# The Transpose Technique to Reduce Number of Transactions of Apriori Algorithm

**Narinder Kumar[1], Anshu Sharma[2], Sarabjit Kaur[3]**
[1] Research Scholar, Dept. Of Computer Science & Engineering, CT Institute of Technology & Research, Jalandhar, Punjab 144008, India
[2&3] Assistant Professor, Dept. Of Computer Science & Engineering, CT Institute of Technology & Research, Jalandhar, Punjab 144008, India
Email.- bhagatnanu1990@gmail.com, asharma106@yahoo.com, r_sarabjitkaur35@rediffmail.com

**Abstract:** Data mining is one of the essentially used and interesting research areas. Mining association rule is one of the important research techniques in data mining field. Many algorithms for mining association rules are proposed on the basis of Apriori algorithm and improving the algorithm strategy but most of these algorithms not concentrate on the structure of database. The proposed technique includes transposition of database with further enhancement in this particular transposition technique. This approach will reduce the total scans over the database and then time consumed to generate the association rules will be less.

**Key Words:** Association rules, Apriori , Transpose, Execution time and Iterations.

## 1. INTRODUCTION:

Data Mining is known as the process of analyzing data to extract interesting patterns and knowledge. Data mining is used for analysis purpose to analyze different type of data by using available data mining tools. Association rule is based on discovering frequent item sets and frequently used by retail stores. The role of data mining (KDD) is very important in many of the field such as the analysis of market basket, classification, etc [1]. Nowadays, databases for the storage purpose are required to be increase in size as per need of technology. Data extraction from these databases is needed to be done in a specific manner for better use in future. Databases are of big size and data mining provides help to identify the particular information. Data mining is discovering the latest knowledge from the available sources [2].

There are various algorithms used for Association rules. Frequent pattern tree or FP-tree then based on this structure an FP-tree-based pattern fragment growth method was developed, FP-growth. Echivalence class clustering and bottom up lattice traversal is known as ECLAT algorithm. It uses vertical data layout [3]. Apriori algorithm is an advantageous algorithm presented to mine the items which are frequently purchased and also association rules are generated. The Apriori includes prior knowledge of items which are frequent. This is a step wise search in which by using the previous item sets, the next level item sets are found. Firstly, L1frequent -1 item sets after collecting the occurrence of items by scan the whole data is mined. Then accumulate all the items with minimum support [4]. This result is shown by L1. This frequent-1 item set required to generate the next level item sets until it reaches to null value. It needs one full scanning of dataset to find the result of Lk. Apriori property is available in algorithm to reduce the search [5]. Apriori algorithm can be implemented on various applications and discusses how effectively e-commerce application can be used with Apriori algorithm to help in business decisions by knowing the customer buying behavior analysis especially in the retail sector. The role of Apriori algorithm is also explained for finding the item-sets which are frequent and association rules generation is also included. The dataset includes analysis of the set of products purchased by the customer in a period of time is selected. Two main measurements are used in finding the frequent item-sets and strong rules of association. For all the transactions support is calculated which defines the association of dataset or item set [6].

ARM used to find out the interesting correlation between the items. In 1993 aggrawal proposed the association rule. After ARM many algorithm are generated that is apriori algorithm and improvement in apriori algorithm. Han and Fu change the minimum support threshold for association rule; the algorithm that is F-P algorithm there is no need of generating the candidate item in this algorithm. Some of these algorithms very slow to show the result in reasonable time [7].

## 2. LITERATURE REVIEW:

Ms. Rina Raval et.al [2013] explained survey of few better apriori algorithms are shown in this paper. From these surveys they find the good ideas. Survey included that many improvements are needed basically on pruning in Apriori to improve efficiency of algorithm. In the technique of Intersection and intersection Record filter is used with the record filter approach where to calculate the support and count the common transaction that contains in each elements of candidate set. In this technique only those transactions are considered that contain at least k items. Second last approach considers frequency and profit of items and generates association rules. Last technique suggests utilization of attributes such as profit, weight to associate with frequent item set for better information gain for user and business standpoint [8].

Jaishree Singh et.al [2013] described an Improved Apriori algorithm which minimize the scanning time by cutting down unusable transaction records as well as minimize the redundant generation of sub-items during pruning the candidate item sets, which can form directly the set of frequent item sets and eliminate candidate having a subset that is not frequent. The performance of Apriori algorithm is improved so that we can mine association information from massive data faster and better. Although this improved algorithm has optimized and efficient but it has overhead to manage the new database after every generation of $L_k$. So, there should be some techniques which have very less number of scans of database [9].

Ajay Kumar [2007] in this paper author proposed a new Parallel Partition Prime Multiple Algorithm for association rules mining. Proposed algorithm addresses the shortcoming of previously proposed parallel buddy prima algorithm. Therefore it minimizes the time and data complexity. The parallel Partition Prime Multiple algorithm can be improved to provide dynamic load balancing. With the increasing data it is important to create more efficient algorithms to extract knowledge from the data. The volume of data size is rising much faster than CPU execution speeds which have strong effects on the performance of software algorithms. The performance of Parallel Computing however cannot increase linearly as the number of the parallel nodes grows [10].

Yun Sing Koh et.al [2008] represented that rare association rule mining has received a great deal of attention in the recent past. In this research, they use transaction clustering as a pre-processing mechanism to generate rare association rules. The basic concept underlying transaction clustering stems from the concept of large items as defined by traditional association rule mining algorithms. They show that pre-processing the dataset by clustering will enable each cluster to express their own associations without interference or contamination from other sub groupings that have different patterns of relationships. The results show that the rare rules produced by each cluster are more informative than rules found from direct association rule mining on the unpartitioned dataset [11]. Aakansha Saxena, et.al [2014] explained that frequent pattern mining is one of the most important task for discovering useful meaningful patterns from large collection of data. Mining of association rules from frequent pattern from massive collection of data is of interest for many industries which can provide guidance in decision making processes such as cross marketing, market basket analysis, promotion assortment etc. The techniques of discovering association rule from data have traditionally focused on identifying relationship between items predicting some aspect of human behavior, usually buying behavior. Projected tree method is efficient in terms of speed but utilizes more space. In this paper, the study includes three classical frequent pattern mining methods that are Apriori, Eclat, FP growth and discusses some issues related with these algorithms [12].

S.Vijayarani et.al [2013] introduced that frequent pattern mining is the process of mining data in a set of items or some patterns from a large database. The resulted frequent set data supports the minimum support threshold. A frequent pattern is a pattern that occurs frequently in a dataset. Association rule mining is defined as to find out association rules that satisfy the predefined minimum support and confidence from a given data base. If an item set is said to be frequent, that item set supports the minimum support and confidence. In this paper, they have studied about how the éclat algorithm is used in data streams to find out the frequent item sets. Éclat algorithm need not required candidate generation [13].

## 3. APRIORI ALGORITHM:

The Apriori includes prior knowledge of items which are frequent. This is a step wise search in which by using the previous item sets, the next level item sets are found. Firstly, L1frequent -1 item sets after collecting the occurrence of items by scan the whole data is mined. Then accumulate all the items with minimum support. This result is shown by L1. This frequent-1 item set required to generate the next level item sets until it reaches to null value. It needs one full scanning of dataset to find the result of Lk. Apriori property is available in algorithm to reduce the search.

**Apriori algorithm having mainly two steps which are**:

1) Join step: The frequent item sets are joined step wise to find the candidates.

2) Prune step: When the items have support less as compared to the given minimum support those items are deleted under this step. Also, the item set having no sub set frequent is deleted.

Step to find the minimum support in Apriori algorithm:

Min support count= number of transaction * sup count.

For example: When the percentage values of support and confidence is given 60% and 60% respectively and number of transactions are 5 then

The result will be 5*60/100 = 3.

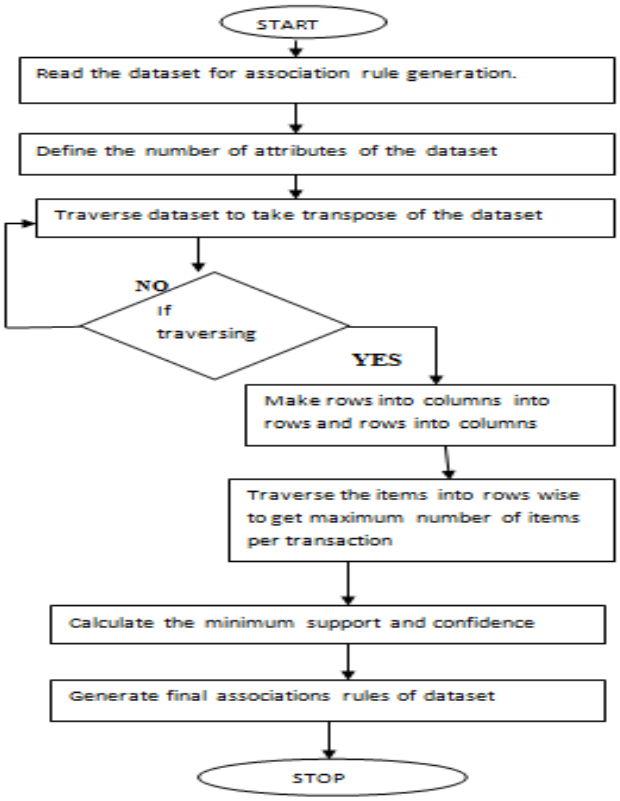
The performance of the classical apriori algorithm decreases because the algorithm requires multiple scans and when database size is very large, then it become very time taken process.

$$
\begin{aligned}
&\text{Apriori}(T,\epsilon)\\
&\quad L_1 \leftarrow \{\text{large } 1-\text{itemsets}\}\\
&\quad k \leftarrow 2\\
&\quad \text{while } L_{k-1} \neq \emptyset\\
&\qquad C_k \leftarrow \{a \cup \{b\} \mid a \in L_{k-1} \wedge b \notin a\} - \{c \mid \{s \mid s \subseteq c \wedge |s| = k-1\} \not\subseteq L_{k-1}\}\\
&\qquad \text{for transactions } t \in T\\
&\qquad\quad C_t \leftarrow \{c \mid c \in C_k \wedge c \subseteq t\}\\
&\qquad\quad \text{for candidates } c \in C_t\\
&\qquad\qquad count[c] \leftarrow count[c] + 1\\
&\qquad L_k \leftarrow \{c \mid c \in C_k \wedge count[c] \geq \epsilon\}\\
&\qquad k \leftarrow k+1\\
&\quad \text{return } \bigcup_k L_k
\end{aligned}
$$

**Fig 1: Pseudo code of apriori algorithm**

## 4. RESEARCH METHODOLOGY:

In this work, improvement is been proposed in traditional apriori algorithm to reduce number of transactions which directly reduction in processing time. The proposed technique is based on transpose technique of apriori algorithm In the transpose technique of apriori algorithm the dataset which is given as input will be transpose and instead to traversing each item in each transaction, algorithm will traverse transaction after every iteration. This directly reduce number of iterations because after each iteration algorithm will access transaction and in each transaction number items get traversed. As explained in flowchart 1, the dataset is given as input and transpose of input dataset is taken. The algorithm traverse rows and columns this traversing continue unless final association rules are created and the creation of association rules depends upon minimum support and confidence. When the final association rules are generated which is of transaction then the union is then of the items of the transaction to create final results. The proposed technique is implemented in MATLAB and it is been analyzed that number of iteration are reduced to 20 % as compared to traditional apriori algorithm




**Flowchart 1: Proposed technique**

## 5. EXPERIMENTAL RESULTS:

In this work, traditional apriori and improved apriori algorithm is implemented by taking market basket dataset. The interface is designed in MATLAB using drag and drop based GUI toolbox.

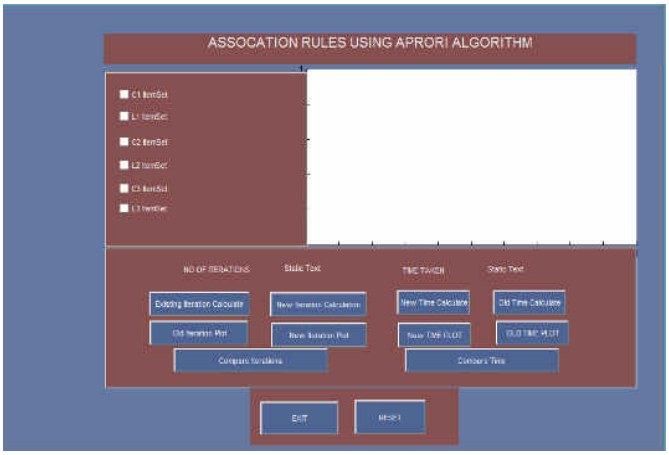



**Fig. 2: Interface of Implementation**

As illustrated in figure 2, the interface is shown which is designed to analyze performance of proposed and existing apriori algorithm. The performance is analyzed in terms of execution time and number of transactions

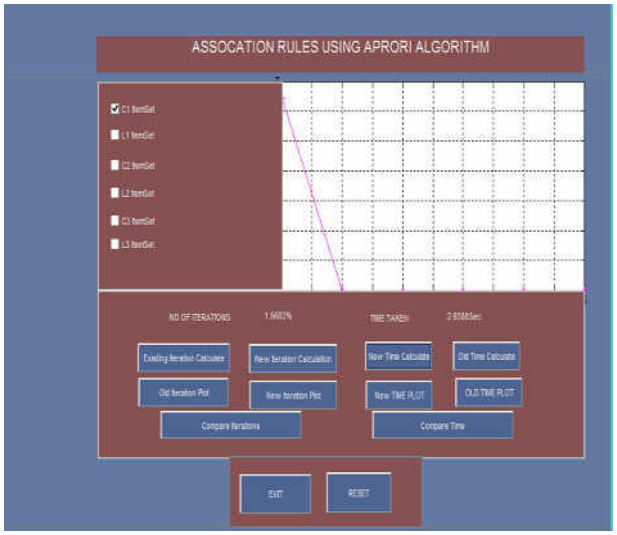



**Fig Fig 3: Execution of proposed algorithm**

As illustrated in figure 3, the proposed apriori algorithm is executed by taken market basket dataset. The proposed algorithm has 1.33 second execution time and 2 iteration for the generation of first level association rules .

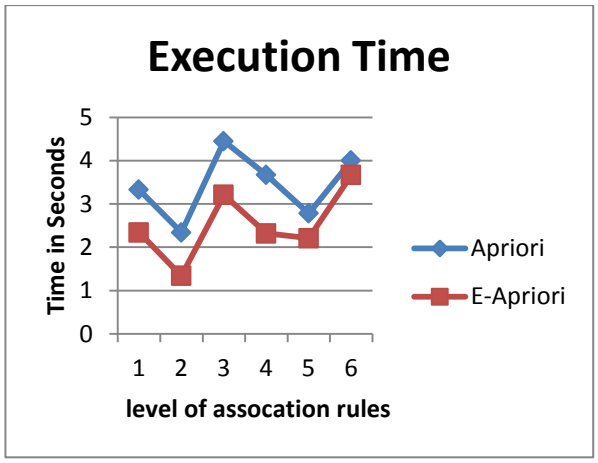



**Fig 4: Execution time comparison**

As shown in figure 4, the execution time of proposed and existing apriori algorithm is compared and it is been analyzed that execution time of proposed algorithm is less as compared to traditional apriori algorithm.

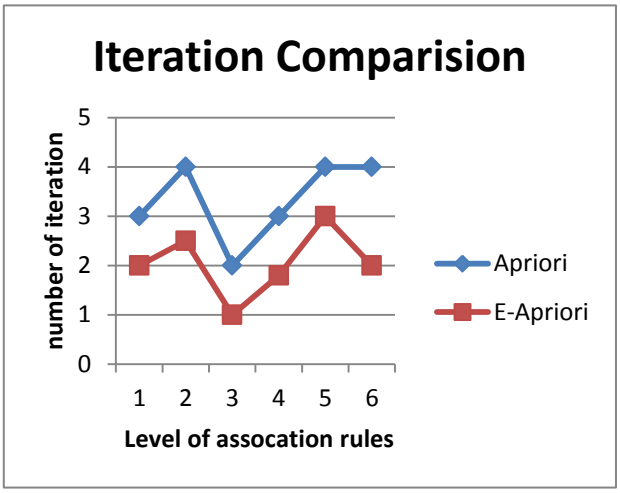



**Fig 5: Iteration Comparison**

As shown in figure 5, the comparison is made between traditional apriori algorithm and proposed apriori algorithm and it is been analyzed that number of iterations are less in proposed algorithm as compared to traditional algorithm.

## 6. CONCLUSION:

Association rule mining is one of the important techniques of research used in the data mining. For generating association rules, the Apriori algorithm is considered as the most efficient one. Apriori algorithm with a different technique is discussed which will acquire enhancement in Apriori algorithm with the transposition of database. Further improvement will be done in transposition technique using some different calculations of minimum support count. This approach will reduce the total scans over database as well as take less time to generate the association rules.

## ACKNOWLEDGMENT

I wish to express my appreciation and gratitude to the number of people who have helped me in gathering useful information for my work.

**REFERENCES:**


1. Dr. Kanwal Garg and Deepak Kumar, "Comparing the Performance of Frequent Pattern Mining Algorithms", International Journal of Computer Applications (0975 –8887) Volume 69–No. 25, May 2013.
2. Yun Sing Koh and Russel Pearsnd, "Rare Association Rule Mining via Transaction Clustering", Australian Computer Society, Inc. Glenelg, Australia. Conferences in Research and Practice in Information Technology (CRPIT), Vol.87, 2008.
3. Jiawei Han and Micheline Kamber, 2 Data Mining: Concepts and Techniques
4. Jaishree Singh, H. R. : "Improving Efficiency of Apriori Algorithm Using Transaction Reduction" International Journal of Scientific and Research Publications, (2013).
5. D. Gunaseelan, P. Uma : "An improved frequent pattern algorithm for mining association rules", International Journal of Information and Communication Technology Research Volume 2 No. 5, (2012)
6. Jugendra Dongre, Gend Lal Prajapati, S.V.Tokekar : "The role of apriori algorithm for finding the association rules in data mining". IEEE, (2014)
7. Rahul Mishra and Abhay, "Comparative Analysis of Apriori Algorithm and Frequent Pattern Algorithm for Frequent Pattern Mining in Web Log Data", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 3 (4), 2012, 4662 – 4665
8. Gagandeep Kaur and Shruti Aggarwal, "Performance Analysis of Association Rule Mining Algorithms", International Journal of Advanced Research inComputer Science and Software Engineering, Volume 3, Issue 8,August 2013
9. Daniel Hunyadi, "Performance comparison of Apriori and FP-Growth algorithms in generating association rules", Proceedings of the European Computing Conference, 2010.
10. Jaishree Singh1, H. R., "Improving Efficiency of Apriori Algorithm Using Transaction Reduction" International Journal of Scientific and Research Publications, Volume 4, issue 5, 2013
11. kumar, a.," Effect of Partition prime algorithm on Data Mining.", Research Scholar, Mewar University., 2007.
12. Aakansha Saxena, and Sohil Gadhiya, "A Survey on Frequent Pattern Mining Methods Apriori, Eclat, FP growth", Volume 2, Issue 1, ISSN: 2321-9939, 2014 IJEDR
13. S Vijayarani and P Sathya, "Mining Frequent Item Sets over Data Streams using Eclat Algorithm" IJCA Proceedings on International Conference on Research Trends in Computer Technologies 2013 ICRTCT(4):27-31, February 2013.