Comparative Analysis of Top-Down & Bottom-up Approach (Apriori Algorithm)

Gaurav Pandey[1] N. K. Gupta[2] Mohit Paul[3]
[1]M.Tech Student
[1,2,3]Department of Computer Science & Information Technology
[1,2,3]Sam Higginbottom Institute of Agriculture, Technology and Sciences, Allahabad

Abstract— Tremendous amount of data being collected is increasing speedily by computerized applications around the world. Hidden in the vast data, the valuable information is attracting researchers of multiple disciplines to study effective approaches to derive useful knowledge from within. Among various data mining objectives, the mining of frequent patterns has been the focus of knowledge discovery in databases. This thesis aims to investigate efficient algorithm for mining including association rules and sequential patterns. Many algorithms have been proposed from last many decades including horizontal layout based techniques, vertical layout based techniques, and projected layout based techniques. But most of the techniques suffer from repeated database scan, Candidate generation (Apriori Algorithms), memory consumption problem and many more for mining frequent patterns. As in retailer industry many transactional databases contain same set of transactions many times, to apply this thought, in this thesis present an improved Apriori algorithm that guarantee the better performance than classical Apriori algorithm.

Key words: Top-Down, Bottom-Up, Support, Confidence, Association Rule, and Frequent Itemsets

I. INTRODUCTION

With the enhance in Information Technology, the size of the databases created by the organizations due to the accessibility of low-cost storage and the development in the data capturing technologies is also increasing. These association sectors include retail, fuel, telecommunications, utilities, manufacturing, transport, credit cards, insurance, banking and many others, extracting the valuable data, it required to explore the databases completely and efficiently. Knowledge discovery in databases (KDD) helps to identifying valuable information in such huge databases. This valuable information can help the decision maker to make perfect future decisions. KDD applications deliver measurable benefits, including reduced cost of responsibility business, improved profitability, and improved quality of service. Therefore Knowledge Discovery in Databases has developed into one of the most active and exciting research areas in the database community.

Data mining is an essential step in the process of knowledge discovery in databases, in which intelligent methods are applied in order to extract patterns. Other steps in knowledge discovery process include pre-mining tasks such as data cleaning (removing noise and inconsistent data) and data integration (bringing data from multiple sources to a single location and into a common format), as well as post mining tasks such as pattern evaluation (identifying the truly interesting patterns representing knowledge) and knowledge presentation (presenting the discovered rules using visualization and knowledge representation techniques).
In data mining, association rule learning is a popular and well researched method for discovering interesting relations between variables in large databases. Piatetsky-Shapiro describes analyzing and presenting strong rules discovered in databases using different measures of interestingness. Based on the concept of strong rules, Agrawal et al introduced association rules for discovering regularities between products in large scale transaction data recorded by point-of-sale (POS) systems in supermarkets

A. Fundamental Components of Data Mining Technology:

It is fundamentally important to declare that the prime key to understand and realize the data mining technology is the ability to make different between data mining, operations, applications and techniques [2], as shown in Fig 1.

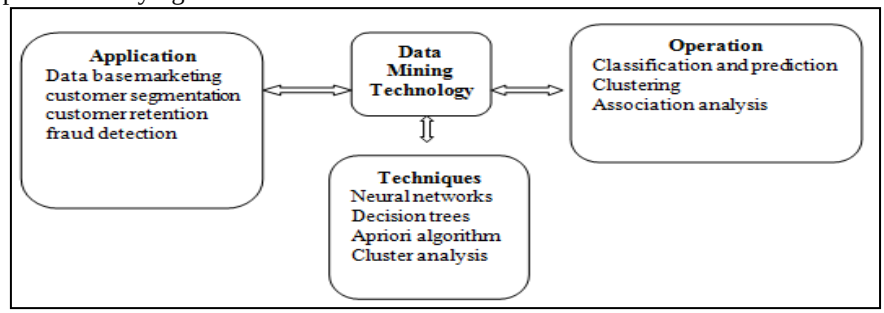

Fig. 1: Components of Data Mining

II. ASSOCIATION RULE MINING

The association rule of data mining is a elementary topic in mining of data [3]. Association rule mining discovery frequent patterns, associations, correlations, or fundamental structures along with sets of items or objects in transaction databases, relational databases, and other information repositories [4].

A lot of studies have been done in the region of association rules mining. First introduced the association rules mining in [5,6]. Many studies have been conducted to address various conceptual, implementation, and application issues relating to the association rules mining task.

The overall performance of mining association rules is determined primarily by the first step. The second step is easy. After the large itemsets are identified, the corresponding association rules can be derivative in straightforward manner. Our main consideration of the thesis is First step i.e. to find the extraction of frequent item sets [9].

III. METHODOLOGY

A. *Classical Apriori Algorithm:*

Apriori employs an iterative approach known as a level-wise search [15], where k-itemsets are used to explore (k+1)-itemsets. First, the set of frequent 1-itemsets is found. This set is denoted L1.L1 is used to find L2, the set of frequent 2-itemsets, which is used to find L3, and so on, until no more frequent k-itemsets can be found. The finding of each Lk requires one full scan of the database. In order to find all the frequent itemsets, the algorithm adopted the recursive method. The main idea is as follows [16]

```
L1 = {large 1-itemsets};
  for (k=2; Lk-1≠Φ; k++) do
  {
    Ck=Apriori-gen (Lk-1); //the new
candidates
    for each transactions t∈D do//scan D
for counts
  {
    Ct=subset(Ck, t);
    // get the subsets oft that are
candidates
    for each candidates c∈ Ct do
      c.count++;
  }
    Lk={c∈ Ck |c.count≥minsup}
  }
    Return=UkLk;
```

Fig. 2:

B. *Limitations of Apriori:*

- In case of large dataset, this algorithm is not efficient [19].
- Apriori algorithm requires large no of scans of dataset [19].
- In case of large dataset, Apriori algorithm produces large number of candidate itemsets. Algorithm scan database repeatedly for searching frequent itemsets, so more time and resource are required in large number of scans so it is inefficient in large datasets [20].

C. *Novel Approach:*

The major objective of the research is to develop and propose a new idea for mining the association rules out of transactional data set. The proposed method is based on Improved Apriori approach. The proposed method is more efficient than classical Apriori algorithm. To achieve the research objective successfully, a series of sequence progresses and analysis steps have been adopted. Figure 5 depicts the method to mine frequent itemsets from the transactional data set using the new method.

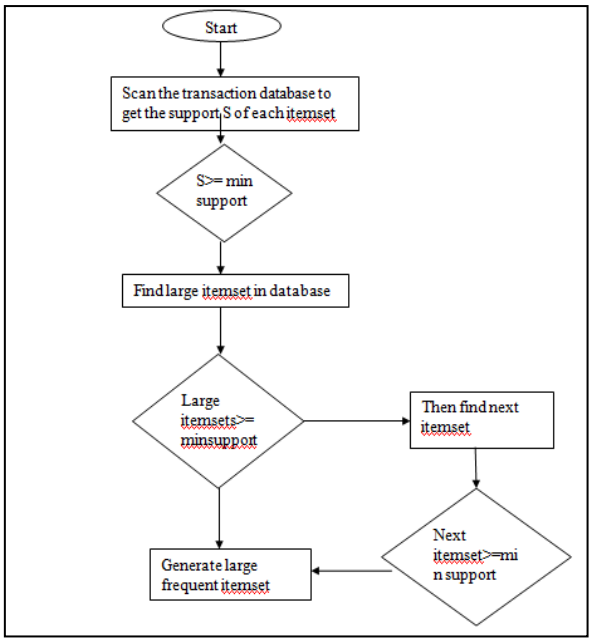

Fig. 3: Flow Chart of Improve Apriori Algorithm

```
Input: database (D), minimum support (min_sup).
Output: frequent item sets in D.
  L1= frequent item set (D)
  j=k; /* k is the maximum number of element in a
transaction from the database*/
            for k= maxlength to 1
            {
            for i=k to 2
             {
            for each transaction Ti of order i
            {
            if (Ti has repeated)
            {
              Ti.count++;
            }
             m=0;
             while (i<j-m)
            {
             if (Ti is a subset of each transaction Tj-m
of order j-m)
              Ti.count++; m++; }
             If (Ti.count >=min_sup)
           Rule Ti generated

          }
```

Fig. 4:

D. *Testing and Result:*

The experiments that we have performed to evaluate the new proposal. For the estimation purpose we have conducted several experiments using the existing data set. Those experiments performed on computer with Core 2 Duo 2.00 GHZ CPU, 2.00 GB memory and hard disk 80 GB. This algorithm was developed by java language using net beans IDE 7.3.1 and for the unit of measuring the time and no. of iteration

E. *Time Comparison with No of Transaction:*

As a result of the experimental study, revealed the performance of our improved Apriori with the Classical

Apriori algorithm. The run time is the time to mine the frequent itemsets. The experimental result of time is shown in Figure 5-1 reveals that the proposed scheme outperforms the Apriori approach.

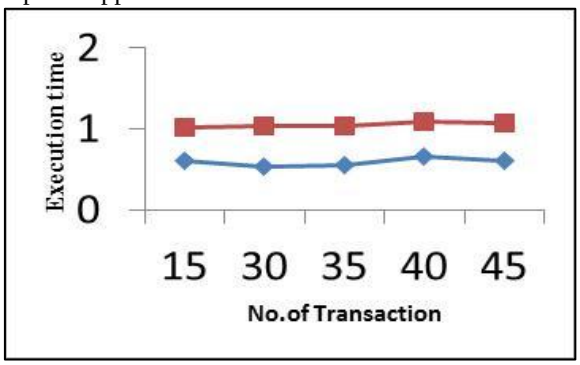

Fig. 5: Depicting Execution Time With Respect To No. Of Transaction

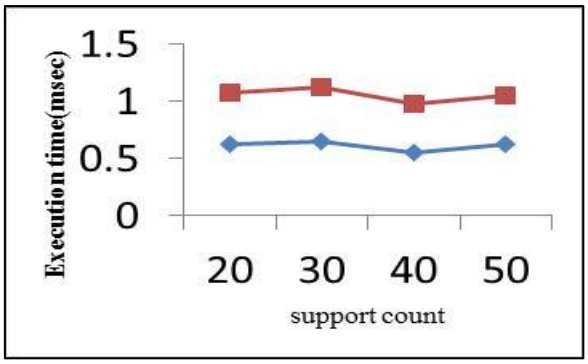

Fig. 6: Depicting Relationship of Support Count With Time Consumption

IV. EXPERIMENTAL ENVIRONMENT

The following snapshots are showing the output after the execution of the complete algorithm. This snapshot display menu in which user interact easily. We are import transact and configure file and run the both algorithms. The below diagram shows snapshots of all work that we perform.

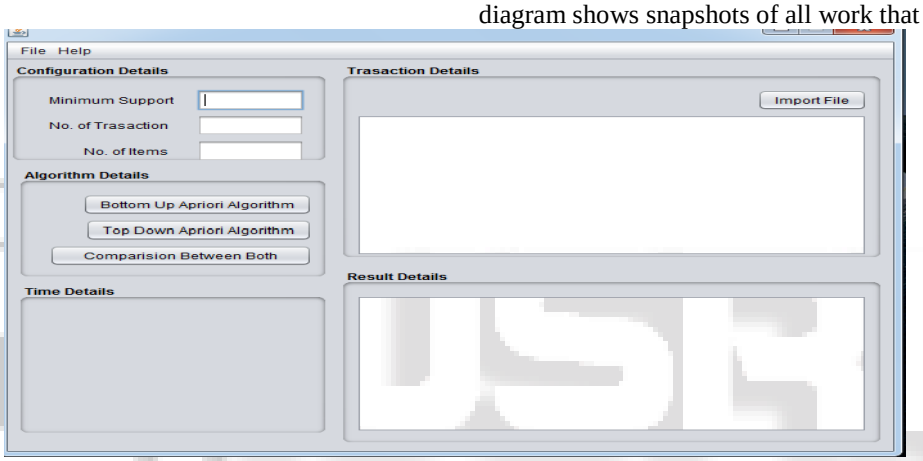

Fig. 7: Snapshots to GUI

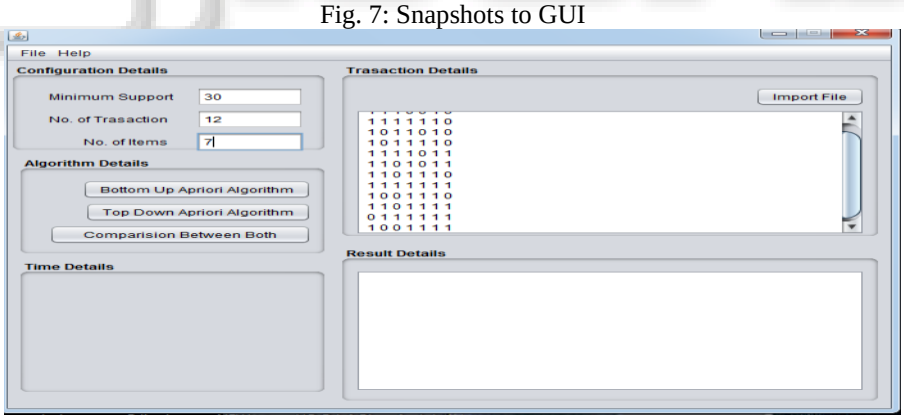

Fig. 8: Snapshot to Transaction and Configuration Details

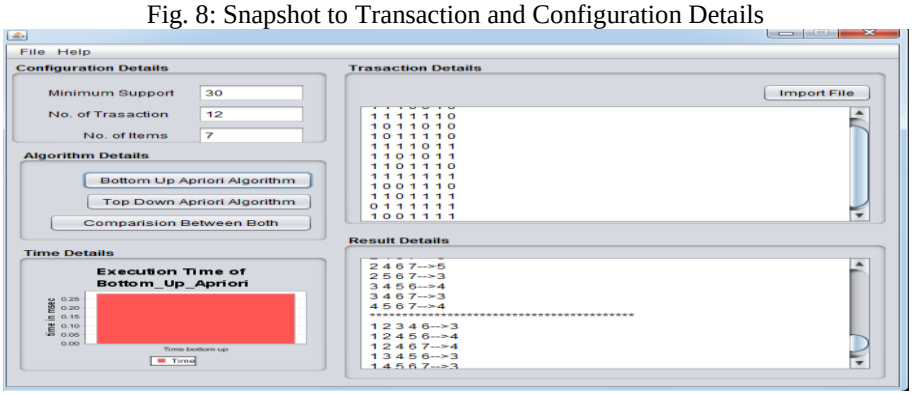

Fig. 9: Snapshot to Bottom up Apriori Algorithm

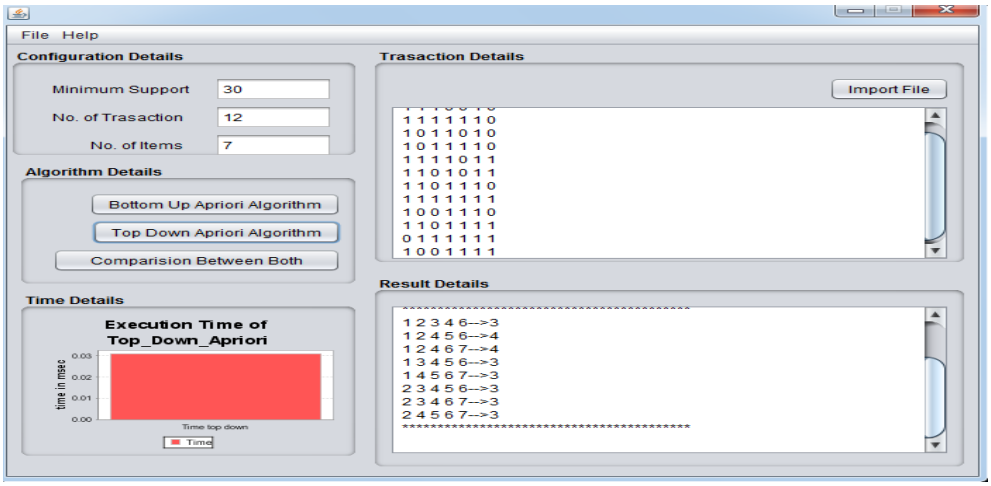

Fig. 10: Snapshot to Top Down Apriori Algorithm

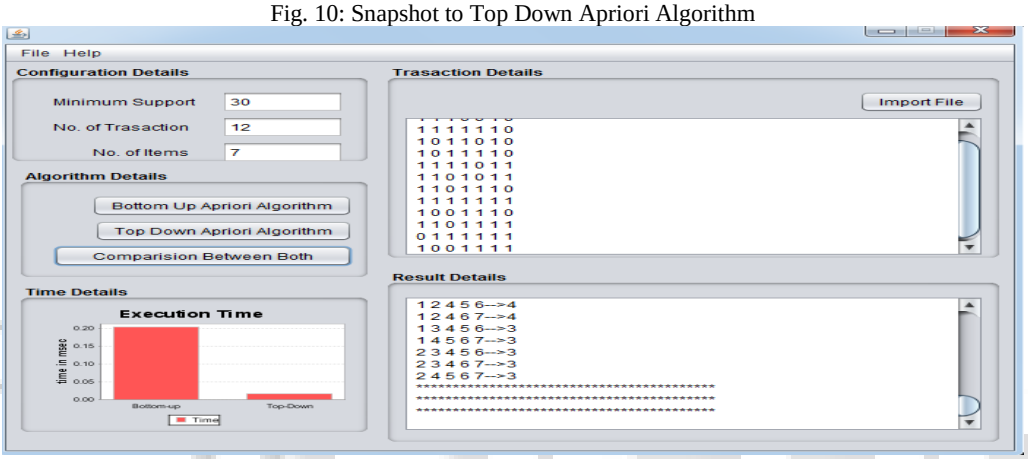

Fig. 11: Snapshot to Comparison between Both Algorithms

V. CONCLUSION

In this thesis, we measured the following factors for creating our new idea, which are the time and the no of iteration, these factors, are affected by the approach for finding the frequent itemsets. Work has been done to develop an algorithm which is an improvement over Apriori with using an approach of improved Apriori algorithm for a transactional database. According to our clarification, the performances of the algorithms are strongly depends on the support levels and the features of the data sets (the nature and the size of the data sets). Therefore we employed it in our scheme to guarantee the time saving and reduce the no of iteration. Thus this algorithm produces frequent itemsets completely. Thus it saves much time and considered as an efficient method as proved from the results.

REFERENCES

[1] Tan P.N., Steinbach M., and Kumar V: Introduction to data mining, Addison Wesley Publishers, 2006.

[2] Han J. & Kamber M.: Data Mining Concepts and Techniques, First edition, Morgan Kaufmann publisher, USA 2001.

[3] Ceglar, A., Roddick, J. F: Association mining ACM Computing Surveys, volume 38(2) 2006.

[4] Jiawei Han, Micheline Kamber , Morgan Kaufmann : Data mining Concepts and Techniques , 2006.

[5] A. Savasere, E. Omiecinski and S. Navathe. : An efficient algorithm for mining Association rules in large databases, In Proc. Int"l Conf. Very Large Data Bases (VLDB), Sept. 1995, p.p 432–443.

[6] Agrawal.R and Srikant R.: Fast algorithms for mining association rules,In Proc. Int"l Conf. Very Large Data Bases (VLDB), Sept. 1994, p.p 487–499.

[7] Lei Guoping, Dai Minlu, Tan Zefu and Wang Yan: The Research of CMMB Wireless Network Analysis Based on Data Mining Association Rules, IEEE conference on Wireless Communications, Networking and Mobile Computing (WiCOM),ISSN :2161- 9646 Sept. 2011,p.p 1-4.

[8] Divya Bansal, Lekha Bhambhu : Execution of APRIORI Algorithm of Data Mining Directed Towards Tumultuous Crimes Concerning Women, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 9, ISSN: 2277 128X September 2013 .

[9] Shweta, Dr. Kanwal Garg: Mining Efficient Association Rules Through Apriori Algorithm Using Attributes and Comparative Analysis of Various Association Rule Algorithms International Journal of Advanced Research in Computer Science and Software Engineering 3(6), June – 2013, pp. 306-312.

[10] Suraj P. Patil1, U. M. Patil2 and Sonali Borse: The novel approach for improving Apriori

algorithm for mining association Rule, World Journal of Science and Technolog 2(3), ISSN: 2231 – 2587, 2012, p.p75- 78.

[11] Toivonen .H. :Sampling large databases for association rules,In Proc. Int‟l Conf Very Large Data Bases (VLDB), Bombay, India, Sept. 1996, p.p 134–145.

[12] Yanfei Zhou, Wanggen Wan, Junwei Liu, Long Cai: Mining Association Rules Based on an Improved Apriori Algorithm 978-1-4244-585 8-5/10/ IEEE 2010.

[13] Luo Fang: The Study on the Application of Data Mining Based on Association Rules, International Conference on Communication Systems and Network Technologies (IEEE) ,may 2012,p.p 477 - 480 .

[14] Chengyu and Xiong Ying: Research and improvement of Apriori algorithm for Rules, In Intelligent Systems and Applications (ISA), 2nd International Workshop on, may 2010, p.p 1 -4.

[15] Jaishree Singh, Hari Ram, Dr. J. S. Sodhi: Improving Efficiency of Apriori Algorithm Using Transaction Reduction, In proceeding of International Journal of Scientific And Research Publication (IJSRP), ISSN 2250-3153, Volume 3, Issue 1, January 2013,p.p1-4.

