ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

ASSOCIATION RULE MINING FOR IDENTIFYING OPTIMAL CUSTOMERS USING MAA ALGORITHM

[1]G. BABU , [2]DR. T. BHUVANESWARI
[1,2]Assistant Professor, Department of Computer Applications
[1]Adhiparasakthi Engineering College,Melmaruvathur,Tamilnadu..
[2]L.N.Government College, Ponneri- 601204.

ABSTRACT

Identifying customers which are more likely potential for a product and service offering is an important issue. In customers identification data mining has been used extensively to predict potential customers for a product and service. Modern companies and organizations efficiently implement a CRM strategy for managing a company interactions and relationships with customers. CRM systems have been developed and designed to support the areas of marketing, service process and sales. Many literature studies are available to preserve the customer relationship but small drawbacks occur in the existing methods. One method to maintain the customer relationship is frequency based method i.e., The Company will give declination to the customer based on the historical data that is the customers how many times come to that company. These methods are not effective. Because the customers give revenue to that company is less. So the company revenue is affected. In the data mining field, association rules have been researched for more than ten years ; however, the degree to which the support threshold effectively discovers interesting association rules has received little attention. Most of the research effort in the scope of association rules has been oriented to simplify the rule set and to improve performance of the algorithm. With the recent advancement of Internet and Web Technology, web search has taken an important role in the ordinary life. To discover interesting patterns or relationship between data in large database association rule mining is used. Association rule mining can be an important data analysis method to discover associate rules in CRM. The Apriori algorithm is a proficient algorithm for determining all frequent customers in CRM. But these are not the only problems that can be found and when rules are generated and applied in different domains. Troubleshooting for them should also take into consideration the purpose of association model and the data they come from. Some of drawbacks like non interesting rules, low algorithm performance arts are found in the algorithm. Several past studies addressed the problem of mining association rules with different Supports will not be appropriate in large dataset and they cannot generate more useful rules. This paper suggests a new framework of algorithm MAA that overcomes the limitations associated with existing methods and enables the finding of association rules based on Apriori Algorithm among the presence and/or absence of a set of items without a preset minimum support threshold and Minimizing Candidate Generation. The proposed work is an efficient algorithm for generating frequent itemsets and is optimized to takes less time compare to the existing algorithms. The main aim of this algorithm is to reduce execution time and memory utilization as compared to the existing algorithms. The framework has been tested on several datasets.. The result obtained shows that the proposed algorithm takes 25% less time compared to the Apriori algorithm in all instances. The performance of the algorithm is influenced by the dimensions of the data set and support factor and it is compared with performance of FP-growth and DynFP-Growth algorithms. The algorithm used to discover coherent rules which has been used in CRM model .Then, the mined information is used to calculate the company profit and frequency (the number of times the particular customer visit the company). By using association rule mining, the profit and frequency value of each customer is computed. Based on the mining result, the companies provide offers to customer using swarm intelligence technique known as particle swarm optimization.. This offer does not affect the company revenues as well as satisfying the customers. This process will make a good relationship between the customers and organizations and to satisfy the customers forever with company's rules.

Keywords: *CRM, PSO, Data Mining, Association Rule Mining, Algorithm, DynFP-Growth*

Journal of Theoretical and Applied Information Technology
31st August 2014. Vol. 66 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

I. INTRODUCTION

This research is concerned with study and analyzes the data mining technique and particle swarm optimization in order to improve the impacts of the effect are business data processing and scientific computing. During the initial years of the development of computer techniques for business, computer professional were concerned with designing files to store the data so that information could be efficiently retrieved. The advent of computer technology has significantly influenced our lives and two major impacts of this effect are Business Data Processing and Scientific Computing. During the initial years of the development of computer technology for business, computer professionals were concerned with designing files to store the data to store the data so that information could be efficiently retrieved. There were restrictions on storage size for storing data and on the speed of accessing the data. Needlessly to say the activity was restricted to a very few, highly qualified professionals. Then came an era when the task was simplified by a DBMS. The responsibilities of intricate tasks, such as declarative aspects of the programs were passed on to the data base administrator and user could pose his query in simpler languages such as query languages. Thus, almost any business – small, medium or large scale - began using computers for day-to-day activities. Every organization is now accumulating a large volume of daily transaction data and storing them as archival files. As a result, masses and masses of data – megabytes, gigabytes, and terabytes – are piling up in the electronic vaults of companies, governments and research institutions. A question that naturally arose is whether the enormous data that is generated and stored as archives can be used for improving the efficiency of business performance. Early 1990's, the answer to this was 'not much'. No one was really interested in utilizing data which was accumulated during the process of daily activities. Once the transaction process is over, these dada were dumped into archival files. Such a collection of data, whether their origin is business enterprise or scientific experiment, has recently spurred a tremendous interest in the areas of knowledge discovery and data mining. As a result, a new discipline in computer science, Data Mining, gradually evolved. Data is the exploration and analysis of large data set, in order to discover meaningful patterns and rules. The key idea is to find effective ways to combine the computer's power efficiency and effectiveness of customer relationship management system. The advent of computing technology has significantly influenced our lives and two major

to process data with the human eye's ability to detect patterns.

2. DATA MINING – AN OVERVIEW

2.1 Definition

"Data mining" is defined as a sophisticated data search capability that uses statistical algorithms to discover patterns and correlations in data [4]. The term is an analogy to gold or coal mining; data mining finds and extracts knowledge ("datanuggets") buried in corporate data warehouses, or information that visitors have dropped on a website, most of which can lead to improvements in the understanding and use of the data. The data mining approach is complementary to other data analysis techniques such as statistics, on-line analytical processing (OLAP), spreadsheets, and basic data access.

2.2. The evolution of data mining

Data mining techniques are the result of a long research and product development process. The origin of data mining lies with the first storage of data on computers, continues with improvements in data access, until today's technology allows users to navigate through data in real time. In the evolution of business data into useful information, each step builds on the previous ones. Table 1 shows the evolutionary stages from the perspective of the user. In the first stage, Data Collection, individual sites collected data used to make simple calculations such as summations or averages. Information generated in this step answered business questions related to figures derived from data collection sites, such as total revenue or average total revenue over a period of time. Specific application programs were created for collecting data and calculations.

The second step, Data Access, used databases to store data in a structured format. At this stage, company-wide policies for data collection and reporting of management information were established. Because every business unit conformed to specific requirements or formats, businesses could query the information system regarding branch sales during any specified time period. Once individual figures were known, questions that probed the

Journal of Theoretical and Applied Information Technology
31st August 2014. Vol. 66 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

performance of aggregated sites could be asked. For example, regional sales for a specified period could be calculated. Thanks to multi-dimensional databases, a business could obtain either a global view or drill down to a particular site for comparisons with its peers (Data Navigation). Finally, on-line analysis tools provided real-time feedback and information exchange with collaborating business units (Data Mining). This capability is useful when sales representatives or customer service persons need to retrieve customer information on-line and respond to questions on a real-time basis.

Information systems can query past data up to and including the current level of business. Often businesses need to make strategic decisions or implement new policies that better serve their customers. For example, grocery stores redesign their layout to promote more impulse purchasing. Telephone companies establish new price structures to entice customers into placing more calls. Both tasks require an understanding of past customer consumption behavior data in order to identify patterns for making those strategic decisions—and data mining is particularly suited to this purpose. With the application of advanced algorithms, data mining uncovers knowledge in a vast amount of data and points out possible relationships among the data. Data mining helps businesses address questions such as, "What is likely to happen to Boston unit sales next month, and why?" Each of the four stages were Revolutionary because they allowed new business questions to be answered accurately and quickly [4].

The core components of data mining technology have been developing for decades in research areas such as statistics, artificial intelligence, and machine learning. Today, these technologies are mature, and when coupled with relational database systems and a culture of data integration, they create a business environment that can capitalize on knowledge formerly buried within the systems.

3. ASSOCIATION RULE MINING

3.1. Definition

Association rules are one of the major techniques of data mining. The volume of data is increasing dramatically as the data generated by day-to-day activities. Therefore, mining association rules from massive amount of data in the database is interested for many industries which help in much business can decision making processes, such as cross marketing, Basket data analysis, and promotion assortment. It helps to find the association relationship among the large number of database items and its most typical application is to find the new useful rules in the sales transaction database, which reflects the customer purchasing behaviour patterns, such as the impact on the other goods after buying a certain kind of goods. These rules can be used in many fields, such as customer shopping analysis, customer relationship management system,storage planning and classifying the users according to the buying patterns, etc.

3.2 Basic Concept of Association Rules

The basic objective of finding association rules is to find all co-occurrence relationships called associations. Since it was first introduced in 1993 by Agrawal et. al, it has attracted a great deal of attention. Many efficient algorithms, extensions and applications have been reported. The classic application of association rule mining is market basket data analysis, which aims to discover how items purchased by customers in a supermarket (or store) are associated. Association rules are of form X -> Y, where X and Y are collection of items and intersection of X and Y is null. For example it may find that "95 percent of customers who bought bread (X) also bought milk (Y)" A rule may contain more than one item in the antecedent and consequent of the rule. Every rule must satisfy two users specified constrains: one is the measure of statistical significance called support and the other is a measure of goodness called confidence.

The problem of mining association rules can be stated as follows: Let $I = \{i_1, i_2, \ldots, i_m\}$ be a set of items. Let $T = (t_1, t_2, \ldots, t_n)$ be a set of transactions (the database), where each transaction t_i is a set of items such that t_i □ I. An association rule is an implication of the form, $X \rightarrow Y$, where $X \subset I$, $Y \subset I$, and $X \cap Y =$ □ X (or Y) is a set of items, called an itemset.A transaction t_i □ T is said to contain an itemset X if X is a subset of t_i (this also say that the itemset X covers t_i). The support count of X in T (denoted by X. count) is the number of transactions in T that contain X. The strength of a rule is measured by its support and confidence.

Journal of Theoretical and Applied Information Technology
31st August 2014. Vol. 66 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Support: The support of a rule, $X \rightarrow Y$, is the percentage of transactions in T that contains $X \cup Y$, and can be seen as an estimate of the probability, $Pr(X \cup Y)$. The rule support thus determines how frequent the rule is applicable in the transaction set T. Let n be the number of transactions in T. The support of the rule $X \rightarrow Y$ is computed as follows:

$$Support = (X \cup Y).Count / n$$

Support is a useful measure because if it is too low, the rule may just occur due to chance. Furthermore, in a business environment, a rule covering too few cases (or transactions) may not be useful because it does not make business sense to act on such a rule (not profitable).

Confidence: The confidence of a rule, $X \rightarrow Y$, is the percentage of transactions in T that contain X also contain Y. It can be seen as an estimate of the conditional probability, $Pr(Y \mid X)$. It is computed as follows:

$$confidence = (X \cup Y).count / X.count$$

Confidence thus determines the predictability of the rule. If the confidence of a rule is too low, one cannot reliably infer or predict Y from X. A rule with low predictability is of limited use.

3.2.1 Apriori Algorithm

A large number of association rule mining algorithms have been developed with different mining efficiencies. Any algorithm should find the same set of rules though their computational efficiencies and memory requirements may be different. The best known mining algorithm is Apriori algorithm. The Apriori algorithm works in two steps :

1. Generate all frequent itemsets: A frequent itemset is an itemset that has transaction support above minimum support.

2. Generate all confident association rules from frequent itemsets: A confident association rule is a rule with confidence above minimum confidence.

The key element that makes association rule mining practical is the minsup threshold. It is used to prune the search space and to limit the number of frequent itemsets and rules generated. However, using only a single minsup implicitly assumes that all items in the data are of the same nature and/or have similar frequencies in the database. This is often not the case in real-life applications. In many applications, some items appear very frequently in the data, while some other items rarely appear. If the frequencies of items vary a great deal, It will encounter two problems :

1. If the minsup is set too high, it will not find rules that involve infrequent items or rare items in the data.
2. In order to find rules that involve both frequent and rare items, this has to set the minsup very low.

There are number of algorithms used to generate association rules such as Apriori algorithm, Eclat algorithm, FP-growth algorithm.

Apriori is a classic algorithm for learning association rules. Apriori is designed to operate on databases containing transactions (for example, collections of items bought by customers, or details of a website frequentation). Other algorithms are designed for finding association rules in data having no transactions or having no timestamps. As is common in association rule mining, given a set of itemsets (for instance, sets of retail transactions, each listing individual items purchased), the algorithm attempts to find subsets which are common to at least a minimum number C of the itemsets. Apriori uses a "bottom up" approach, where frequent subsets are extended one item at a time (a step known as candidate generation), and groups of candidates are tested against the data. The algorithm terminates when no further successful extensions are found. The purpose of the Apriori Algorithm is to find associations between different sets of data. It is sometimes referred to as "Market Basket Analysis". Each set of data has a number of items and is called a transaction. The output of Apriori is set of rules that tell us how often items are contained in sets of data [57]. The algorithm is described in the following Pseudo Code

Ck : Candiate itemset of size k
Lk : frequent itemset of size k
L1 = {frequent items};
For (k=1; Lk !=∅; k++) do begin
Ck-1 = candiates generated from Lk;
For each transaction t in database do
Increment the count of all candiates in Ck+1 that are contained in t
Lk+1 = candiates in Ck+1 with min_support
End
Return ∪kLk

Journal of Theoretical and Applied Information Technology
31st August 2014. Vol. 66 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

4. RELATED WORK IN ARM

Wang explores the data generalization concept of data mining as a way to hide details Information, rather than discovering trends and patterns. Once the data are masked, standard data mining techniques can be applied without modification. This method used in this paper not only hides the sensitive data but also provides an appropriate mining result [32].

Jafari reveals that certain specific sensitive association rules are hidden by decreasing its support or confidence than the pre-defined minimum support and minimum confidence. To decrease the confidence of a rule X⇒Y, either increases the support of X, i.e., the left hand side of the rule, but not support of X∪Y, or decrease the support of the item set X ∪ Y. In the second case, if it only decreases the support , the right hand side of the rule, it would reduce the confidence faster than simply reducing the support of X (Y [32].

Homburg, C et. al.,say that customer satisfaction research has focused primarily on the “disconfirmation of expectations” model, which states that feelings of satisfaction arise when Consumers compare their perceptions of a product's performance to their expectations. As per the marketing concept, any product should be considered as a “total product” that includes the core benefits from the product as well as auxiliary dimensions associated with that product. The customer is satisfied when he/she feels that the product's performance is equal to or more than what was expected (confirmation). Hence, customers have higher retained wills to communicate with enterprises and to develop a future relationship. Customer satisfaction should therefore be understood as relationship specific and individualized [34].

The RFM analytic model is proposed by Hughes (1994), and it is a model that differentiates important customers from large data by three variables (attributes), i.e., interval of customer consumption, frequency and money amount. The detail definitions of RFM model are described as follows:

- Recently of the last purchase (R). R represents recently, which refers to the interval between the time that the latest consuming behavior happens and the present. The shorter the interval is, the bigger R is.
- The frequency of the purchases (F). F represents frequency, which refers to the number of transactions in a particular period, for example, two times in one year, two times of one quarter or two times in one month. The many the frequency is, the bigger F is.
- The monetary value of the purchases (M). M represents monetary, which refers to the consumption money amount in a particular period. The much the money is, the bigger M is [35].

Mohd H.N.M. Nasir et al. [10] have discussed the causes of failures of the CRM system. To rectify them, they have proposed a CRM prototype using Human Computer Interaction (HCI). For the purpose of capturing user's requirements, they have acquired and analyzed the background, current conditions and environmental interactions of a multi-national company. The analysis mainly intends to determine the relationship between the stages of patterns and internal-external influences. Interviews, naturalistic documentation and studying user documentation were also done to gather blended data. Using all these data, the prototype had been developed with the incorporation of User-Centered Design (UCD) approach, Hierarchical Task Analysis (HTA), metaphor and identification of users' behaviors and characteristics. The performance of technique was measured using usability.

Mon Fong has discussed about different data types and mining algorithms [28]. He says that all mining algorithms are not best suited for all data types. Selecting a mining algorithm not only depends on the pattern to be extracted but also on the data types of the data on which the mining algorithm has to be applied. Data is transformed into the required format. For data transformation ‘data type’s generalization process’ is used.

Yen Ting talks about the ontology driven data mining on a medical driven database[29]. The database contains information about patients undergoing treatment for chronic kidney disease. Ontology was used as an aid to provide facts about the attributes of the database and also used for controlling vocabulary and the attributes with more impact are selected manually. Association Rule mining is applied on these selected attributes and antecedents are selected. Then the antecedents are taken as new class variables and association rule is applied. The final output was good when compared to that of naive method of association rule mining.

Journal of Theoretical and Applied Information Technology
31st August 2014. Vol. 66 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

5. DESIGN AND DEVELOPMENT OF PROPOSED ALGORITHM (MAA)

Drawbacks of Apriori algorithm

- Large Number of in-frequent itemsets are generated which increase the space complexity.
- Too many database scans are required because large number of itemsets are generated.
- As the number of database scans are more the time complexity increases as the database increases.

Due to these drawbacks there is a necessity in making some modification in the Apriori algorithm in the proposed system. The modified algorithm is described in the following steps:

Input:

D, a information of transaction

Min_sup, the minimum support count threshold

- Within the initial iteration of the algorithmic rule, every item may be a member of the set of candidate 1-itemset C1. The algorithmic rule merely scans all the transaction to count the quantity of occurrences of every item.
- The set of frequent item sets, L1, is set by comparing the candidate count with minimum support count that contains candidate 1-itemsets satisfying minimum support.
- To come up with the set of frequent 2-itemsets, L2, the algorithmic rule generates a candidate set of a pair of-itemset and so the transactions in D are scanned and therefore the support count of every candidate item set in C2 is accumulated and so repetition of the step 2.
- Then D2 is set from L2.
- Generate C3 candidates from L2 and scan D2 for count of every candidate.
- At the tip of the pass, verify that of the candidate item sets are literally massive, and people become the seed for following pass.
- This method continues till no new massive item sets are found.

The modified Apriori algorithm reduces the number of database scans and the redundancy while generating subtests and verifying them in the database. This algorithm needs to scan the database only once and also does not require to find the candidate set when searching for frequent itemset.

6. EXPERIMENTAL EVALUATION

The frequent pattern mining algorithms were implemented and tested on several data sets. The platform's specifications used for this test are Core 2 Duo processor 2.00 GHZ, 2 GB memory and hard disk 80 GB , Windows 7. In order to obtain more realistic results a Microsoft SQL 2000 Server for Windows 7 was used and accessed through the standard ODBC interface. To study the performance and scalability of the algorithms generated data sets with 10,000 to 50,000 transactions, and support factors between 5% and 40% were used. Any transaction may contain more than one frequent itemset. The number of items in a transaction may vary, based on dimension of a frequent itemset. Also, the number of items in an itemset is variable. Taking into account these considerations, the generated data sets depend on the number of items in a transaction, number of items in a frequent itemset, etc. The necessary parameters to generate the test data sets are defined in Table 6.1:

Table 6.1 Parameters

|D|	Number of transactions
|T|	Average size of the transactions
|L|	Number of maximal potentially large itemsets
|N|	Number of items

The sample data set has been generated for a number of items N = 100 and a maximum number of frequent itemsets |L| = 3000. |T| was chosen to be 10. Some of the results of the comparison between the Apriori, MAA(Modified Apriori Algorithm), FP-growth and DynFP-growth algorithms for support factor of 5% and for different data sets are presented in Table 6.2

Journal of Theoretical and Applied Information Technology
31st August 2014. Vol. 66 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 6..2 The Results For Support Factor Of 5%

Transactions(K)	Execution Time in sec			
	Apriori	DynFP-	FP-growth	MAA
10	13.94	2.32	3.76	2.06
20	21.98	3.98	6.88	3.12
30	48.37	8.23	14.63	6.10
40	66.50	12.10	20.90	11.26
50	107.65	19.50	34.30	15.36
80	198.30	37.90	64.80	33.89
110	1471.40	55.00	95.50	53.25
150	3097.20	98.90	174.60	86.85
190	5320.60	152.70	273.60	148355
300	9904.80	284.00	526.70	274.50
400	17259.20	458.10	849.70	258.25
520	20262.60	610.20	[illegible]	

Table 6.2 shows that the execution time of the algorithms grows with the dimension of the data set. The best performance is obtained by the MAA algorithm. The figure 6.1 shows that the execution time for the MAA algorithm is constant for a certain data set when the support factor decreases from 40% to 5% while, at the same time, the execution time of the MAA increases dramatically. For a support factor of 30% or greater and a data set of 40,000 transactions, the modified algorithm has better performances than the Apriori algorithm, but for a support factor of 20% or less its performance decreases dramatically. Thus, for a support factor of 5% the execution time for the Apriori algorithm is three times longer than the execution time of the FP-growth algorithm and up to five times longer than DynFP-growth.

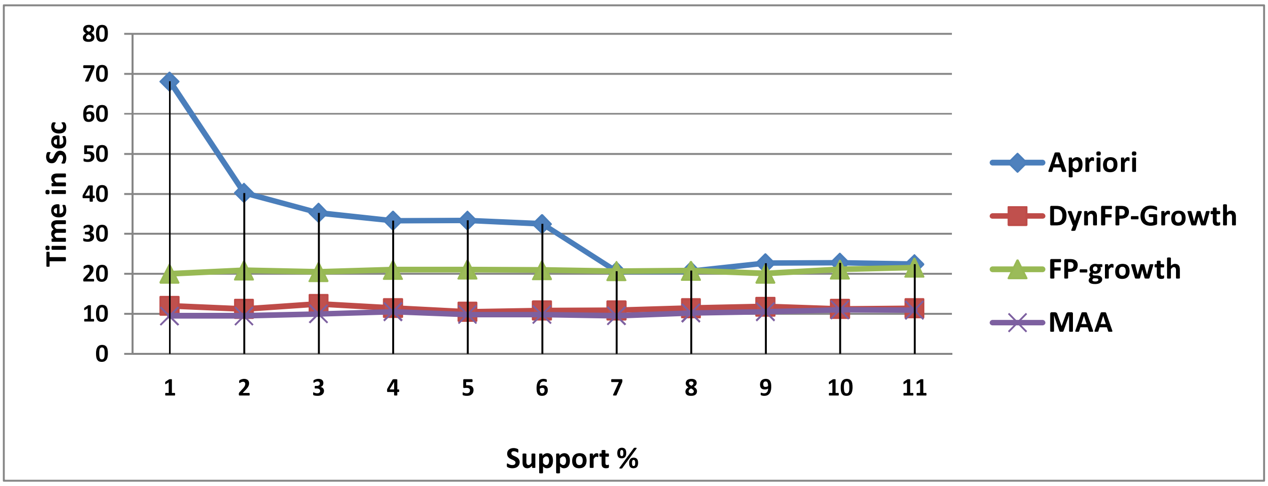


Figure 6.1. Scalability function of support for D1 40K database

Figure 6.1: The results for D1 150k by Support

Journal of Theoretical and Applied Information Technology
31st August 2014. Vol. 66 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

From the experimental data presented it can be concluded that the MAA algorithm is better than the DynFP-growth and FP-growth algorithm. First of all, the FP-growth algorithm needs at the most two scans of the database, while the number of database scans for the candidate generation algorithm (Apriori) increases with the dimension of the candidate itemsets. Also, the performance of the FP-growth algorithm is not influenced by the support factor, while the performance of the Apriori algorithm decreases with the support factor. Thus, the candidate generating algorithms (derived from Apriori) behave well only for small databases (max. 50,000 transactions) with a large support factor (at least 30%). In other cases the algorithms without candidate generation DynFP-growth, FP-growth and MAA algorithm behave much better. Several dataset have been used as test data to determine the performance and accuracy of the modified algorithm based on time and memory. The average results for both the execution time and the database pass yields 38% and 33% respectively in favor of the modified one. However, in some test data the outcome is in accordance with the original algorithm. It has been observed that as the number of items per transaction decreases the favorable result will be from the original algorithm since the pruning of candidate keys is closer to the first k+1 while implementing the modified one with the k(n) – 1 where n is the maxi-mum set size with set size frequency >= minimum support

6.1 Time Comparison

As a result of the experimental study, the performance of modified algorithm has been compared with the FP growth , DnFP growth and Apriori.The run time is the time to mine the frequent itemsets. The experimental result of time is shown in Figure 5.6 reveals that the proposed scheme outperforms the modified Apriori approach. The graph shows that in terms of execution time, the modified apriori executes less time compared to the other mining algorithms.. Moreover, in terms of database passes, the modified apriori provides less database access compared with the original one that makes its execution faster.

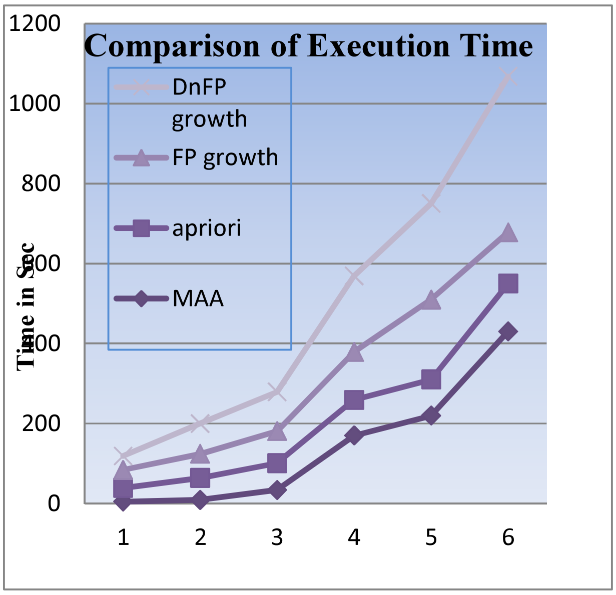
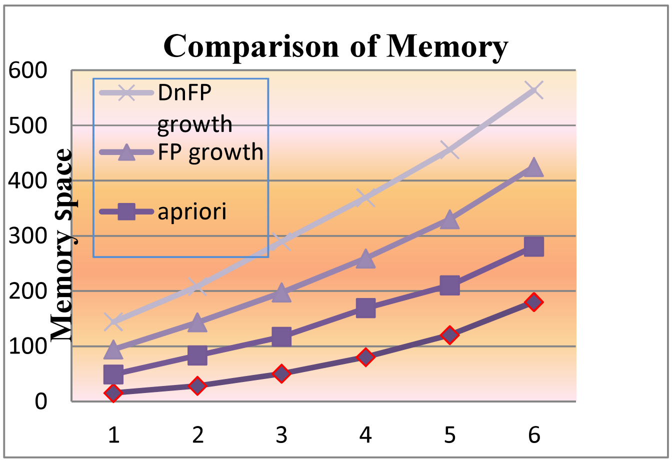


Figure 6.2: The Execution Time for Sample Dataset

6.2 Memory comparison

As it is clear from figure 5.7, the memory consumption for the modified algorithm has been producing high value at all level support because it produces candidate itemsets. The memory consumption for Dyn FP growth-at higher support levels is approximately compared to the new approach because as the support increase the probability of finding the maximal itemset whose repetition is greater than the minimum support is less thus its working become same as the FP-



ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Fig 6.3 . The Memory Usage At Various Support Levels On Sample Dataset

Several experiments have performed to evaluate the performance modified one against FP-growth,Dyn FP growth and Apriori, for generating the association rules. To perform the experiments different values of support were set because with different value of support the number of the frequent itemsets is different, and the running time and the memory consumptions are affected by the value of the support.From the experimental data presented it can be concluded that modified Apriori algorithm (MAA) takes less time for generating frequent item and is efficient than other algorithms and it speeds up the data mining process. Therefore it is said from the above results the modified Apriori algorithm performs better than the other association rule mining algorithms

7. CONCLUSION

In this paper we have discussed on modified apriori algorithms which provides a new technique in generating rules in data mining. This algorithms is more efficient than the traditional algorithm and provide faster results in terms of time Time and memory complexity. The above comparison clearly states that new modifications in the Apriori can improve the efficiency of the apriori. The main attribute that always will be in consideration is the number of database scans. As the number of transaction grows the size of the database increases due to which the number of scans increases. Many algorithms above have suggested a new technique which requires only one database scan. These methods can also further be modified to increase the efficiency of apriori. Also candidate set generation is another important aspect that should be more focused on. The itemsets generation step of apriori many times generates itemset which are not frequent and most of the time not required. Some algorithms have suggested to scan each transaction and combination of items should be based on items present in the transaction only. But this method will require keeping a track of items which have been traversed, and for doing this again we will require more number of comparison that leads to greater time complexity. These are some areas which can be focused and which would help in improving the association rule mining's apriori better and optimized algorithm .

REFERENCES

[1] , Wonyoung Kim and Ungmo Kim, “Mining Frequent Itemsets with Normalized Weight in Continuous Data Streams”, Journal of Information Processing Systems, Vol.6, No.1, March 2010.

[2] Geetha Mary and Sriman Narayana Iyengar, “A Frame Work for Ontological Privacy Preserved Mining”, International Journal of Network Security & Its Application , Vol.2, No.1, pp.9-20, January 2010.

[3] Kanimozhi Selvi Chenniangirivalsu Sadhasivam and Tamilarasi Angamuthu, "Mining Rare Itemset with Automated Support Thresholds", Journal of Computer Science, Vol. 7, No. 3, pp. 394-399, 2011.

[4] Aghaebrahimi, Zahiri and Amiri, “Data Mining Using Learning Automata”,World Academy of Science, Engineering and Technology, Vol. 49, No.60, pp.308-311, 2009.

[5] Mahesh, Mahesh T R and Vinayababu, “Using Data Mining Techniques for Detecting Terror-Related Activities on the Web”, Journal of Theoretical and Applied Information Technology, Vol.16, No.2, pp.99 -104, June 2010.

[6] Bakır, Batmaz, Gunturkun, Ipekçi, Koksal and Ozdemirel, “Defect Cause Modeling with Decision Tree and Regression Analysis”, World Academy of Science, Engineering and Technology, Vol.24, No.1, pp.1-4, 2006

[7] Choong Kwai Fatt and Edward Wong Sek Khin, "The Social-Technical View of Knowledge Management in Services Industries", Journal of Social Sciences, Vol. 6, No.2, pp. 256-264, 2010.

[8] Clement Sudhahar, Duraipandian Israel, Prabhu Britto and Selvam, "Service Loyalty Measurement Scale: A Reliability Assessment", Journal of American Applied Sciences, Vol. 3, No. 4, pp. 1814-1818, 2006.

[9] Michael Torggler, “The Functionality and Usage of CRM Systems”, International Journal of Human and Social Sciences, Vol.4:3, No.22, pp.164-172, 2009.

[10] Mohd H.N.M. Nasir , Mustaffa Kamal and Wan A.W. Rozali, "Human Computer Interaction Approach in Developing Customer Relationship Management", Journal of Computer Science, Vol. 4, No. 7, pp. 557-564, 2008.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[11] Ranjana Vyas, Lokesh Kumar Sharma and U.S. Tiwary, "Exploring Spatial ARM (Spatial Association Rule Mining) for Geo-Decision Support System", Journal of Computer Science, Vol. 3, No. 11, pp. 882-886, 2007.
[12] Thangavel , Qiang Shen and Pethalakshmi, “Application of Clustering for Feature Selection Based on Rough Set Theory Approach”, The International Journal of Artificial Intelligence and Machine Learning, Vol.6, No.1, pp.19-27, January 2006.
[13] Alizadeh, Ghazanfari and Fathian, “Learning FCM by Data Mining in a Purchase System”, International Journal of Industrial Engineering & Production Research, Vol.20, No. 1, pp.11-20, 2009.
[14] Perumal Pitchandi and Nedunchezhian Raju, "Improving the Performance of Multivariate Bernoulli Model based Documents Clustering Algorithms using Transformation Techniques", Journal of Computer Science, Vol. 7, No. 5, pp. 762-769, 2011
[15] Alizadeh, Ghazanfari and Fathian, “Learning FCM by Data Mining in a Purchase System”, International Journal of Industrial Engineering & Production Research, Vol.20,No. 1, pp.11-20, 2009
[16] Khalid Raza, “Application of Data Mining in Bioinformatics”, Indian Journal of Computer Science and Engineering, Vol.1, No. 2, pp. 114-118, 2010.
[17] Olga V. Krivobokova, “Evaluating Customer Satisfaction as an Aspect of Quality Management”, World Academy of Science, Engineering and Technology, Vol.53, No.92,pp.565-568, 2009.
[18] Saravanan Muthaiyah, "Malaysian Service Sectors Adopt FEDI to Boost Exports and Imports", American Journal of Applied Sciences, Vol. 1, No. 3, pp. 160- 167, 2004
[19] Cheng Fang Hsu, and Shinn-Jong Lin, “mCRM’s New Opportunities of Customer Satisfaction,” International Journal of Human and Social Sciences, Vol. 3:2, No.19,pp.133-137, 2008.
[20] Ching-Hsue Cheng and You-Shyang Chen, “Classifying the segmentation of customer value via RFM model and RS theory”, Expert Systems with Applications, Vol.36, No.3, pp.4176-4184, 2009 .
[21] Abdullah, Al-Mudimigh, Farrukh Saleem and Zahid Ullah, “Developing an Integrated Data Mining Environment in ERP-CRM Model – a Case Study of MADAR” , International Journal of Education and Information Technologies, Vol. 3, No.2, pp.135- 144, 2009.
[22] Yu-Teng Chang, “Applying Data Mining to Telecom Churn Management”, International Journal of Reviews in Computing, Vol.1, No.10, December 2009.
[23] Abdullah, Al-Mudimigh, Zahid Ullah and Farrukh Saleem, “A Framework of an Automated Data Mining Systems Using ERP Model”, International Journal of Computer and Electrical Engineering, Vol. 1, No. 5, pp.651-655, December 2009
[24] Alok Mishra and Deepti Mishra, “Customer Relationship Management: Implementation Process Perspective”, Acta Polytechnica Hungarica, Vol. 6, No. 4, pp.83-99, 2009