An Approach to Discover Frequent Itemset and Promising Itemset Using Association Rule Mining

Sejal P. Patel[1], Dr. S. M. Shah[2]

[1] ME Student, Department of Computer Science and Engineering, GEC Modasa, Gujarat, India
[2] HOD, Department of Computer Science and Engineering, GEC Modasa, Gujarat, India

ABSTRACT

Data mining is part of KDD (knowledge discovery process). It is used to discover interesting data from large amount of data. One of the major concept of data mining is association rule mining and frequent itemset mining. Proposed approach is used to discover frequent itemset and promising frequent itemset and keep only useful data in database. So size of database is reduced. Frequent itemset is itemset having frequently Occurrence in transaction database. Promising item set is item set became frequent in nearest future. By using promising itemset list avoid rescanning of original database. In dynamic database only scan new inserted or increment data in database. Based on this frequent item set generate rule used in decision making strategy in retail business or many application.

Keyword: - *Transaction database, Frequent Itemset, Promising Frequent Itemset, Association Rule mining*

1. INTRODUCTION

Following the growth of explosive growth of the amount of data gathered by transactional systems, a challenge for finding new techniques to extract useful patterns from such a huge amount of data arose. Data mining emerged as the new research area to meet this challenge. Data mining known as knowledge discovery of data base is an efficient way of extracting required knowledge from given or available large amount of dataset. It becomes very helpful in analyzing various patterns and various item set. One of the major applications of data mining is association rule mining technique. Association rule mining is a technique which finds correlation between two item sets.

1.1 Association Rule Mining

The association rule mining problem is to find out all the rules in the form of X => Y, where X and Y ⊂ I are sets of items, called itemsets.[2] The association rule discovery algorithm is usually decomposed into 2 major steps. The first step is find out all large itemsets that have support value exceed a minimum support threshold and the second steps is find out all the association rules that have value exceed a minimum confidence threshold.[2] We can discover rule using two parameter i.e. 1) support 2) confidence. An association rule is an implication of the form $X \Rightarrow Y$ (s, c), where X and Y are frequent itemsets in a transactional database and $X \cap Y = \emptyset$, s is the percentage of records that contain both X and Y in the database, called support of the rule, and c is the percentage of records containing X that also contain Y, called the confidence of the rule [2].

Association rule mining is to find all association rules the support and confidence of which are above or equal to a user-specified minimum support and confidence, respectively. Using support count we get frequent itemset. Frequent itemset are those itemset which are frequently appearing in dataset.

2. PROBLEM STUDY

2.1. Need of Frequent Itemset Mining

Efficient algorithms for mining frequent itemsets are crucial for mining association rules as well as for many other data mining tasks. The major challenge found in frequent pattern mining is a large number of result patterns [1]. For the static database traditional association rule mining can work but for dynamic database traditional association rule mining is a tradeoffs, for dynamic database when new transactions are inserted into the database.

This may introduce new association rules and some existing association rules would become invalid [3]. As a brute force approach, apriori may be reapplied to mining the whole dynamic database when the database has been changed. However, this approach is very costly even if small amount of new transactions is inserted into a database. Thus, the association rule mining for a dynamic database is an important problem.

In this paper, new incremental algorithm, called promising frequent itemset algorithm, is introduced. The goal of this work is to solve the efficient updating problem of association rules after a nontrivial number of new records have been added to a database. Our approach introduces a promising frequent itemset for an infrequent itemset that has capable of being a frequent itemset after a number of new records have been added to a database. This can reduce a number of times to scan an original database.

3. RELATED WORK

Ratchadaporn Amornchewin, Worapoj Kreesuradej Proposed a Promising frequent item set algorithm. Promising frequent item set algorithm is a new incremental algorithm to solve the efficient updating problem of association rules after a nontrivial number of new records have been added to a database. The maintenance of association rules for dynamic databases is an important problem in large data. [3]

Sunithavanamala, L.Padma Sree, S.Durga Bhavani Introduce the MSAppriori_VDB Algorithm. This new algorithm uses vertical database format that reduces the number database scans to one. This would lead to the generation rare rules efficiently. This algorithm uses multiple minimum supports which are calculated based on the frequency of occurrence. Hence the approach reduces the burden of assumption about minimum support threshold. The main issue of this algorithm is Large Size of Tid List. It is require more memory. [6]

Gurnee Kaur Introduce the Improve Apriori Algorithm for Improving the Efficiency of Apriori Algorithm in Data Mining This paper proposed an optimized method for Apriori algorithm which reduces the size of database along with reducing the number of candidate item sets generated. This improved algorithm is more efficient but it has overhead to manage the new database after every generation of Lk. [8]

Ms. Anju K.Kakkad, Ms. Anita Zala proposed Modified Approach of Promising Frequent Itemset Algorithm Based on Bucket Sort Approach in incremental Association Rule mining. This paper introduce new idea of incremental association rule mining which does not scan original database. Without scanning original database it will scan only incremented database. This algorithm with bucket sort approach give accurate result. This algorithm deal with update database. So, maintaining association rules for a dynamic database is an important issue. [5]

Anju Kakkad, Anita Zala Introduce the Promising Frequent Itemset Algorithm for Discover of Frequent Itemset and Promising Frequent Itemset Using Incremental Association Rule Mining Over Stream Data Mining. This approach Solve the efficient updating problem of association rules after a nontrivial number of new records have been added to a database. This algorithm has disadvantages With Space complexity and cannot work with decremented database. [4]

Kavitha J.K, Manjula D, Kasthuri Bha J.K Proposed an Efficient Incremental Rule Mining (EIRM) Algorithm. Which is Effective and Efficient Rule Mining Technique for Incremental Dataset. The algorithm also takes itemset count at each stage into consideration and split the frequent itemset as promising itemset and infrequent itemset as unpromising itemset. [7]

3.1. Proposed methodology

In an observation the itemset will be frequent itemset in updated database if it is member of large itemset in original database or incremental database. The main problem of incremental update is changing of frequent itemset that cause to re-execute from original database again. In this paper we present the new idea to avoid scanning the original database. Then we compute not only frequent itemset but also compute itemset that may be potentially large in an incremental database called "Promising frequent Itemset".

Proposed approach works for finding promising frequent itemset and frequent itemset. Focusing on promising frequent itemset extraction , proposed approach provide an even better approach for promising itemsets by dividing the support count twice. This will give chance to more items to be frequent as they can become part of promising frequent.

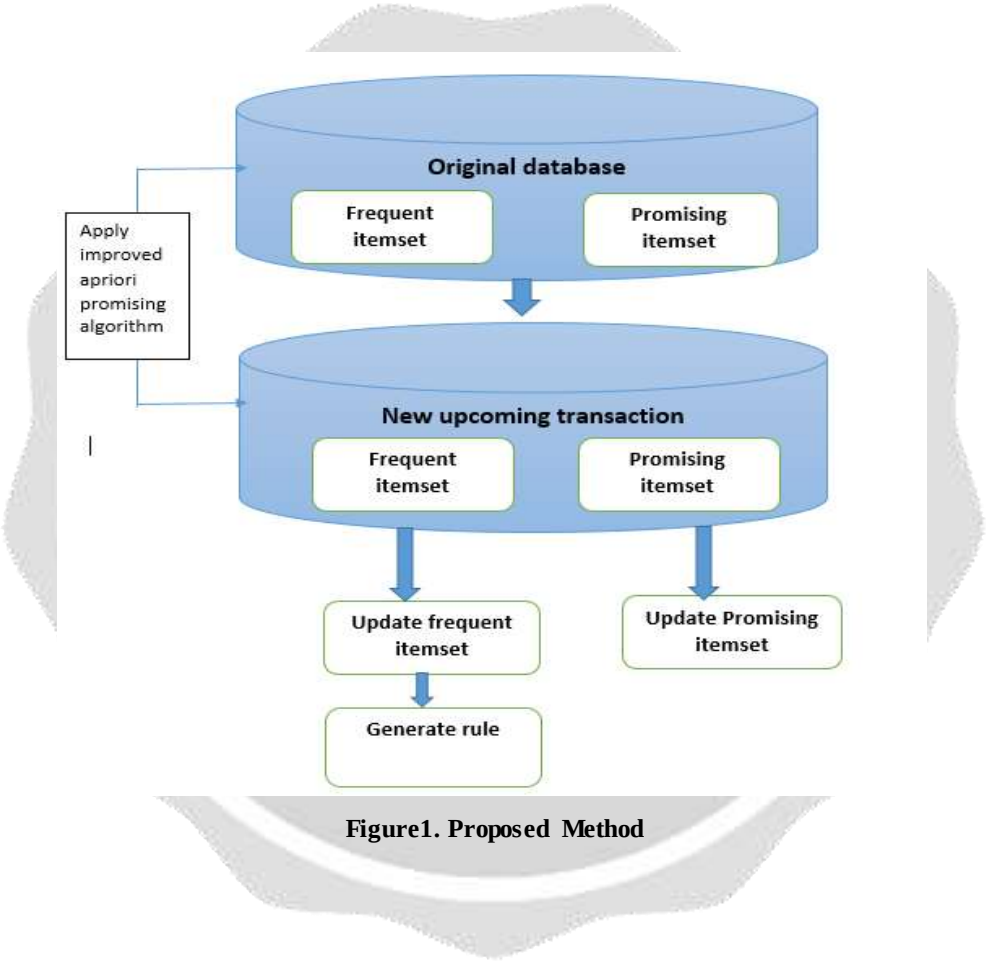


Figure1. Proposed Method

Algorithm Step:

Step 1: Take transaction database as input

Phase1

Step2: Give Minimum support count

Step3: Based on Support generate Frequent Itemset and Promising Frequent itemset

Step4: Frequent Itemset is

Itemset.Support >= min sup threshold

Step 5: Promising frequent itemset is

Itemset.Support >=promising itemset support<min sup threshold

Step 6: New Transaction Data Base Goto Phase 1

Phase2

Step7: Promising Frequent Itemset is find in Increment Database

min_supDB – (maxsupp/ (total size) * inc_size)

Step8: Compare Frequent Itemset and Promising itemset in original and new transaction database

Step9: Find which promising Item set is Become Frequent and Which Frequent Itemset is Become Promising Item set.

Step11: for finding frequent itemset from promising frequent itemset do Following:

For each itemset

{Promsing.itemset.support + upcoming same Itemset.Support >=min sup threshold.}

Move items to frequent itemset.

Step12: Update list of frequent itemset and Promising itemset

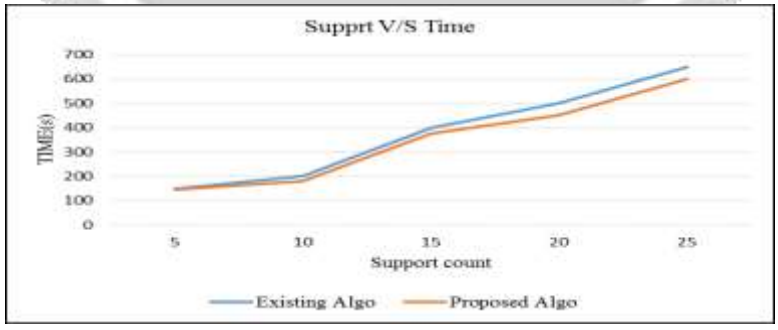
Step13: Using frequent itemset and confidence threshold Generate rule

3.2. Experiment Detail

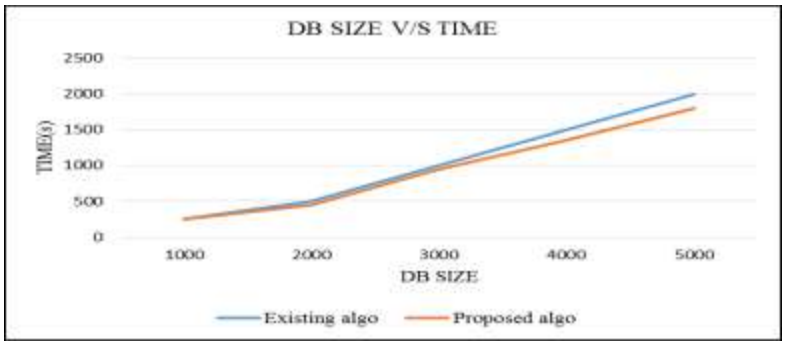
To evaluate the performance of promising frequent algorithm, the algorithm is implemented and tested on a PC With a core processor, and 1 GB main memory. The experiments are conducted on a synthetic dataset. The technique for generating the dataset is proposed by Agrawal and etc. The synthetic dataset comprises 1, 00,000 transactions has 10 items on average and the maximal size itemset is 4.

Proposed algorithm reduce time According to dynamic user Specified Support threshold Based on Support Count of an Itemset and size of data base is increase.

`



Comparison Based on Support value V/S Execution Time



Comparison Based on Database Size V/S Execution time

4. CONCLUSIONS

It can be concluded that the promising frequent itemset are as important as frequent itemset algorithm as they have potential to be frequent and with the help of scanning only that itemset having support count either greater or greater than lowest support count increases the speed of calculation and reduces the execution time in efficient manner. Enhancement in space complexity and decrement data size also in Future.

5. REFERENCES

[1]. Efficient Analysis of Frequent itemset Association Rule Mining Methods [M. Inbavalli, Dr. Tholkappia Arasu, International Journal of Scientific & Engineering Research, Volume 6, Issue 4, April-2015]
[2].A Survey on Frequent Itemset Mining with Association Rules [Endu Duneja, A.K. Sachan International Journal of Computer Applications (0975 – 8887) Volume 46– No.23, May 2012]
[3].Incremental Association Rule Mining Using Promising Frequent Itemset Algorithm [RatchadapornAmornchewin, WorapojKreesuradej, IEEE 2007]
[4].Discovery Of Frequent Itemset and Promising Frequent Itemset Using Incremental Association Rule Mining over Stream Data Mining. [Anju Kakkad, Anita Zala, Ijcsmc, Vol. 3, Issue. 5, May 2014, Pg.923 – 929]
[5]. Incremental Association Rule Mining by Modified Approach of Promising Frequent Itemset Algorithm Based on Bucket Sort Approach. [Ms. Anju K.Kakkad, Ms. Anita Zala, Ijarcce November 2013].
[6].Efficient Rare Association Rule Mining Algorithm. [Sunithavanamala, L.PadmaSree, S.DurgaBhavani, Ijera May-Jun 2013].
[7].Effective and Efficient Rule Mining Technique for Incremental Dataset. [Kavitha J.K, Manjula D, Kasthuri Bha J.K, JATIT, November 2013].
[8].Improving the Efficiency of Apriori Algorithm in Data Mining [Gurneet Kaur International Journal of Science, Engineering and Technology Volume 02 Issue 05 June- 2014]
[9]. IBM_Quest_Synthetic_Data_Generator
[10]. A Survey: On Various Incremental Association Rule Mining Approaches [Darshak I. Sojitra1, Dr. K. M. Patel 2, International Journal of Advance Engineering and Research Development Volume 2, Issue 1, January -2015]