai kamarajUniversity

Abstract— Data mining is one of the widest area where many researches takes place to mine desired and hidden data. There are many different approaches to find the hidden data. This paper deals with Frequent Pattern growth algorithm which follows association rule concept to group the required data items. Using this method of mining time can be reduced to a greater extent. This paper contains implementation of a real time system; the implementation is about making a survey on the group of people and their mobile connection's service providers. End result contains the set of people from a particular age group with their support and confidence for the service provider they have chosen. Based on which any decisions can be made by service providers to enhance their business and attain many customers.

Keywords- Data mining, Frequent pattern, Support confidence.

I. INTRODUCTION

Data mining is a process of analyzing data from different perspectives and summarizing it into useful information. It uses sophisticated data analysis tools to discover previously unknown, valid patterns and relationships in large data sets. These tools can include statistical models, mathematical algorithms and machine learning methods. The algorithms used in data mining are used in such a way that it can improve their performance automatically through experience such as neural networks or decision trees. This data mining in simple terms can be said as knowledge mining, since data mining deals with extracting or mining of knowledge. It attempts to discover hidden rules underlying the data and for this reason it is also called as data surfing. It is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses. Data mining tools predict future trends and behaviours, allowing businesses to make proactive, knowledge-driven decisions. Data mining tools can answer business questions that traditionally were time consuming to resolve. Databases with hidden patterns, finding predictive information that expert may miss because it lies outside their expectations. Most companies already collect and refine massive quantities of data. Data mining techniques can be implemented rapidly on existing software and hardware platforms to improve the value of existing information resources, and can be integrated with new products and systems. Examples of profitable applications illustrate its relevance to today's business environment as well as a basic description of how data warehouse architectures can evolve to deliver the value of data mining to end users. Mining frequent patterns in transaction databases, time-series databases, and many other kinds of databases has been studied popularly in data mining research. Most of the previous studies adopt an Apriori-like candidate set generation-and-test approach. However, candidate set generation is still costly, especially when there are prolific patterns and/or long patterns. In this study, we propose a novel frequent pattern tree (FP-tree) structure, which is an extended prefix- tree structure for storing compressed, crucial information about frequent patterns, and develop an efficient FP-tree-based mining method, FP-growth, for mining the complete set of frequent patterns by pattern fragment growth. Efficiency of mining is achieved with three techniques: (1) a large database is compressed into a highly condensed, much smaller data structure, which avoids costly, repeated database scans, (2) our FP-tree based mining adopts a pattern fragment growth method to avoid the costly generation of a large number of candidate sets, and (3) a partitioning-based, divide-and-conquer method is used to decompose the mining task into a set of smaller tasks for mining confined patterns in conditional databases, which dramatically reduces the search space. Our performance study shows that the FP-growth method is efficient and scalable for mining both long and short frequent patterns, and is about an order of magnitude faster than the Apriori algorithm and also faster than some recently reported new frequent pattern mining methods.

The most commonly used techniques in data mining are:

- Artificial neural networks: Non-linear predictive models that learn through training and resemble biological neural networks in structure.
- Decision trees: Tree-shaped structures that represent sets of decisions. These decisions generate rules for

the classification of a dataset. Specific decision tree methods include Classification and Regression Trees (CART) and Chi Square Automatic Interaction Detection (CHAID) .

• Genetic algorithms: Optimization techniques that use process such as genetic combination, mutation, and natural selection in a design based on the concepts of evolution.

II. ASSOCIATION RULE

Association means grouping of related items from a set. A simple example is analyzing a large database of supermarket transactions with the aim of finding association rules. This is called Association rules mining or market basket analysis. Association rule mining is used to:

- Find frequent patterns.
- Find associations.
- Find correlations.

Among the set of items or objects in transactional databases, relational databases. This can be used by the retailers, entrepreneur in order to make any advertisement, improvement in their business. The market based analysis find customer's purchasing habits. This analysis is done onto the customer basket to identify the frequent combination of products. Market Basket Analysis is a technique that assists in understanding what items are likely to be purchased together according to the association rules, primarily with the aim of identifying cross-selling opportunities. A super market can use this technique to organize and place products frequently sold together into the same area. The direct marketers can use the MBA to find what new products to offer their customers. The application of market basket analysis is generally facilitated by the use of the data mining tools. Using this analysis product in demand can be identified by marketers and "combined take rates" of the products can be known. The combined take rates are defined as - how often the items are bought together. In a data base, this can be answered with a query. When there are 100 products, it will take thousands of queries to get the "most popular basket". Association rule proposed the support-confidence measurement framework and reduced association rule mining to the discovery of frequent item sets.

Two basic Entities of association rule are:

Support-> Support is a measure of what fraction of the population satisfies both the antecedent and the consequent of the rule.

Confidence->Confidence is a measure of how often the consequent is true when the antecedent is true.

Five different algorithms are used in development of association rules. They are AIS, SETM, Apriori, AprioriTID, Apriori Hybrid.

III. APRIORI

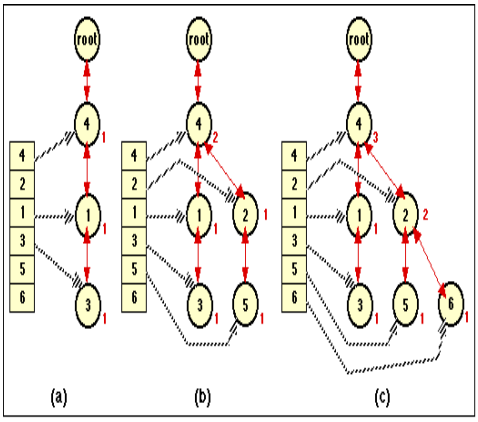
Apriori is an efficient association rule mining algorithm that explores level-wise mining. The Apriori Algorithm is an influential algorithm for mining frequent item sets for Boolean association rules. All combination of items in a set of transactions that occurs with a specified minimum frequency. These combinations are called frequent itemsets. Apriori calculates the probability of an item being present in a frequent itemset, given that another item or items is present. Apriori discovers patterns with frequency above the minimum support threshold. Therefore, in order to find associations involving rare events, the algorithm must run with very low minimum support values. However, doing so could potentially explode the number of enumerated itemsets, especially in cases with a large number of items. This could increase the execution time significantly. Classification or anomaly detection may be more suitable for discovering rare events when the data has a high number of attributes. There are few drawbacks in using Apriori algorithm like multiple scans of transaction database are to be made, huge number of itemsets were to be checked, tedious workload in counting support.

IV. FREQUENT PATTERN (FP)

Frequent Pattern growth algorithm is an algorithm which is a refined model of Apriori algorithm. In Apriori multiple transactions are done to extract the frequent itemsets, whereas in FP growth approach the multiple transactions are being reduced by forming a tree structure which contains the frequent itemsets. It is based on a prefix tree representation of the given database of transactions (called an FP-tree), which can save considerable amounts of memory for storing the transactions.

To illustrate the method, considering the simple data set:

{1, 3, 4}
{2, 4, 5}
{2, 4, 6}



The basic idea of the FP-growth algorithm can be described as a recursive elimination scheme: in a preprocessing step delete all items from the transactions that are not frequent individually, i.e., do not appear in a user-specified minimum number of transactions. Then select all transactions that contain the least frequent item (least frequent

among those that are frequent) and delete this item from them. Recurse to process the obtained reduced database, remembering that the item sets found in the recursion share the deleted item as a prefix. On return, remove the processed item also from the database of all transactions and start over, i.e., process the second frequent item etc. In these processing steps the prefix tree, which is enhanced by links between the branches, is exploited to quickly find the transactions containing a given item and also to remove this item from the transactions after it has been processed.

V. FP GROWTH IMPLEMENTATION

We have implemented this Frequent Pattern algorithm to get a survey in Service providers. The collected information by us contains many records like name, address, mobile number, age, salary, family members, preferred service provider etc which belong to different databases. All these details from the collected information are being used to manipulate or mine to find out the hidden data into it. Here we have manipulated the Age, Income and Service provider. Using all the above details, the support and confidence of a group of

People belonging to a particular age group having salary to a certain amount are mined. This result can be used by any service provider to improve their services to attract those users who prefer other service providers. They can also use this result in enhancing their business by making any business promotions.

Advantages:-

- Business promotions can be made.
- Enhancing of business circle can be made.

VI. OUTPUT

AGE	INCOME	SERVICE	SUPPORT	CONFIDENCE
20-30	10000	Airtel	30.12	75.00
20-30	20000	Airtel	25.11	60.00
30-40	10000	Aircel	40.12	50.08
30-40	20000	Aircel	35.30	60.35

ID	AGE	INCOME	SERVICES
100	20-30	10000	Airtel
101	20-30	20000	Aircel
102	30-40	10000	Airtel
103	20-30	20000	Airtel
104	30-40	10000	Aircel
105	20-30	10000	Aircel

VI. CONCLUSION

The hidden data in a large data warehouses is mined easily by segregating based on their occurrence. The frequently used items are further processed to get a reduced database with

VII. REFERENCES

[1] Introduction to Data mining with case studies **– G.K.Gupta**
[2] Data mining techniques – **Arun K pujari**
[3] Data Mining concepts and techniques- **Jiawei Han Micheline Kamber**
[4] R. Agrawal, T. Imielinski, and A.N. Swami, "Mining association rules between sets of items in large databases," Proceedings of ACM SIGMOD International Conference on Management of Data, ACM Press, Washington DC, pp.207-216, May 1993.
[5] Mohammed J. Zaki, "Scalable Algorithms for Association Mining," IEEE Transactions on Knowledge and Data Engineering, vol.12, no. 3,pp. 372-390, May/June 2000.
[6] J. Han, J. Pei, and Y. Yin, "Mining Frequent Patterns without Candidate Generation," Proceedings of ACM SIGMOD International Conference on Management of Data, ACM Press, Dallas, Texas, pp. 1-12, May 2000.
[7] J. Pei, J. Han, H. Lu, S. Nishio, S. Tang, and D. Yang, "Hmine: Hyper-Structure Mining of Frequent Patterns in Large Databases," Proceedings of IEEE International Conference on Data Mining, pp. 441-448, 2001.
[8] Pietracaprina, and D. Zandolin, "Mining Frequent Item sets Using Patricia Tries," FIMI '03, Frequent Itemset Mining Implementations, Proceedings of the ICDM 2003 Workshop on Frequent Item set Mining Implementations, Melbourne, Florida, Dec. 2003.