

# Performance Analysis of Frequent Pattern Mining Algorithm Using Apriori on Medical Data

Rahul Shukla<sup>1</sup>

Ph.D. Student Computer Science & Engineering  
Mewar university Chittorgarh, 321901, India

Anil Kumar Solanki<sup>2</sup>

Head Information Technology Deptt.  
BIET, Jhansi, India

---

**Abstract**— frequent pattern mining is an essential data mining task, with a goal of discovering knowledge in the form of repeated patterns. Many efficient pattern mining algorithms have been discovered to enhance the performance of Apriori Algorithm, The purpose of these algorithms to determining the frequent pattern. The main issue for any algorithm is to reduce the Execution time. In this paper we compare the frequent pattern mining Algorithm.

**Keywords**— Apriori, Association rule, Frequent Patterns, Support, Memory Based

---

## I. INTRODUCTION

Data mining is a way of mining the potentially useful information from the Data. Pattern recognition is a type of classification where an input pattern is classified into one of the several classes based on its similarity to these predefined classes. The current evolution of Data mining function and products is the result of years of influence from many disciplines including databases, information retrieval, statistical algorithm and machine learning, multimedia and graphics

This evolution is due to the support of three technologies that are sufficiently mature: massive data collection, high performance computing and data mining algorithm. Frequent pattern mining has applications ranging from intrusion detection and Market basket analysis, to credit card fraud prevention and drug discovery.

Knowledge Discovery in databases is the process of identifying a valid, potentially useful and ultimately understandable structure in data. This process involves selecting or sampling data from a data warehouse, cleaning or pre-processing it, transforming or reducing it, applying a data mining component to produce a structure, and then evaluating the derived structure. Thus, the structure that are the outcome of the data mining process must meet certain conditions so that these can be considered as knowledge. These conditions are validity, understand ability, utility, novelty and interestingness. Frequent pattern are the sequence of Items that occur frequently, close to each other and are extracted from the transactional database.

Frequent pattern mining is one of the most important and well researched techniques of data mining. The Association rule defined as “the task is to find the association between the presences of various item within the dataset.” The Association rule solely dependent on the Discovery of Frequent pattern. The presence of one set of items in a transaction implies other set of items.

## II. LIMITATION OF APRIORI

One is to find those item sets whose occurrences exceed a predefined Support in the database; those item sets are called frequent Pattern The Apriori Algorithm can be further divided into two sub-part candidate large item sets generation process and frequent item sets generation process. Frequent item set or large item set are those item sets whose support count exceeds the value of support threshold. Due to Number of passes apriori takes the more time. It scan the Database many time for Frequent pattern Discovery.

## III. APRIORI ALGORITHM FOR FREQUENT PATTERN MINING

Apriori algorithm is the very well-known algorithm for finding the association rule and is used in many commercial products. It uses the property of large Item set. “Any subset of large item set must be a large”. The large item set satisfies the minimum support requirement, so do all its subset.

The apriori algorithm is to generate candidate item set of particular size and then scan the database to count these to see if they are large. During the scan  $k$ , candidate of size  $k$ ,  $C_k$  are counted. Only those candidates that are large are used to generate candidate for the next pass. The  $L_k$  is used to generate  $C_{k+1}$ . An item set is considered as a candidate only if all its subset also is large. The apriori algorithm works in the two processes.

### A. Generation of Apriori:

This part of algorithm find the candidate itemset. it generates the superset of large item set.

### B. Pruning:

This part removes the candidate item set that have subset of size  $k-1$  that are not considered as large.

*C. Apriori algorithm pseudo code:*

```
Procedure Apriori_Generation ( $L_{i-1}$ )
{
// $L_{i-1}$ : Large item set of size i-1
 $C_i$  //candidate of size i
 $C_i = \phi$ ;
For each  $I \in L_{i-1}$  do
For each  $J \neq I \in L_{i-1}$  do
If i-2 of the elements in I and J are equal then
 $C_k = C_k \cup \{I \cup J\}$ ;
}

Procedure Pruning (I, D, s)
{
I //Item set
D //Dataset of transaction
S //Support
 $K=0$ ;
 $L = \phi$ ;
 $C_1 = I$ ;
Repeat
 $K=k+1$ ;
 $L_k = \phi$ ;
For each  $I_i \in C_k$  do
 $C_i = 0$ ; //Initial counts for each Item set are 0
For each  $t_j \in D$  do
For each  $I_i \in C_k$  do
If  $I_i \subset t_j$  then
 $C_i = C_i + 1$ ;
For each  $I_i \in C_k$  do
If  $C_i \geq (s \times |D|)$  do
 $L_k = L_k \cup I_i$ ;
 $L = L \cup L_k$ ;
 $C_{k+1} = \text{Apriori\_Generation}(L_k)$ ;
Until  $C_{k+1} = \phi$ ;
}
```

#### IV. MEMORY BASED APPROACH ON APRIORI ALGORITHM

This is the Memory efficient algorithm because it takes the very less memory when the query is fire, it takes those Frequent pattern whose length is greater than or equal to n. it is very much space saving. This algorithm is based on the apriori algorithm. The procedure of apriori algorithm searches the candidate item set in each horizontal transaction of any size. In this procedure count the support of candidate item set of size n, and search its existence in horizontal transaction whose size may be greater than less than or equal to the n. and in the memory based approach count the support of candidate Item set only in the transaction whose size is greater than or equal to the size of candidate item set, because it may exist only in those transaction whose size is greater than or equal to n. This approach takes very less run time as compared to Apriori Algorithm to generation of frequent pattern

#### V. EXPERIMENTS

In this section, we evaluate the performance of Apriori and Memory based Algorithm. To make the evaluation, we check the performance of Apriori and memory based approach on the different support count and fixed support count with different size of dataset

##### A. *DataSet*

In tune with our application, we have taken a dataset of Hospital with 633 items. In the analysis process we considered 2000 record in Horizontal Transaction format to generate the frequent pattern. In horizontal Transaction Each Transaction contains the multiple medicines in the single row. The transaction format contains the transaction Id (TID) and no. of Item (Medicine). The view of horizontal Transaction is given below in Table 1.

TID	Medicine
1	STERILE WATER 5ML,COLISPAS DROP,PERINORM INJ
2	MEROTEC 250 MG INJ
3	kefragard 0.75 inj
4	combiflam tab
5	MONOCEF SB 1GM INJ
6	RANTAC INJ,DISPO VAN 2ML SYREING,DISPO VAN 5ML SYRIENG,DISPO VAN 1 ML SYRIENG,Dany Ing.,DISPO VAN 50ML SYRINGE,AUGPEN-300 INJ

TABLE 1: Dataset of Hospital

VI. TIME COMPARISON OF APRIORI AND MEMORY BASED APPROACH

As a result of the experimental study, show that the performance of Apriori and Memory Based algorithm. The run time is the time to mine the frequent Pattern. We have 2064 transaction and The experimental is applied on 2000 transaction of hospital data(1st May to 27th May 2015) with different support count result of Execution time is shown in the Fig-1 reveals that the Memory Based is better than Apriori approach. In the analysis we use the different support count(5,6,7,8).In the study we observe that when the support count is less both the algorithm takes more time because there are large number of candidate item set those may become the frequent although memory based approach takes less Time for every value of support count when we increase the transaction the difference between the execution time is also Increase.

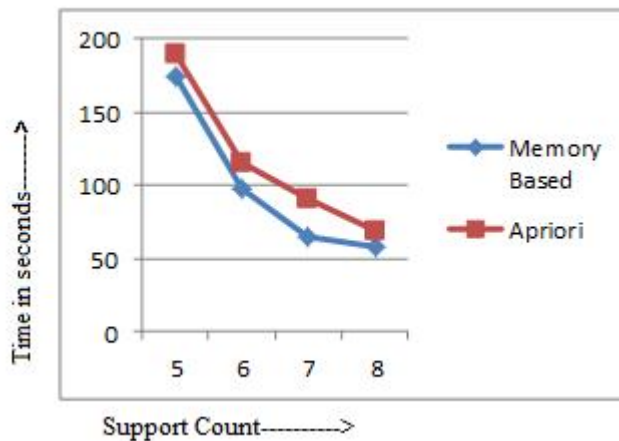


FIGURE 1: Execution Time Comparison on different support count

Now the experimental is again performed on different size of dataset with fixed support count 8 .The result of Execution time is shown in the Fig-2 reveals that the Memory Based performs better than Apriori

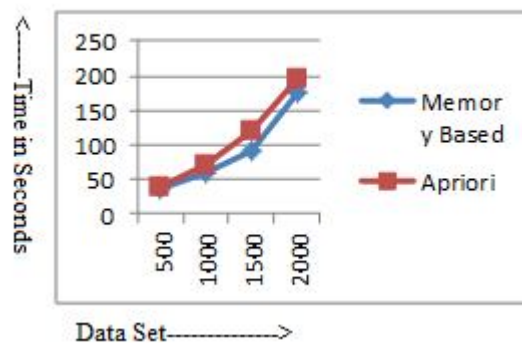


FIGURE 2: Execution Time Comparison on different size dataset

## VII. CONCLUSIONS

The association rules play a major role in many data mining applications, trying to find interesting patterns in data bases. Apriori is the simplest algorithm which is used for mining of frequent patterns from the transaction database. The main drawback of Apriori algorithm is that the candidate set generation is costly, especially if a large number of patterns and/or long patterns exist. Apriori algorithm uses large item set property, easy to implement, but it repeatedly scan the database. Apriori takes more time to scan the large Frequent patterns. The Memory based approach is used for efficient mining of frequent patterns in large databases. Memory based approach we count the support of candidate set only in the transaction record whose length is greater than or equal to the length of candidate set and it is more efficient than apriori algorithm and also takes lesser time and gives better performance.

## REFERENCES

- [1]. Rahul Mishra, Abha choubey, Comparative Analysis of Apriori Algorithm and Frequent Pattern Algorithm for Frequent Pattern Mining in Web Log Data, (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 3 (4) , 2012
- [2]. Huiping peng “Discovery of Interesting Association Rules on Web Usage Mining” 2010 International Conference.
- [3]. Agrawal R, Srikant R., "Fast Algorithms for Mining Association Rules", VLDB. Sep 12-15 1994, Chile, 487-99, pdf, ISBN 1-55860-153-8.
- [4]. Paresh Tanna, Dr. Yogesh Ghodasara, Using Apriori with WEKA for Frequent Pattern Mining, (IJETT) – Volume 12 Number 3 - Jun 2014
- [5]. Agrawal, R., Imielinski, T., and Swami, A. N. 1993. Mining association rules between sets of items in large databases. In Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, 207-216
- [6]. C.Borgelt. “Efficient Implementations of Apriori and Eclat”. In Proc. 1st IEEE ICDM Workshop on Frequent Item Set Mining Implementations, CEUR Workshop Proceedings 90, Aachen, Germany 2003.
- [7]. Krutika. K. Jain, Anjali . B. Raut, Review paper on finding Association rule using Apriori Algorithm in Data mining for finding frequent pattern; International Journal of Engineering Research and General Science Volume 3, Issue 1, January-February, 2015.
- [8]. Himani Bathla, Ms. Kavita Kathuria, Apriori Algorithm and filtered associator in association rule mining; International Journal of Computer Science and Mobile Computing; Vol. 4, Issue. 6, June 2015, pg.299 – 306.
- [9]. Shelly Ahuja, Gurpreet Kaur, Review of Association Rule Mining Using Apriori Algorithm; International Journal of Advanced Research in Computer Science and Software Engineering; Volume 4, Issue 3, March 2014.
- [10]. R. Agrawal, S. Gollapudi, A. Kannan, and K. Kenthapadi. Enriching textbooks with images. In CIKM, 2011.
- [11]. Anagnostopoulos, A., Becchetti, L., Castillo, C., Gionis, A., and Leonardi, S. Power in unity: forming teams in large-scale community systems. In CIKM (2010), 599–608
- [12]. Gitanjali J, C.Ranichandra,M.Pounambal, APRIORI algorithm based medical data mining for frequent disease identification, IPASJ International Journal of Information Technology (IJIT); Volume 2, Issue 4, April 2014.
- [13]. Zhang Danping, Yu Haoran and Zheng Linyu, Apriori Algorithm Research Based on Map-Reduce in Cloud Computing Environments; Open Automation and Control Systems Journal, 2014, 6, 368-373
- [14]. Shruti Aggarwal, Ranveer Kaur, “Comparative Study of Various Improved Versions of Apriori Algorithm”, International Journal of Engineering Trends and Technology (IJETT) - Volume4 Issue4- April 2013
- [15]. Adriano Veloso, Matthew Eric Otey, Srinivasan Parthasarathy, Wagner Meira Jr.; Parallel and Distributed Frequent Itemset Mining on Dynamic Datasets; Int'l Conf. on High Performance Computing; 2003.
- [16]. Haixun Wang, Wei Fan, Philip S. Yu, Jiawei Han; Mining Concept-Drifting Data Streams using Ensemble Classifiers; ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining; August 2003.
- [17]. Ran Wolff, Assaf Schuster; Association Rule Mining in Peer-to-Peer Systems; IEEE Transactions on Systems, Man and Cybernetics, Part B, Vol. 34, Issue 6; December 2004.