IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 1 Issue 6, August 2014.
www.ijiset.com
ISSN 2348 – 7968

A Comparative Analysis of Improved Version of Apriori Algorithm of Data Mining

Sana Irfan

M.tech Scholar, Deparment of Computer Science, Bharat Institute of Technology
Meerut, Uttar Pradesh, India

Abstract

Apriori algorithm in data mining is one of the most frequently used algorithm for finding the frequent itemset in database. Efficiency has been concerned for several years in data mining for finding frequent itemsets. Various improved version of Apriori algorithm are proposed. This paper presents an idea about the improved versions of Apriori algorithm and their performance comparison with Apriori algorithm.

Keywords*-Data Mining, Apriori Algorithm, Frequent Item-set generation, FP-Growth, HDO, APFT*

1. Introduction

Apriori is a seminal algorithm proposed by R.Agarwal and R.Srikant [1] [2]. Many research have been done to make mining frequent itemsets [4] efficient and scalable. The name of the Apriori based on the fact that the algorithm uses *prior Knowledge* to find the frequent itemset. It uses a property *'Any subset of a large itemset must be large'*Apriori algorithm works in two steps: join and prune step. In join step a set $L_k$is find and in prune step the itemset are compared with fixed minimum support and those that are less than the minimum support are deleted from L_k and a superset C_k is formed. Apriori algorithm suffers from several deficiencies: two many scans of database and generate a huge number of candidate sets. To overcome these deficiencies various techniques like hash-based technique, partitioning, transaction reduction, sampling, improved version of Apriori algorithm is proposed.

Table 1: Notation

L_k	Set of large k-itemsets
C_k	Set of candidate k-itemsets
$\overline{L_k}$	Set of Infrequent Itemsets

2. HDO Apriori

High Definition Oriented (HDO) Apriori [3] proposed idea of deletion of infrequent itemsets with lower dimensions. We can obtain a higher efficiency than that of original Apriori algorithm when dimension of data is high. HDO Apriori made changes in prune function of Apriori algorithm. By this functionality it can delete infrequent itemset. This algorithm introduces a new variable $L_k$in which infrequent itemsetare inserted into it instead of being deleted. To indicate whether an is frequent or not a flag item is added to indicate. To tackle the deficiency of Apriori Algorithm it purposed an opinion of deleting multiple infrequent itemsets during just one scan of candidate itemsets.

3. APFT Algorithm

APFT algorithm [5] is a combination of two algorithm that is FP Growth and Apriori algorithm from FP-Growth it uses a FP-Tree structure that defines in it. APFT algorithm construct a FP-tree in the beginning like the FP-Growth [7], such that all the branches include the items and generates the candidate itemsets using the Apriori algorithm's candidate-generate method. APFT algorithm removes the drawbacks of FP-Growth algorithm that FP-Growth algorithm fails when database is sparse or lot of frequent pattern result in more use of main memory. AFPT algorithm works on two steps; it constructs the FP-Tree as FP-Growth do. In second step, it use Apriori algorithm to mine the FP-tree. FP-tree use the divide and conquer strategy for mining process In second step it requires additional node table named NTable. NTable have two fields: Item-Name and Item support. APFT does not require extra spaces on the mining process. So, APFT has a better space scalability.

ISSN 2348 – 7968

4. Problem Statement

Compare the performance of different purposed improved Apriori algorithms and find how much efficiency they improve. Efficiency depends upon the various factors like value of minimum support[6], accuracy, number of records, etc.

5. Experimental Result

In this experiment, value of minimum support is change and its effect is seen on original Apriori and the HDO Apriori.

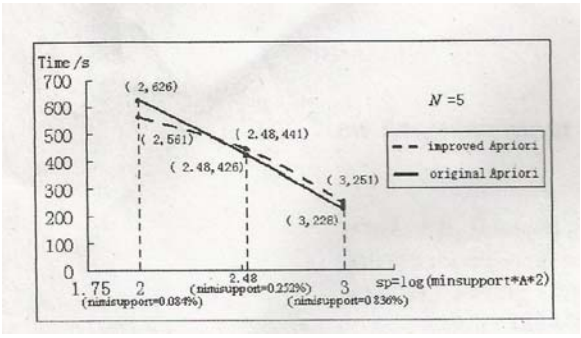

Figure 1: The comparison of runtime at different minimum supports between the original Apriori and the improved Apriori.

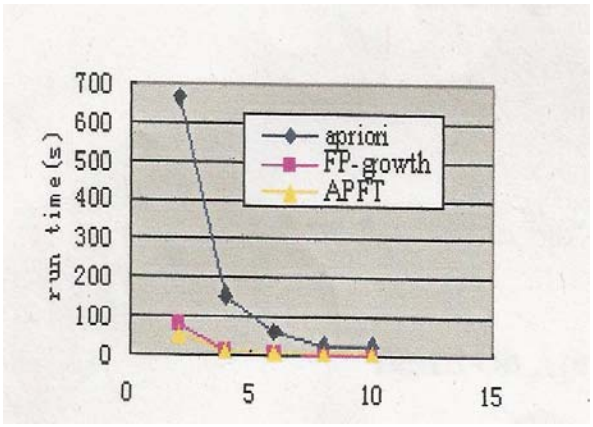

Figure 2: The comparision of runtime at different minimum supports between APFT, FP-growth, Apriori.

6. Conclusion

This paper represents the study of different improved version of Apriori Algorithm and their performance comparison with original Apriori Algorithm. Although every improved Apriori Algorithm purposed an improvement in efficiency of original Apriori Algorithm but still it needs research to find how much improvement these algorithm purposed.

7. References

[1] R. Agrawal and R. Srikant.Fast algorithms for mining association rules in large databases. Research Report RJ 9839, IBM Almaden Research Center, San Jose, California, June 1994.
[2] R. Agrawal, T. Imielinski, and A. Swami Database mining: A performance perspective.IEEE Transactions on Knowledge and Data Engineering, 5(6):914{925, December 1993. Special Issue on Learning and Discovery in Knowledge Based Databases.
[3] Lei Ji, Baowen Zhang, Jianhua Li, “A New Improvement on Apriori Algorithm”, IEEE,2006
[4] Margaret H.Dunham, “Data mining Introductory and Advanced Topics”, Pearson Education 2008.
[5] QihuaLan, Defu Zhang, Bo Wu, “A New Algorithm For Frequent Itemsets Mining Based On Apriori And FP-Tree”, IEEE, 2009
[6] J. Han, M. Kamber, “Data Mining Concepts and Techniques”, Morgan Kaufmann Publishers, San Francisco, USA, 2001, ISBN 1558604898.
[7] Ketan Shah, SunitaMahajan. 2009. A new efficient formulation for frequent item-set generation. In Proceedings of the International Conference on Advances in Computing, Communication and Control, Mumbai, India, January 23-24, 2009.

Author Sana Irfan is pursuing M. Tech (CSE) from Bharat Institute of technology, Meerut, Uttar Pradesh, India. Her research areas include Data Mining.