

# Enhancement in Apriori Algorithm using Transpose Technique to Improve Performance

Harkamal Kaur

Mr. Abhishek Tyagi

Dept. of Computer Science and Engineering,  
Lovely Professional University,  
Phagwara, Punjab

**Abstract**—Data mining process is the one of the most important and interesting research area. In the transaction database, 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 number of scans on the dataset and then less time will consumed to generate the association rules.

**Keywords**—Association rules, Support count, Apriori Algorithm, Transpose technique.

## I. INTRODUCTION

Data mining, popularly known as knowledge discovery in databases, it is the nontrivial extraction of implicit, previously unknown and potentially useful information from the data in databases [1]. It is the process of analyzing data to extract interesting patterns and knowledge. Data mining uses a variety of data analysis tools to discover patterns and relationships in data which is used to make predictions. Various data mining tasks are classified as [2]: exploratory data analysis, descriptive modeling and predictive modeling, discovering patterns and rules and retrieval by content. It is currently used in a wide range of applications like healthcare, market basket analysis, education system, manufacturing engineering, scientific discovery, e-commerce and decision making.

Association rule mining [12] is a data mining technique which is used to find the interesting association relationships among a large set of data items. The uncovered relationships can be represented in the form of association rules or set of frequent item-sets. Many algorithms have been proposed in association rule mining used to mine frequent item-sets. The algorithms vary mainly in generation of candidate item-sets and about the calculation of the supports for the candidate item-sets.

Number of algorithms [13] are available as: Compacting Data Sets (CDS) approach in which first duplicate transactions are being merged and then intersection between item-sets is done and unneeded subsets are deleted, DHP (Direct Hashing and Pruning) algorithm uses a hash technique for finding candidate item-sets and highly depends on the hash table size [10], and the Frequent Pattern Growth Algorithm adopts a

divide-and-conquer strategy. It first compresses the database representing frequent items into a frequent pattern tree and then divides the compressed database into a set of conditional databases, and Apriori algorithm [11] is an influential algorithm for mining frequent item-sets and for finding strong association rules. This algorithm uses prior knowledge and works on the principle that every non-empty subsets of a frequent item-set must itself be a frequent item-set. Apriori algorithm uses an iterative approach known as level-wise search, where k-item-sets are used to explore (k+1)-item-sets.

Apriori algorithm can be implemented on various applications and also reviews about finding the behavior of customers in retail sector using association rules. Apriori algorithm also finds the tendency of customers on the basis of frequently purchased item-sets [14].

## II. LITERATURE REVIEW

Chanchal Yadav et al. (2013) proposed that various approaches are used to overcome the drawbacks of the Apriori algorithm as to improve its efficiency. The proposed approach is presented which decreases pruning operation of candidate item-set. Data consistency is improved by resolving the problems like bad data or duplicate data and instead of finding the whole dataset, focusing on finding association in the filtered dataset. Other technique is also used to overcome the exceed limit of memory size by both frequent and infrequent item-sets by dividing dataset into horizontal partitions. The proposed idea reduces the size of each transaction and takes less time in comparison to Apriori algorithm to handle the data [3].

D. Gunaseelan, P. Uma (2012) for mining frequent pattern they proposed an improved algorithm using transposition of the database with minor modification of the Apriori-like algorithm. The main advantage of the proposed technique is the database stores in transposed form and in each iteration database is filtered and reduced by generating the transaction id for each pattern. The proposed technique reduces the storage space as well as the huge computing time. It has been presented the experimental results using synthetic data and also it has been compared results with the classical Apriori algorithm. Hence, the proposed technique is very beneficial for discovering patterns from large datasets [4].

Andre Bergmann proposed that knowledge discovery in databases has a great potential in data mining manufacturing. Useful patterns are extracted through various steps and future values are predicted on the basis of data mining regression and windowing technique. Time series model is created for forecasting which is of key importance in process and manufacturing engineering. Forecasting is performed to predict a machine failure and feature extraction is applied to compress the time series, keeping the important information while removing noise and correlations [5].

Mohammed AI-Maolegi, Bassam Arkok (2014) proposed another approach by reducing the number of transactions to be scanned for candidate item-sets generation which overcome wasting time for scanning the whole database as in Apriori algorithm. Before scanning all transaction records to count the support of each candidate, use before generated frequent item-set(Li) to get the transaction IDs of the minimum support count between X and Y items. Repeat the steps until no new frequent item-sets are identified. To generate the candidate support count in improved algorithm time consumed is less than as compared with the Apriori algorithm [6].

Zhuang Chen et al. (2011) analyses the shortcomings of Apriori algorithm and studies the current improvement strategies of Apriori. The improved algorithm named as BE-Apriori is presented which includes pruning optimization and transaction reduction strategies. Pruned optimization strategy have used temporary table to count the frequency of items all in the frequent item-sets and, transaction reduction strategy to compress the size of transaction and reduce the scale of database scanning. The improved algorithm has decreased the number of frequent item-sets generated and also reduced the running time. The experimental results about dataset retail have show the advantages of proposed algorithm of low system overhead, good operating performance and higher efficiency, as compared with pure Apriori [7].

Maza Dimitrijevic, Zita Bosnjak (2010) Comprehensive analysis of web usage association rules is conducted on a website of an educational institution and association rules are applied as a data mining technique to extract useful knowledge from web usage data. A set of basic pruning schemes are proposed to reduce the size of rule set and also non-interesting rules are removed. The analysis of association rules confirmed the hypothesis that it is not time consuming to discover interesting and useful association rules in web usage data. The pruned rule set is analyzed from the user's point of view and proposed actions that a webmaster may decide to take based on knowledge extraction from rules in order to enhance a web site and improve browsing experience of visitors [8].

Lamine M. Aouad et al. (2009) they present a new distributed approach and give a comparison of proposed approach with a classical Apriori-like distributed algorithm. In the proposed approach mining frequent item-sets on distributed datasets over the grid and only a local pruning strategy is considered. This reduces the communication and synchronization costs. It

is showed that in classical distributed schemes intermediate communication steps are computationally inefficient locally, then constraints the global performance. This approach greatly enhances the performance and achieves high scalability compared to the grid implementation of a distributed Apriori founded algorithm, namely the FDM (fast distributed mining) approach [9].

### III. CLASSICAL TECHNIQUE

Association rule mining is employed in many attractive application areas including engineering, marketing, medicine, and more. The base paper reviews about 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 frequent item-sets and generating the association rules.

The dataset of market basket 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 association rules which have support and confidence values respectively. The support for all the transactions is calculated which defines the association of dataset or item-set.

Apriori includes two main steps: join step includes generating candidates by joining among the frequent item-sets level-wise and prune step includes discarding item-sets if support is less than minimum threshold value and also discard item-sets containing infrequent subsets.

Procedure for finding the frequent item-sets in Apriori algorithm:

1. Search for all individual elements (1-element item-set) that have a minimum support of  $s$ .
2. Repeat:
  - a. From the results of previous search for  $i$ -element item-set, search for all  $i+1$  element item-sets that have a minimum support of item-set.
  - b. The set of all interesting frequent ( $i+1$ ) item-sets are obtained.
3. Until item-set size reaches maximum.

Procedure for defining the mining association rules using Apriori algorithm:

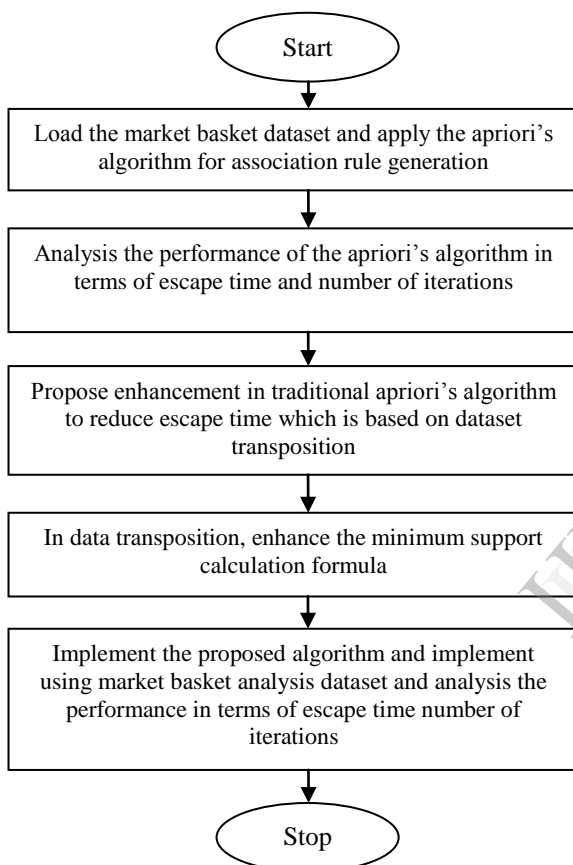
- Use Apriori to generate frequent item-sets of different sizes.
- At each iteration divide each frequent item-set  $x$  into two parts antecedent (LHS) and consequent (R.H.S) which represents a rule of the form  $LHS \rightarrow RHS$
- The confidence of such a rule is  $\text{support}(X) / \text{support}(LHS)$ .

All rules whose confidence is less than minimum confidence are discarded.

#### IV. PROPOSED TECHNIQUE

Proposed method will use the transposition of dataset and then find out the frequent item sets using Apriori algorithm steps. Reduce the escape time of Apriori's algorithm for association rule generation. To do that, this technique will further enhanced by modify the minimum support calculation formula of transpose technique. In this work, also define one more threshold value which is called "number of items in the transactional database". If in the particular transition the number of items below the threshold value then that transition will be deleted from the dataset.

##### A. Flowchart



The approach will reduce the number of scans on the dataset. If the number of scans reduced, less time is consumed to generate the association rules. The proposed idea will be implemented in MATLAB which is widely used in all areas of applied mathematics, in education and research at universities, and in the industry.

#### V. CONCLUSION

In the transaction database, mining association rule is one of the important research techniques in data mining field. For generating the association rules, Apriori algorithm is considered as the most efficient algorithm. In this paper, 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 number of scans over database and take less time to generate the association rules.

##### REFERENCES

- [1] N. Pandhy, Dr. P. Mishra, and R. Panigrahi, "The Survey of Data Mining Applications and Feature scope," International Journal of Computer Science, Engineering and Information Technology (IJCSSEIT), Vol.2, No.3, June 2012.
- [2] Introduction to Data Mining and Knowledge Discovery, Third Edition ISBN: 1-892095-02-5, Two Crows Corporation, 10500 Falls Road, Potomac, MD 20854 (U.S.A.), 1999.
- [3] Chanchal Yadav, Shuliang Wang, Manjot Kumar, "An Approach to Improve Apriori Algorithm Based On Association Rule Mining," 2011 IEEE.
- [4] D. Gunaseelan, P. Uma, "An improved frequent pattern algorithm for mining association rules," International Journal of Information and Communication Technology Research, 2012 Volume 2 No. 5.
- [5] Andre Bergmann, "Data Mining for Manufacturing: Preventive Maintenance, Failure Prediction, and Quality Control," Unpublished.
- [6] Mohammed Al-Maolegi, Bassam Arkok "An Improved Apriori Algorithm for Association Rules", International Journal on Natural Language Computing, 2014, Vol.3, No.1.
- [7] Zhuang Chen, Shibang Cai, Qiulin Song, Chonglai Zhu, "An Improved Apriori Algorithm Based on Pruning Optimization and Transaction Reduction," 2011 IEEE.
- [8] M. Dimitrijevic, and Z.Bosnjak, "Discovering interesting association rules in the web log usage data," Interdisciplinary Journal of Information, Knowledge, and Management, 5, 2010, pp.191-207.
- [9] Lamine M. Aouad, Nhien-An Le-Khac, Tahar M. Kechadi, "Performance study of distributed apriori-like frequent item-sets mining," Springer, Knowledge Information System, 2009.
- [10] "Using hash based Apriori algorithm to reduce the candidate 2-item-sets for mining," <http://www.jgrcs.info> Volume 2, No. 5, April 2011.
- [11] Apriori Algorithm, [http://www3.cs.stonybrook.edu/~cse634/lecture\\_notes/07a/priori.pdf](http://www3.cs.stonybrook.edu/~cse634/lecture_notes/07a/priori.pdf)
- [12] Association Analysis: Basic Concepts and Algorithms. <http://www-users.cs.umn.edu/~kumar/dmbook/ch6.pdf>
- [13] Shivangi Srivastava, Ganesh Khadanga, Divya Gupta "Mining Recurrent Pattern Identification on Large Database," International Journal on Computer Science and Engineering (IJCSSE), Vol. 6 No. 04 Apr 2014.
- [14] Jugendra Dongre, Gend Lal Prajapati, S.V.Tokekar, "The role of apriori algorithm for finding the association rules in data mining," 2014 IEEE.