[VNSGU JOURNAL OF SCIENCE AND TECHNOLOGY] Vol.4. No. 1, July, 2015
153 - 156,ISSN : 0975-5446

Improved Apriori Algorithm using Bottom – up Approach

PRAJAPATI Priti H.
SRIMCA
Uka Tarsadia University
priti.prajapati@srimca.edu.in

SHETH Jikitsha R.
SRIMCA
Uka Tarsadia University
jikitsha.sheth@srimca.edu.in

Abstract

Basic purpose of data mining is to find knowledge from huge amount of data. Organization needs knowledge for decision making which will be beneficial for success. Data mining provides frequent patterns from given data. Process of decision making can be done very easily by analyzing frequent patterns. Apriori algorithm is basic algorithm to find frequent item sets from large amount of data. Other techniques are also available which can make Apriori algorithm more efficient. In this research paper, we have discussed one bottom up approach which improves basic Apriori algorithm. It provides frequent item sets in less time as compare to basic Apriori algorithm.

Keywords: Data mining, Association rule, Frequent itemset, Knowledge, Pattern

1. Introduction

Knowledge is collection of information which can useful to understand different situation. It is important to extract knowledge from large amount of data as organization want to take decision based on past experience. For that purpose, organization stores data in data warehouse. Data mining is the core process of "Knowledge Discovery in Database" which provides different interesting and hidden pattern from given dataset [1]. Frequent itemsets are set of items that frequently appear together in a transactional dataset. It is very difficult and time consuming process to find frequent itemsets from data of organization because they are in terabytes. Data mining integrate different techniques from many disciplines like information retrieval, machine learning, neural networks, database technology and statistics [4]. There are many area where data mining like Banking and Financial sectors, Retail Industry, Telecommunication industry, Health care, Scientific applications and many more.

2. Literature Review

Two basic algorithms are available to find frequent itemsets: Apriori Algorithm and FP – Growth Algorithm [1].

FP-growth is a representative pattern growth approach [1]. It is Depth First Approach and uses different data structure which is known as FP-tree. It finds frequent item sets without candidate itemset generation. It follows Divide-and-Conquer methodology. It is two steps approach.

1) Construct FP-tree using 2 passes over the data
2) Extract frequent itemsets directly from the FP – tree

In first pass, it scans data to find support for each item. Discard infrequent items and sort frequent items in decreasing order based on respective support. We can get FP-tree after completion of first pass. In second pass, it reads at a time one transaction and map it to a path. Based on that find frequent itemsets from FP-tree. Here, defined fixed order is used and maintain pointers between two nodes which contains same item.

Apriori follows candidate generation approach. It is Breadth First Search Algorithm. It is the most classical and important algorithm for mining frequent itemsets [1]. It uses prior knowledge of frequent itemset properties. It follows iterative approach where candidate set of previous iteration used to generate current result set. First, the set of frequent 1-itemsets is found by scanning the database to accumulate the count of each item, and collecting those items that satisfy minimum support. The resulting set is denoted L_1. Next, L_1 is used to find L_2 which contains 2-itemsets. Further L_2 is used to find L_3 and so on. The process is continue until no more frequent k-itemsets can be found. There are two major steps of Apriori algorithm which are join and prune. A set of candidate k-itemsets is generated by joining Lk-1 with itself which is denoted by C_k. If an itemset is not frequent, any large subset from it is also non-frequent; this condition prune from search space in database [3]. Here itemset is considered frequent itemsets if its support is greater than or equal to minimum support specified by user [4]. There are various versions of Apriori algorithm available like Apriori, AprioriTid, and AprioriHybrid. Apriori and AprioriTid are not considered transactions to find frequent itemsets [5]. Other techniques are also available which can improve efficiency of basic Apriori algorithm such as Hash-based technique, Transaction reduction, Partitioning, Sampling and Dynamic itemset counting.

3. Limitations of Apriori and FP – Growth Algorithms

Though Apriori algorithm is very basic and simple algorithm, it has some limitations. At each stage of process, it generates candidate sets and count occurrences of each itemsets. E.g. if there are 10^4 from frequent 1-itemsets, it need to generate more than 10^7 candidates into 2-length which in turn they will be verified and accumulate [1]. It costs much memory and also algorithm scans whole database at each stage to find occurrences which is time consuming process. Thus it is not suitable for large amount of data. FP- growth algorithm takes less time as compare to Apriori algorithm. In case of large amount of data, FP – tree may not fit in memory. It is also expensive to build such kind of tree because of its complexity.

This research paper proposed modified Apriori algorithm which can generate frequent itemsets using less memory in less time as compared to original algorithm.

4. Improved Apriori Algorithm

In this section, we discussed suggested bottom-up approach to improve basic Apriori algorithm.

Method:
Input: Transaction set *D* and Minimum support *min*
Output: L, List of frequent itemsets

1) Consider list of items, C, where C = {C_1, C_2, C_3,…..C_N }.
2) Prepare list of itemset which includes all unique items with occurrences S. Choose only that itemsets, L_1, where S >= min.
3) Reconstruct list of items C where C ∈ L_1.
4) Repeat above two steps until L_N = ϕ.

5. Analysis and Comparison of Result

We have checked basic and proposed Apriori algorithm for 10 datasets with 100, 200, 300, 600, 700, 1400, 1500, 3000, 6000, 12000 transactions respectively. Each datasets contains 6 to 9 unique items. For dataset which contains 100 transactions, execution time taken by original algorithm was 109 milliseconds where time taken by proposed algorithm was 78 milliseconds. Thus proposed Apriori reduce the time consumption by 28.44% from the original Apriori for first dataset and by 70.49% for dataset which contains 12000 transactions. Figure – 1shows result of experiment.

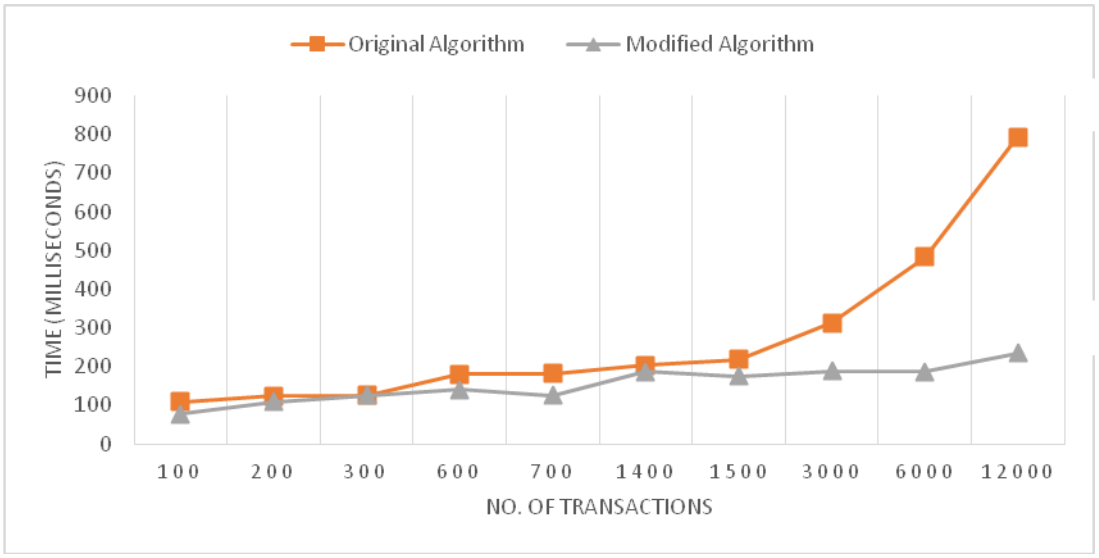

Figure 1 – Result of Original and Modified Apriori Algorithm based on No. of Transactions

Assume that for N transactions, t_1 is time taken by original algorithm, t_2 is time taken by proposed algorithm and $\Delta t = t_1 - t_2$. After observing result as available in figure – 1, we found that Δt is inversely proportional to number of transactions. Thus $\Delta t \propto \frac{1}{N}$ for the proposed algorithm.

6. Conclusion

In this research paper, we have discussed new bottom-up approach to improve basic Apriori algorithm which focuses on consumption of time to find frequent itemsets. This approach also reduce frequency of database scan at each stage. Here we found that execution time taken by basic algorithm is less than execution time taken by proposed algorithm. Thus it is very useful for large amount of data stored into data warehouse.

References

[1] S. Rao, P. Gupta, Implementing Improved Algorithm Over APRIORI Data Mining Association Rule Algorithm, International Journal of Computer Science And Technology, 2012, p.p. 489-493

[2] J. Singh, H. Ram, Dr. J. S. Sodhi, Improving Efficiency of Apriori Algorithm Using Transaction Reduction, International Journal of Scientific and Research Publications, 2013, p.p. 1-4

[3] H. Najadat, M. Al-Maolegi, B. Arkok, An Improved Apriori Algorithm for Association Rules, International Research Journal of Computer Science and Application, 2013, p.p. 1-8

[4] J. Han, M. Kamber, Data Mining: Concepts and Techniques, third ed., Morgan Kaufmann Publishers

[5] S. Maheswari, P. Jain, Novel Method of Apriori Algorithm using Top Down Approach, 2013, p.p. 18-21

[6] S. Aggarwal, R. Kaur, Comparative Study of Various Improved Versions of Apriori Algorithm, 2013, p.p. 687-690

[7] R. Raval, I. Rajput, V. Gupta, Survey on Several Improved Apriori Algorithms, IOSR Journal of Computer Engineering, 2013, 57-61

[8] J. Yabing, Research of an Improved Apriori Algorithm in Data Mining Association Rules, International Journal of Computer and Communication Engineering, 2013, p.p. 25-27

[9] C. Kaur, Association Rule Mining using Apriori Algorithm: A Survey, International Journal of Advanced Research in Computer Engineering & Technology, 2013, p.p. 2081-2084

[10] J. Jha, L. Ragha, Educational Data Mining using Improved Apriori Algorithm, International Journal of Information and Computation Technology, 2013, p.p. 411-418

[11] V. Mangla, C. Sarda, S. Madra, Improving the efficiency of Apriori Algorithm in Data Mining, International Journal of Engineering and Innovative Technology, 2013, p.p. 393-396

[12] S. Maheswari, P. Jain, The Research on Top Down Apriori Algorithm using Association Rule, 2014, 839-842