An Enhancement of Association Rule Mining Algorithm

Gurpreet Batra, Alpana Vijay Rajoriya

Abstract- One of the well-researched and most important techniques of mining data is Association Rule Mining. Association Rules as the name itself indicates includes finding correlations among sets of items in transaction database. Most famous algorithm of association rule mining is Apriori is used for knowledge discovery. The proposed work is based on finding association rules considering the multidimensionality of the attributes and reducing the computation time that will increase the efficiency. Proposed work will improve the existing Apriori algorithm and will reduce some of the drawbacks of the existing algorithm.

Keywords: Association rules, confidence, support count, Apriori Algorithm.

I. INTRODUCTION

The process that finds beneficial patterns from huge data is known as data mining. With the development of information technology, the amount of databases as well as amount of data in different areas has been generated in extremely large amount. Increased research in databases as well as in information technology has led to the storage of this data and manipulating this valuable data for further decision making. Data mining is basically a process that extracts the valuable information and potential patterns from a huge amount of data. The process of data mining includes analysing data from different viewpoints and then summarizing into advantageous information that is used to raise revenues and to cut the costs. Data mining software is one of the analytical tools used for analysing data and allowing users the analysis of data from so many different angles, making categories of it and summarizing the identified relationships. One of the significant and well researched data mining technique is Association rule mining. Potential association rules can extracted from the data that seems to be unrelated in relational database and these rules are in IF THEN form. The main motive of the association rule mining is to find out frequent patterns in a dataset. The patterns that occurs frequently are known as frequent patterns. Example would be like "If someone buys a bottle of whisky then he is 82% likely to buy a bottle soda".

Manuscript Receive d on December 2014.

Er. Gurpreet Batra, M.Tech Department of (CSE), Lovely Professional University, Phagwara, Punjab, India.

Ms. Alpana Vijay Rajoriya, Asst. Prof., Department of Computer Science and Engineering, Lovely Professional University, Phagwara, Punjab, India.

Two parts of association rules: an antecedent, consequent. IF part is the antecedent. THEN part is consequent.

$$\text{Antecedent} \longrightarrow \text{Consequent}$$

$$\text{Support}(A \longrightarrow B) = P(A \text{ and } B)$$

$$\text{Confidence}(A \longrightarrow B) = P(A \mid B) = \frac{\text{Support}(A \longrightarrow B)}{\text{Support}(A)}$$

Association rules are made by analysing data for frequent if/then patterns and using the measures support and confidence to identify the most important relationships. How frequently the items appear in the database is indicated by support. The number of times the if/then statements have been found to be true is indicated by confidence. Apriori is the most classical and famous algorithm of mining frequent patterns. Prior knowledge of frequent itemset properties is used by the algorithm that's why the name of the algorithm is Apriori. It is designed for working on transactional databases. Apriori uses a categorical attributes and employs approach of "bottom up" extending one item at a time of frequent subset. When no more successful extensions are found then it leds to the termination of algorithm. Important role is played by these rules in areas like data analysis of shoping basket, store layout, product combinations and designing catalog.

II. RELATED WORK

Farah Hanna AL-Zawaidah et al. (2011)proposed Feature Based Association rule mining. It uses the features of the items and calculating the weight of candidate itemsets to generate association rules. The proposed algorithm involves transforming database means reorganizing of transaction database into a feature matrix. This helps in reducing the I/O accesses and also fastening the process of data mining. It also uses the leverage measure to reduce the number of candidate's itemsets and thus saving the memory from storing useless candidates.[1]. Jun Yang, Zhonghua Li et al (2013) Traditional Apriori algorithm gives equal importance to all the transactions in the database. This reduces the importance and accuracy of the association rules produced by the apriori algorithm. In order to generate more reasonable association rules different level of importance should be given to different items. In this paper, a improved algorithm i.e Feature based apriori is given which is based on the apriori algorithm. The improved algorithm uses features. It means every transaction is provided with its own features that carry more information. In classical apriori all the rules are mined. There was no facility of looking for particular one item. But, in feature based apriori, when association rules are mined then only transactions with same features are scanned and computed. Thus, providing association rules that are more reasonable. The improved

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

algorithm is analyzed under book recommendation system. The feature added to transactions is reader type. So, only transactions with same reader type are used to provide the strong association rules which are more reasonable. [2] Mohammed Al-Maolegi et al. (2014) presents an improvement on apriori . Here, not all the database is scanned again and again. Instead of this, when frequent 1-itemset is generated from the candidate set then those items along with there support count are listed which satisfies the minimum support threshold and also transaction id of all the items which are frequent is recorded. This eliminates the need of searching items in whole database and allows for searching them in the transactions whose ids are already specified. For producing next frequent 2- itemset , the frequent 1- itemset is joined with itself and then same process of recording transaction id and scanning in between those transaction is repeated for finding the all k frequent itemsets. Thus, this paper provides a way of reducing the scanning of large number of transcations in database by only scanning the transactions specified in transaction id.[3]

III. CLASSICAL TECHNIQUE

For Boolean association rules, Apriori is the most classical and famous algorithm of mining frequent patterns. Prior knowledge of frequent itemset properties is used by the algorithm that's why the name of the algorithm is Apriori. It is applied on transactional database. Each transaction has number of items contained in it forming a set of data. Output of the algorithm is sets of rules known as strong association rules. Algorithm employs apriori property stating that any subset of frequent itemsets must be frequent and this property helps in reducing search space. Iterative approach is employed by the algorithm in which k-itemsets are used for exploring (k+1)- itemsets. At first, set of frequent 1-itemsets i.e C1 is created by scanning database for counting occurences of each item, and then after it only those items in C1 satisfying minimum support threshold are collected. The resultant set so obtained is denoted as L1. This L1 is then used by joining it with itself to create the C2 candidate set of 2-itemsets. From C2 the itemsets satisfying support threshold are retained in L2. The process of generating candidate itemsets and L intemsets continues until no frequent k-itemsets can be found. Thus, it is divided into a two-step process which is used to search frequent item sets: join and prune actions.

Step 1: Join:

Joining Lk-1 with itself a set of candidate k-itemset is generated and denoted as Ck.

Step 2: Pruning:

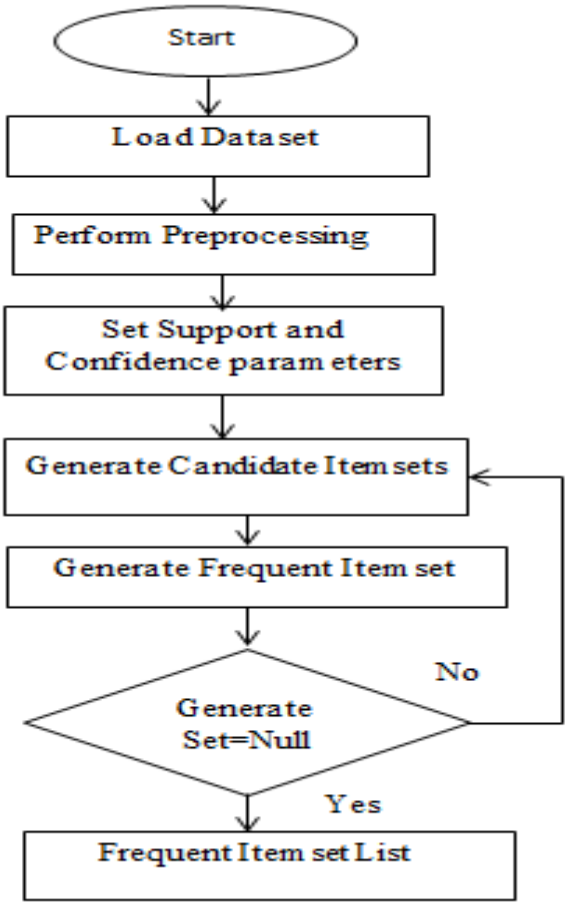
Ck which is a supset of Lk may contain members that may be frequent or not. But all of the frequent k-item sets are members of Ck. Apriori property is used here for reducing the size of Ck. A database scan leds to the determination of Lk by determining the count of each Ck candidate. Thus all candidates that have support count no less than the support threshold are in Lk.

IV. PROPOSED TECHNIQUE

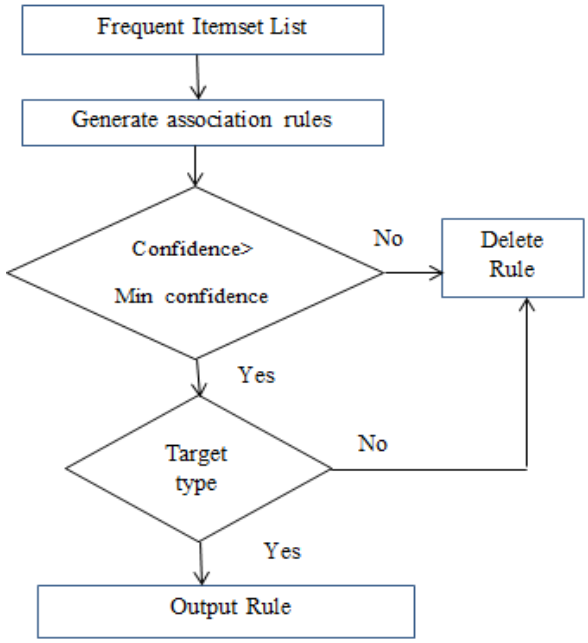
Proposed method using apriori algorithm will produce the association rules by also considering the multidimensionality of the attributes. Proposed method will reduce the number of iterations thus leading to reduced error rate. As the number of iterations will be reduced so computation time will also be reduced.

Flowchart:

- For Finding Frequent Itemsets:



- Generate Strong Rules:



In proposed technique we will also define the target type as a threshold. If the target type of generated rule will not match with the type of the rule generated then that rule will be discarded. Thus, the proposed method will also be very helpful in providing accurate more accurate rules. The proposed idea will be implemented in Weka and Netbeans softwares.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

V. CONCLUSION

Biggest issue in every domain of research is of data mining. Mining data with more and more accuracy and at the same time consuming as less processing time as possible is a very big task. Research that will be developed using rule induction along with association rule mining will be very advantageous in terms of accuracy and processing time. Using this, the number of rules will reduced and more data would be covered. With fast processing time error rate will be reduced from large dataset and time complexity will be reduced with the combine use of rule induction and association algorithm Apriori.

REFERENCES

[1] AL-Zawaidah, Farah Hanna, Yosef Hasan Jbara, and A. L. Marwan. "An Improved Algorithm for Mining Association Rules in Large Databases." World of Computer Science and Information Technology 1.7 (2011): 311-316.
[2] Yang, Jun, et al. "An Improved Apriori Algorithm Based on Features." Computational Intelligence and Security (CIS), 2013 9th International Conference on. IEEE, 2013.
[3] Al-Maolegi, Mohammed, and Bassam Arkok. "An Improved Apriori Algorithm for Association Rules." International Journal on Natural Language Computing (IJNLC), 3.1 (2014).
[4] Yadav Chanchal, Shuliang Wang, and Manoj Kumar. "Amity University Noida, UP, India." Computing, Communications and Networking Technologies (ICCCNT), 2013 Fourth International Conference on. IEEE, 2013.
[5] Moharana, U. C., and S. P. Sarmah. "Determination of optimal kit for spare parts using association rule mining." International Journal of System Assurance Engineering and Management Springer, Sweden (2014): 1-10.
[6] Lee, Dong Gyu, et al. "Discovering medical knowledge using association rule mining in young adults with acute myocardial infarction." Journal of medical systems 37.2 (2013): 1-10.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.