# Re-Adapted Apriori Algorithm in E-Commerce Proposal Coordination

Arivazhagan [1], R. Pragaladan [2]

Research Scholar, Dept. of CS, Sri Vasavi College, Erode, India[1]

Assistant Professor, Dept. of CS, Sri Vasavi College, Erode, India[2]

**ABSTRACT**: Data mining is the process for generating frequent item sets that satisfy minimum support. Mining frequent item set is very fundamental part of association rule mining. Numbers of algorithms are used for generating frequent itemsets. The Apriori algorithm generates the frequent item sets, but the accuracy of that frequent set update is very less, there have been several methods proposed to improve its performance. Apriori algorithm consumes more time for scanning the database repeatedly. In this paper proposed a Re-adapted apriori algorithm to classify the users, improve accuracy, and to improve efficiency, which reduces a lot of time of scanning database and shortens the computation time of the algorithm. The accuracy of the frequent item set is very higher than priory algorithm.

**KEYWORDS:** data mining, apriori algorithm, modified apriori algorithm,re- adapted apriori algorithm.

## I. INTRODUCTION

In everyday life, information is collected almost everywhere. For example, at supermarket checkouts, information about customer purchases is recorded. When payback or discount cards are used, information about customer purchasing behavior and personal details can be linked. Evaluation of this information can help retailers devise more efficient and modified marketing strategies. The majority of the recognized organizations have accumulated masses of information from their customers for decades. With the e-commerce applications growing quickly, the organizations will have a vast quantity of data in months not in years. Data Mining, also called as Knowledge Discovery in Databases, is to determine the trends, patterns, correlations and anomalies in these databases that can assist to create precise future decisions. Physical analysis of these huge amount of information stored in modern databases is very difficult. Data mining provides tools to reveal unknown information in large databases which are already stored. A well-known data mining technique is Association Rule Mining. It is able to discover all the interesting relationships which are called as associations in a database. Association rules are very efficient in revealing all the interesting relationships in a relatively large database with a huge amount of data. The large quantity of information collected through the set of association rules can be used not only for illustrating the relationships in the database, but also for differentiating between different kinds of classes in a database. Association rule mining identifies the remarkable association or relationship between a large set of data items. With a huge quantity of data constantly being obtained and stored in databases, several industries are becoming concerned in mining association rules from their databases.

E-commerce recommendation system has been great development in the theory and practice. However, with the further expansion of e-commerce systems, e-commerce recommendation system is also facing a series of challenges. The major challenges facing e-commerce recommendation system, the key technology for e-commerce recommendation system recommended in the algorithm design and recommended system architecture useful to explore and study. The recommendation system is the user (user) is to provide users with the recommendation of the item (item). The user refers to users of the recommended system, that is, customers in e-commerce activities. The project is the recommended object is to provide products and services to our customers in e-commerce activities, which is the final recommendation system is returned to the users of the recommended content. In e-commerce activities, it is the number of users and the number of items. Recommended system to face the current user, called the target users or active users. The recommendation system work, it is according to certain algorithms, given the target users of the

recommended project. Association rule mining in large amounts of data to find interesting association or contact between the itemsets is an important topic in the research of KDD (Knowledge Discovery in Database). With the large amounts of data constantly collect and store a lot of people in the industry are increasingly interested in mining association rules from their databases. From a large number of business transaction records found interesting relationship that can help many business decision making, such as classification design, cross-shopping and cheap.

It is perhaps the most important model invented and extensively studied by databases and data mining community. Apriori utilizes a complete bottom up search with a horizontal layout and enumerate all frequent item sets. The proposed improved method of Apriori algorithm utilizes top down approach, where the rules are generated by avoiding generation of un-necessary patterns. The major advantage of this method is the number of database scans is greatly reduced, and produced high accuracy and classify the users by their usage.

## II.  LITERATURE SURVEY

### Novel  Method of Apriori Algorithm using Top Down Approach

Pooja Jain et.al, introduced an Association Rule mining is one of the important and most popular data mining techniques. It extracts interesting correlations, frequent patterns and associations among sets of items in the transaction databases or other data repositories. Apriori algorithm is an influential algorithm for mining frequent itemsets for Boolean association rules. Firstly, the concept of association rules is introduced and the classic algorithms of association rule are analyzed. In Apriori algorithm, most time is consumed for scanning the database repeatedly. Therefore, the methods are presented about improving the Apriori algorithm efficiency, which reduces a lot of time of scanning database and shortens the computation time of the algorithm.[1]

### Algorithm:Classical Apriori Algorithm Summary

Agrawal.R et.al, a proposed Apriori employs an iterative approach known as a levelwise search, where k-itemsets are used to explore (k+1)- itemsets. First, the set of frequent 1-itemsets is found. This set is denoted L1 .L1 is used to find L2 , the set of frequent 2- itemsets, which is used to find L3 , and so on, until no more frequent k-itemsets can be found. The finding of each Lk requires one full scan of the database. In order to find all the frequent itemsets, the algorithm adopted the recursive method.[10]

### Improved Version of Apriori Algorithm Using Top Down Approach

Mr.kailash Patidar et.al, introduced As with the advancement of the IT technologies, The amount of accumulated data is also increasing. It has resulted in large amount of data stored in databases, warehouses and other repositories. Thus the Data mining comes into picture to explore and analyze the databases to extract the interesting and previously unknown patterns and rules known as association rule mining. In data mining, Association rule mining becomes one of the important tasks of descriptive technique which can be defined as discovering meaningful patterns from large collection of data. Mining frequent item set is very fundamental part of association rule mining .As in retailer industry many transactional databases contain same set of transactions many times, to apply this thought, in this thesis present an improved Apriori algorithm that guarantee the better performance than classical Apriori algorithm.[2]

### Algorithm:Improved Apriori Algorithm:

Dao-I Lin et.al, Introduced the improved Apriori algorithm is usually used for association mining technique by using top down approach. The top down Apriori algorithms requirements to large frequent item sets and generates frequent candidate item sets. The improved Apriori algorithm which reduce unnecessary data base scan. This algorithm is useful for large amount of item set. Therefore, improved top down algorithm uses less space, less number of iteration.[4]

### An Efficient Data Mining Technique for Generating Frequent Item sets

Luo Fang et.al, proposed a Frequent item generation is a key approach in association rule mining. The Data mining is the process of generating frequent item sets that satisfy minimum support. Efficient algorithms to mine frequent patterns are crucial in data mining. Since the Apriori algorithm was proposed to generate the frequent item sets, there have been several methods proposed to improve its performance. But they do not satisfy the time constraint.

However, most still adopt its candidate set generation-and-test approach. In addition, many methods do not generate all frequent patterns, making them inadequate to derive association rules. The Enhance apriori algorithm has proposed in this paper requires less time in comparison to apriori algorithm. So the time is reducing.[8]

**Algorithm:Apriori Algorithm for Generating Frequent Item sets**

Different algorithms have been proposed for finding frequent item sets. The Apriori Algorithm is a well-known approach which is proposed by Agrawal & Srikant (1994). It is an iterative approach and there are two steps in each iteration. The first step generates a set of candidate item sets. Then, in the second step we count the occurrence of each candidate set in database and prune all disqualified candidates (i.e. all infrequent item sets). Apriori uses two pruning technique, first on the bases of support count (should be greater than user specified support threshold) and second for an item set to be frequent , all its subset should be in last frequent item set.[10]

## III. ALGORITHM

**Apriori Algorithm**

Apriori is a classic algorithm for learning association rules in data mining. Apriori is an influential algorithm for mining frequent itemsets for Boolean association rules. It is an iterative approach and there are two steps in each iteration. The first step generates a set of candidate item sets. Then, in the second step we count the occurrence of each candidate set in database and prune all disqualified candidates (i.e. all infrequent item sets). Apriori uses two pruning technique, first on the bases of support count (should be greater than user specified support threshold) and second for an item set to be frequent , all its subset should be in last frequent item set. The iterations begin with size 2 item sets and the size is incremented after each iteration. The algorithm is based on the closure property of frequent item sets: if a set of items is frequent, then all its proper subsets are also frequent.
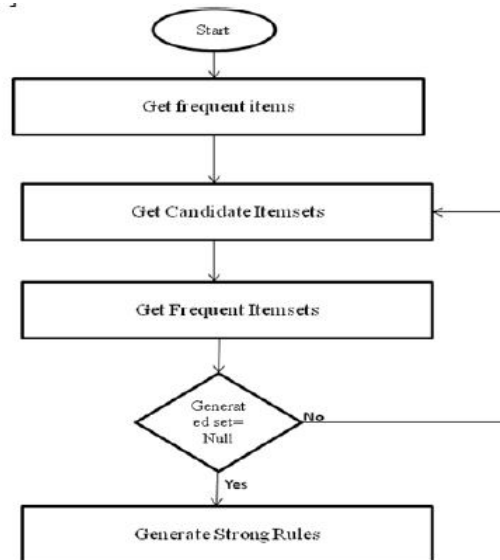


**Fig 1: Flow chart for apriori algorithm**

The Apriori Algorithm is an influential algorithm for mining frequent item-sets for Boolean association rules.
- Frequent Item-sets: The sets of item which has minimum support (denoted by Li for ith-Item-set).
- Apriori Property: Any subset of frequent item-set must be frequent.
- Join Operation: To find Lk, a set of candidate k-item-sets is generated by joining Lk-1 with itself.

Find the frequent item-sets: the sets of items that have minimum support– A subset of a frequent item-set must also be a frequent item-set i.e., if {AB} is a frequent item-set, both {A} and {B} should be a frequent item-set -Iteratively find frequent item-sets with cardinality from 1 to k (k-item-set) Use the frequent item-sets to generate association rules.

Join Step: Ck is generated by joining Lk-1with itself.  Prune Step: Any (k-1)-item-set that is not frequent cannot be a subset of a frequent k-item-set.

**Apriori Algorithm Pseudo-Code**

```
Variables:
C_k: Candidate item-set of size k
L_k: frequent item-set of size k
L_1= {frequent items}:
Process:
For (k = 1; L_k!=∅; k++) do begin
C_{k+1}= candidates generated from L_k;
        For each transaction t in database do
                Increment the count of all candidates in C_{k+1}
        Those are contained in t
        L_{k+1}= candidates in C_{k+1} with min_support
        End
        Return ∪_kL_k;
```

**Disadvantages of Apriori:**

1. Assumes transaction database is memory resident.
2. This algorithm is not efficient in large database.
3. This algorithm requires large number of dataset scans.
4. It only explains the presence and absence of an item in transactional databases.
5. In case of large dataset, Apriori algorithm produce large number of candidate itemsets. Algorithm scan database repeatedly for searching frequent itemsets, so more time and resources are required in large number of scans so it is inefficient in large data set.
6.

**Modified Apriori Algorithm**

The traditional Apriori algorithm is most frequently used by different researchers and groups to mine log data. This algorithm has some problem with their performance we observe that when the item set are increased then the time and memory required is increased exponent manner. To overcome this problem we propose a new Modified Apriori algorithm.

**Modified Apriori Algorithm Pseudo-Code**

```
Variables:
C_k: Candidate item-set of size k
L_k: frequent item-set of size k
L_1= {frequent items};
Process:
For (k = 1; L_k!=∅; k++) do begin
C_{k+1}= candidates generated from L_k;
        For each transaction t in database do
                If ( t== input set) then
                {
                        Increment the count of all candidates in C_{k+1}
                }
        Those are contained in t
        L_{k+1}= candidates in C_{k+1} with min_support
        End
        Return ∪_kL_k;
```

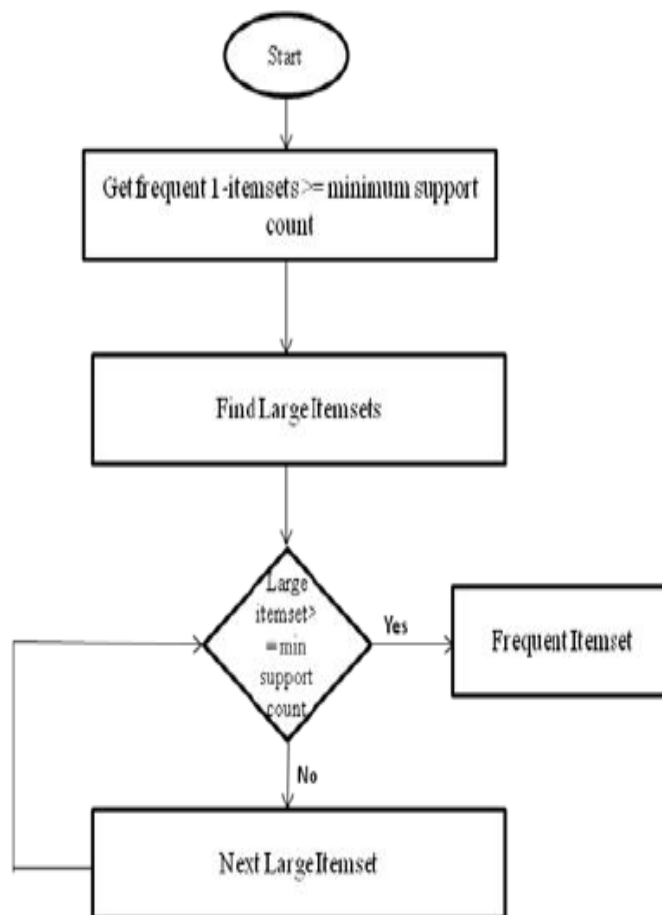**Re- Adapted Apriori Algorithm**

This algorithm reduces the number of scans while generating frequent itemsets. The main objective of this new approach is to build up new idea for generating frequent itemsets in transaction dataset. Top down approach is used for mining association rule. The top down Apriori algorithms uses large frequent item sets and generates frequent candidate item sets. The re-adapted Apriori algorithm reduces unnecessary data base scans. This algorithm is useful for large amount of item set. Re-adapted apriori algorithm uses less space and less number of iterations.

**Advantages of Re-Adapted Apriori Algorithm**
- Reduces unnecessary data base scans
- Useful for large amount of item set.
- Uses less space.
- Less number of iteration.
- Provides high accuracy.



**Flow chart for Re- Adapted Apriori Algorithm**

Input: Binary data matrix X of size p x q, K
Output: Frequent Itemsets and Association rules
//Binary data is transformed to real data using Wiener
transformation on a vector basis.
V = Call function wiener2 (Xi)
// Xi is a vector i of X
//Calculate K clusters (C1, C2, …CK) for V
C1, C2, …CK = Call function kmeans (V, K)
For each cluster Ci
       Cdn : Candidate itemset of size n
       Ln: frequent itemset of size n
       L1 = {frequent items};
       For (n=1; Ln != $\phi$ ; n++)
        Do begin
              Cdn+1 = candidates generated from Ln;
              For each transaction T in database do
Increment the count of all candidates in Cdn+1
                  which are contained in T
       Ln+1= candidates in Cdn+1 with min_support
         End
       UnLn are the frequent itemsets generated
       End
              End

**Re- Adapted Apriori Algorithm Pseudo-Code**
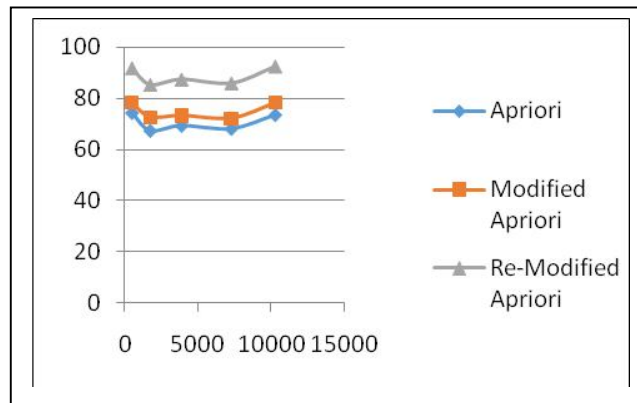


**Accuracy graph of frequent item set mining**

## IV.  ACCURACY ANALYSIS

The performance of algorithms are evaluated using N cross Validation method, based on this method accuracy is calculated using the total number of correctly classified objects verses the total sample produced to classify. The mathematical expression can be written as for calculating the performance in terms of accuracy as:

$$Accuracy = \frac{\text{total no of correctly classified instaces} * 100}{\text{total No of instances}}.$$

Accuracy of the re-adapted algorithm is very high rather than priori and modified priori algorithm by using the proposed method algorithm and accuracy equation.

## V. USER CLASSIFICATION

In the E-commerce recommendation system the user classified based on their website usage by using the re-adapted apriori algorithm.
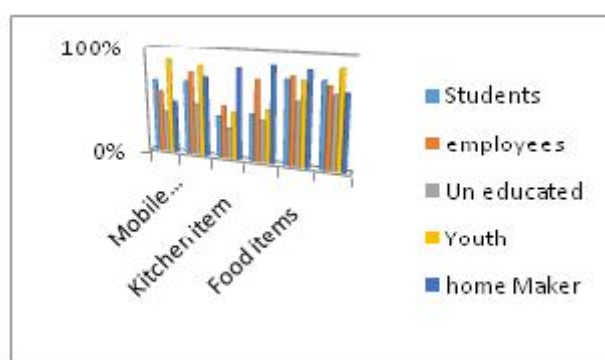


## VI. CONCLUSION AND FUTURE WORK

In this paper, the re-adapted Apriori algorithm is proposed to overcome the deficiency of the classical Apriori algorithm. The classical Apriori algorithm uses the bottom up approach. The new proposed method use the top down approach which reduces the number of database scans and it is useful for large amount of database scan. re-adapted apriori algorithm is efficient than classical apriori algorithm and reduces the time. After implementation of web mining system that is observed that the targeted goals are achieved and successfully omit results as expected from the selected models. In future work the same algorithms are utilized to work with other application for utilizing the properties of both kinds of algorithm. During study some modifications are also proposed in Apriori algorithm that is much efficient and effective with respect to the traditional Apriori algorithm, in future work that is required to enhance more for reducing the search time and building time in the proposed algorithm.

## REFERENCES

[1] Shikha Maheshwari Pooja Jain , Novel Method of Apriori Algorithm using Top Down Approach, International Journal of Computer Applications (0975 – 8887) Volume 77– No.10.
[2] Mr.kailash Patidar, Mr.Gajendra singh, Jatin Khalse, Improved Version of Apriori Algorithm Using Top Down Approach, IJSRD - International Journal for Scientific Research & Development| Vol. 2, Issue 10, 2014 | ISSN (online): 2321-0613.
[3] Geetha,Sk. Mohiddin , An Efficient Data Mining Technique for Generating Frequent Item sets, International Journal of Advanced Research in Computer Science and Software Engineering Volume 3, Issue 4, April 2013 ISSN: 2277 128X.
[4] Dao-I Lin, Fast Algorithms for Discovering the Maximum Frequent Set, New York University September 1998.
[5] Shikha Maheshwari Pooja Jain, The Research on Top Down Apriori Algorithm using Association Rule, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 4, April 2014 ISSN: 2277 128X.
[6] Ms Shweta, Dr. Kanwal Garg, Mining Efficient Association Rules Through Apriori Algorithm Using Attributes and Comparative Analysis of Various Association Rule Algorithms, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 6, June 2013 ISSN: 2277 128X..
[7] Langfang Lou, Qingxian Pan, Xiuqin Qiu , New Application of Association Rules in Teaching Evaluation System, International Conference on Computer and Information Application,2010, pp 13-16.
[8] Luo Fang, Qiu Qizhi ,The Study on the Application of Data Mining Based on Association Rules, International Conference on Communication Systems and Network Technologies 2012 pp.477-480.
[9] Dao-I Lin , Fast Algorithms for Discovering the Maximum Frequent Set.
[10] Agrawal.R and Srikant.R. , Fast algorithms for mining association rules. In Proc Int" Conf. Very Large Data Bases (VLDB), Sept. 1994, pages 487–499.

**BIOGRAPHY**

Mr. Arivazhagan A is Part time research scholar pursuing his M. Phil in the area of Computer Science in Sri Vasavi College, Erode affiliated by Bharathiar University. He has also completed his MCA Post Graduation in Computer Application in kongu Engineering College in Perundurai, Erode affiliated by Anna University. His Research area is Data Mining.

Mr.R. Pragaladan received his B.Sc and M.Sc degree in Computer Science from the Bharathidasan University in 2003 and 2006 respectively. He received his M.Phil degree in Computer Science from Bharathidasan University in the year 2007. He is currently pursuing Ph.D in Computer Science from Bharathiar University. He is presently working as an Assistant Professor, Department Of Computer Science, Sri Vasavi College, Erode – 638316. He has 6 years Teaching and Research Experience. His Research Areas are Cloud Computing, Network Security and Mobile Computing