ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

Re-Modified Apriori Algorithm in E-Commerce Recommendation System

R.Revathi [1], M.Geetha[2]

Research Scholar, Dept. of CS, Muthayammal College of Arts & Science, Rasipuram, Namakkal, India[1]

Associate Professor, Department of BCA, Muthayammal College of Arts & Science, Rasipuram, Namakkal, India[2]

ABSTRACT: Data mining is the process for generating frequent itemsets 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-modified 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, re- modified 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

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

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

This section includes the previous methods that are consumes the frequent item set mining algorithms for web log pattern extraction. In order to find information from the web access log files L.K. Joshila Grace et al provides a study according to his article, Log files contain information about User Name, IP Address, Time Stamp, Access Request, number of Bytes Transferred, Result Status, URL that Referred and User Agent. The log files are maintained by the web servers. By analysing these log files gives a neat idea about the user. This paper gives a detailed discussion about these log files, their formats, their creation, access procedures, their uses, various algorithms used and the additional parameters that can be used in the log files which in turn give way to an effective mining. It also provides the idea of creating an extended log file and learning the user behaviour. Web Usage Mining is the application of data mining techniques to discover interesting usage patterns from Web data, in order to understand and better serve the needs of Web-based applications. Usage data captures the identity or origin of Web users along with their browsing behaviour at a Web site. Web usage mining itself can be classified further depending on the kind of usage data considered. They are web server data, application server data and application level data. Web server data correspond to the user logs that are collected at Web server. Some of the typical data collected at a Web server include IP addresses, page references, and access time of the users and is the main input to the present Research.

One of the most well known and popular data mining techniques is the Association rules or frequent item sets mining algorithm. The algorithm was originally proposed by Agrawal et al. for market basket analysis. Because of its important applicability, many revised algorithms have been introduced since then, and Association rule mining is still a widely researched area. Agrawal et al. presented AIS algorithm in which generates candidate item sets on-the-fly during each pass of the database scan. Large item sets from previous pass are checked if they are present in the current transaction. Thus new item sets are formed by extending existing item sets. This algorithm turns out to be ineffective because it generates too many candidate item sets. Agrawal et. al. developed various versions of Apriori algorithm such as Apriori, AprioriTid, and AprioriHybrid. Apriori and AprioriTid generate item sets using the large item sets found in the previous pass, without considering the transactions. AprioriTid improves Apriori by using the database at the first pass. Counting in subsequent passes is done using encodings created in the first pass, which is much smaller than the database. A further scalability study of data mining was reported by introducing a light-weight data structure called Segment Support Map (SSM) with the purpose of reduces the number of candidate item sets required for counting. Han et. al.,introduced an algorithm known as FP-Tree algorithm for frequent pattern mining. It is another milestone in the development of association rule mining and avoids the candidate generation process with less passes over the database. FP-Tree algorithm breaks the bottlenecks of Apriori series algorithms but suffers with limitations.

III. EXISTING SYSTEM

The existing techniques have their own advantages and disadvantages. This section provides some of the drawbacks of the existing algorithms and the techniques to overcome those difficulties. Among the methods discussed for data mining, Apriori Algorithm is found to be better for association rule mining. Still there are various difficulties faced by Apriori Algorithm. They are,

- It scans the database a number of times. Every time additional choices will be created during the scanning process. This creates additional work for the database to search. Therefore, database must store a huge number

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

of data services. This results in lack of memory to store those additional data. Also, the I/O load is not sufficient and it takes a very long time for processing. This results in very low efficiency.

- Frequent item in the larger set length of the circumstances, leads to significant increase in computing time.
- Accuracy of the frequent item set is very low by using the apriori algorithm.

IV. PROPOSED SYSTEM

In this paper we propose re-modified apriori algorithm to improve the accuracy of the frequent data item and classify the user based on the usage of the web. There have been various researches in the discovery of association rules between the items of a transaction set. However, when defining user behavior patterns for services like a supermarket, an e-learning site, or simply any website, the analysis of the items composing their transactions should alone not be considered. For example, the behavior of a user in a website cannot be measured well by the items the user buys (what); but the way the user buys these items (how) should also be considered, in order to differentiate from other users or to group that user to other users with similar behavior. This leads to a step beyond the simple association rules discovery generated by the users of the service, that is to say, the relationship existing between the whole set of users and each one of the individuals has to be analyzed.

Advantages of proposed system:

It improving the accuracy of the frequent data item set 100% by using the re-modified apriori algorithm and classifies the users based on their data usage.

V. 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.
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.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

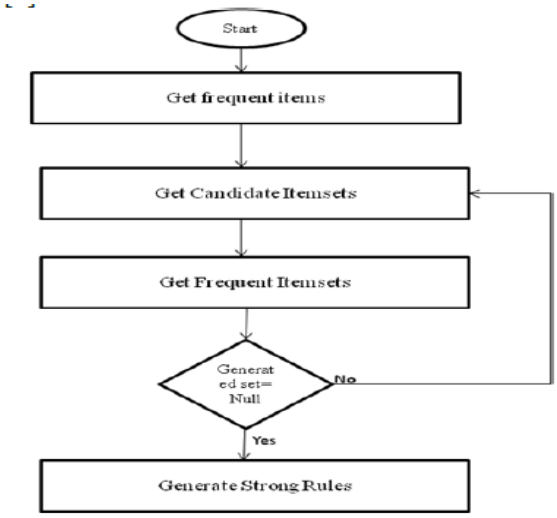


Fig 1: Flow chart for apriori algorithm

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.

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 $\cup_k L_k$;

Apriori Algorithm Pseudo-Code

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.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

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.

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.

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 $\cup_k L_k$;

Modified Apriori Algorithm Pseudo-Code

Re- Modified 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-modified Apriori algorithm reduces unnecessary data base scans. This algorithm is useful for large amount of item set. Re-modified apriori algorithm uses less space and less number of iterations.

Advantages of Re-modified Apriori Algorithm

- Reduces unnecessary data base scans
- Useful for large amount of item set.
- Uses less space.
- Less number of iteration.

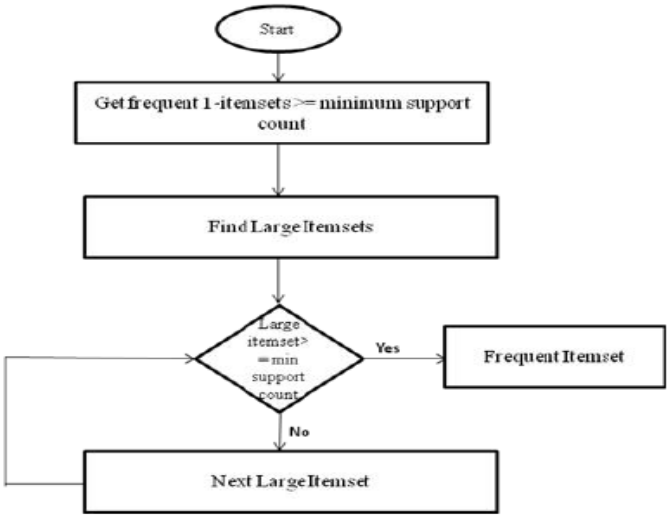
ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

- Provides high accuracy.



Flow chart for Re- Modified 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 != ϕ ; 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- Modified Apriori Algorithm Pseudo-Code

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

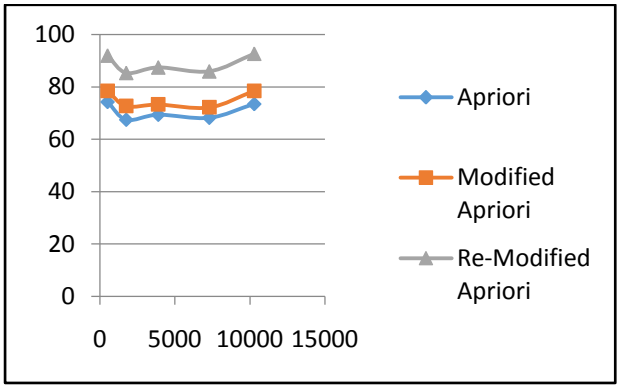
(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

VI. 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:

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

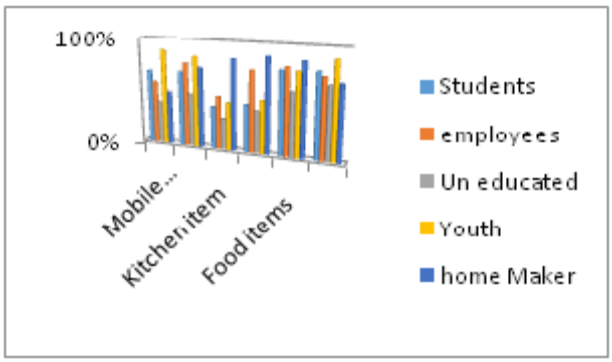


Accuracy graph of frequent item set mining

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

VII. USER CLASSIFICATION

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



ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

VIII. CONCLUSION AND FUTURE WORK

In this paper, the re-modified 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-modified 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"l Conf. Very Large Data Bases (VLDB), Sept. 1994, pages 487–499.

BIOGRAPHY

R.REVATHI. was born on 30-10-1989 in Tamilnadu, India. She received B.sc Computer Science degree in 2011 from Sengunthar college of Arts and Science, Namakkal, Affiliated to Periyar University, Salem, Tamilnadu, India. She received B.Ed,in 2012 from Gnanamani college of Education,Affiliated to Tamilnadu Teacher Educational University,Chennai, Tamilnadu,India. She received M.Sc Computer Science degree in 2014 from Sengunthar college of Arts and Science, Namakkal, Affiliated to Periyar University, Salem, Tamilnadu, India. She is Pursuing M.Phil (full time) degree from Muthayammal College of Arts & Science, in Periyar University Salem, Tamilnadu, India. Her interested research area is Datamining.

M.GEETHA. She received her B.sc Computer Science degree from Bharathidasan University and MS(IT)., degree from Bharathidasan University.She has completed her M.Phil at Annamalai University. She is having 10 years of experience in collegiate teaching and she is the Associate Professor, Department of BCA in Muthayammal College of Arts and Science, Rasipuram affiliated by Periyar University. Her main research interests include Datamining and Warehousing.