

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor & UGC Approved Journal)

Website: [www.ijirset.com](http://www.ijirset.com)

Vol. 6, Issue 8, August 2017

## Analysis of Association Rule Mining Algorithms to Generate Frequent Itemset

G.Suganya<sup>1</sup>, K.J.Paulraj Ananth.<sup>2</sup>

M. Phil Student, Department of Computer Science, Sun Arts & Science College, Keeranor, Thiruvannamalai, India<sup>1</sup>

Assistant Professor, Department of Computer Science, Sun Arts & Science College, Keeranor,  
Thiruvannamalai, India<sup>2</sup>

**ABSTRACT:** Association rule mining algorithm is used to extract relevant information from database and transmit into simple and easiest form. Association rule mining is used in large set of data. It is used for mining frequent item sets in the database or in data warehouse. It is also one type of data mining procedure. In this paper some of the association rule mining algorithms such as apriori, partition, FP-growth, genetic algorithm etc., can be analyzed for generating frequent itemset in an effective manner. These association rule mining algorithms may differ depend upon their performance and effective pattern generation. So, this paper may concentrate on some of the algorithms used to generate efficient frequent itemset using some of association rule mining algorithms.

**KEYWORDS:** Data Mining- Apriori-Association Rule Mining-Frequent Itemset.

### 1. INTRODUCTION

#### DATA MINING

A large amount of data can be accessed in this world at any instance of time. All over the world data is scattered in various field for effective retrieval of relevant data. Data may be in structured or unstructured data. Data mining is the procedure that can be used to extract information from large volume of data. It involves many techniques like clustering, Classification, association rule mining etc., Data mining is also referred as mining knowledge from database. Some of the steps that are involved in knowledge discovery process are:

- Data Cleaning
- Data Integration
- Data Selection
- Data Transformation
- Data Mining
- Pattern Evaluation
- Knowledge Presentation

#### ASSOCIATION RULE

Association is a data mining function that discovers the probability of the co-occurrence of items in a collection. The relationships between co-occurring items are expressed as association rules. Association rules are often used to analyze sales transactions.

### II. ASSOCIATION RULE MINING

The goal of association rule finding is to extract correlation relationships in the large datasets of items. Ideally, these relationships will be causative. Many businesses are interested in extracting interesting relationships in order to raise profit. Scientists, on the other hand, are interested in discovering previously unknown patterns in their field of research. An illustrative example of association rule mining is so-called market basket analysis. This process is based on transactional data, which are huge amounts of records of individual purchases. For example, department stores such as

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor & UGC Approved Journal)

Website: [www.ijirset.com](http://www.ijirset.com)

Vol. 6, Issue 8, August 2017

Wal-Mart<sup>[14]</sup> keep track of all transactions in a given period of time. One transaction could be  $t_1 = \{\text{Italian bread, 1\% milk, energizer batteries}\}$

It can contain an arbitrary (technically, it is limited to what's in the store) number of items. All items purchased belong to a set of items  $\Omega$ . In the computer science jargon, such a set is typically called a universe. Another notational detail is that transaction  $t_1$  can also be called an itemset, because it contains a set of purchased items. Support and confidence can be considered as measures the interestingness of the rules.

## NOTATION AND BASIC CONCEPTS

Let  $\Omega = \{i_1, i_2, \dots, i_m\}$  be a universe of items. Also, let  $T = \{t_1, t_2, \dots, t_n\}$  be a set of all transactions collected over a given period of time. To simplify a problem, we will assume that every item  $i$  can be purchased only once in any given transaction  $t$ . Thus,  $t \subseteq \Omega$  ("t is a subset of omega"). In reality<sup>[14]</sup>, each transaction  $t$  is assigned a number, for example a transaction id (TID). Let now  $A$  be a set of items (or an itemset). A transaction  $t$  is said to contain  $A$  if and only if  $A \subseteq t$ .

Now, mathematically, an association rule will be an implication of the form  $A \Rightarrow B$  where both  $A$  and  $B$  are subsets of  $\Omega$  and  $A \cap B = \emptyset$  ("the intersection of sets  $A$  and  $B$  is an empty set").

### Support:

Support (sometimes called frequency) is simply a probability that a randomly chosen transaction  $t$  contains both itemsets  $A$  and  $B$ .

Mathematically,

$$\text{Support}(A \Rightarrow B) = P(A \wedge B)$$

### Confidence:

Confidence (sometimes called accuracy) is simply a probability that an itemset  $B$  is purchased in a randomly chosen transaction  $t$  given that the itemset  $A$  is purchased. Mathematically, we will use a simplified notation that:

$$\text{Confidence}(A \Rightarrow B) = P(B | A)^{[14]}$$

## III. ASSOCIATION RULE MINING ALGORITHMS

Some of the association rule mining algorithms are

- Apriori
- Partition
- Pincer search
- Border algorithm
- Fp-growth
- Genetic Algorithm etc.,

These are some of the algorithms that are used to find frequent itemset in an efficient manner. In this paper we focused on the very first developed apriori algorithm to find frequent itemset in large set of database.

## IV. APRIORI ALGORITHM

Apriori algorithm is used for frequent itemset mining and association rule learning. The algorithm use a level-wise search, where  $k$ -itemsets (An itemset which contains  $k$  items is known as  $k$ -itemset) are used to explore  $(k+1)$  itemsets, to mine frequent itemsets from transactional database for Boolean association rules. In this algorithm, frequent subsets are extended one item at a time and this step is known as candidate generation process. Then groups of candidates are tested against the data<sup>[4]</sup>. To count candidate item sets efficiently, Apriori uses breadth-first search method and a hash tree structure. It identifies the frequent individual items in the database and extends them to larger and larger item sets

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor & UGC Approved Journal)

Website: [www.ijirset.com](http://www.ijirset.com)

Vol. 6, Issue 8, August 2017

as long as those item sets appear sufficiently often in the database. Apriori algorithm determines frequent item sets that can be used to determine association rules which highlight general trends in the database<sup>[4]</sup>.

---

## Apriori algorithm pseudo code:

---

```
procedure Apriori (T, minSupport)
{
//T is the database and minSupport is the minimum support L1= {frequent items};

for (k= 2; Lk-1 !=∅; k++)
{
Ck= candidates generated from Lk-1

//that is cartesian product Lk-1 x Lk-1 and
//eliminating any k-1 size itemset that is not
//frequent

for each transaction t in database do
{

#increment the count of all candidates in Ck
that are contained in t
Lk = candidates in Ck with minSupport

} //end for each

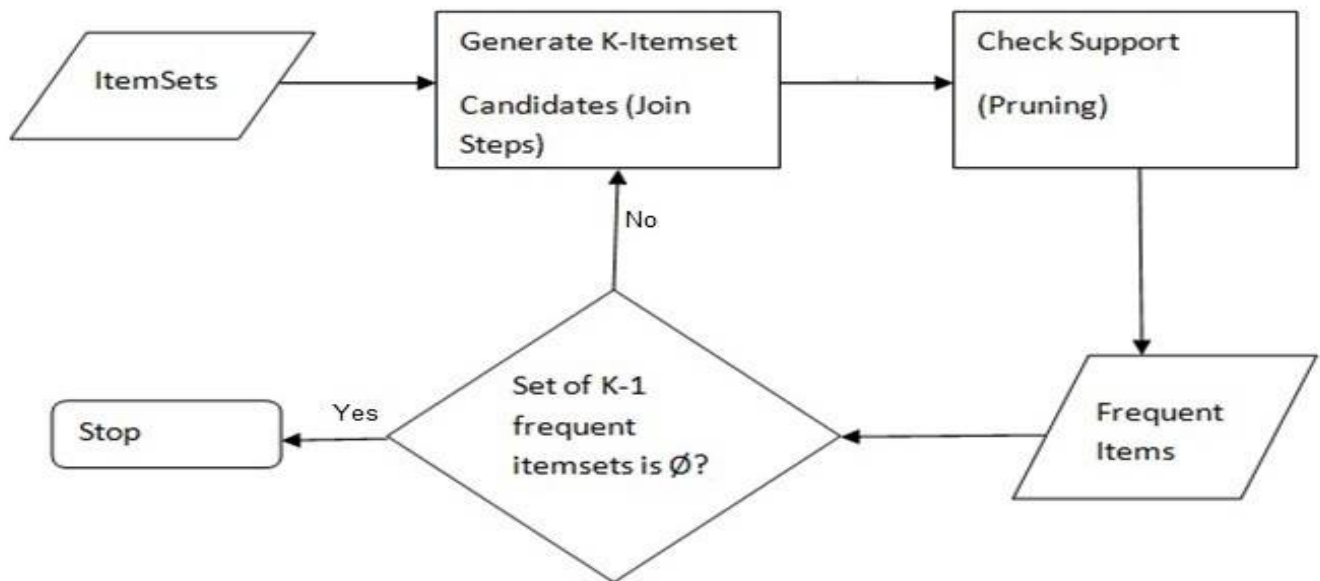
} //end for
return ;
}[4]
```

---

Association rule generation is usually split up into two separate steps:

1. First, minimum support is applied to find all *frequent itemsets* in a database.
2. Second, these frequent itemsets and the minimum confidence constraint are used to form rules. While the second step is straight forward, the first step needs more attention. Finding all frequent itemsets in a database is difficult since it involves searching all possible itemsets (item combinations<sup>[4]</sup>).

**FIND FREQUENT ITEMSET USING APRIORI<sup>[15]</sup>**



**Figure 1**

The above figure shows the frequent itemset generation using Apriori Algorithm. First take the item sets and generate K-itemset candidates and next to check for pruning in the item sets. Then find the most frequent items in set of K-1 frequent itemsets.

**V. LITERATURE REVIEW**

**APRIORI Algorithm: An overview of association rule mining algorithms:** It identifies the frequent individual items in the database and extends them to larger and larger item sets as long as those item sets appear sufficiently often in the database<sup>[4]</sup>.

**Applying correlation threshold on apriori algorithm:** To increase the efficiency of present apriori algorithm a new method for incorporating a new correlation factor threshold is introduced to reduce time complexity and create more frequent itemset in less time<sup>[1]</sup>.

**YAFIM: A parallel frequent itemset mining algorithm with spark:**(YAFIM->Yet Another Frequent Itemset Mining) A parallel Apriori algorithm based on RDD framework-specially designed in memory parallel computing model to support iterative algorithms and interactive data mining. It is achieved by 18x speed up in average of various assumptions. It makes good use of memory<sup>[2]</sup>.

**APRIORI-AP: To address more complex pattern mining problems:**(AP-Accelerated Apriori). It avoids routine configuration of database. It matches and counts up to 18,432 itemsets in parallel. Its solution also outperforms CPU implementations of Elcat especially for large datasets<sup>[3]</sup>.

**APRIORITID ALGORITHM:** Each member has TID of each transaction. Large itemset is present in this transaction. A new set c is generated for finding candidate itemset support. The set c is used to count the support of each candidate itemset<sup>[4]</sup>.

**APRIORIHYBRID ALGORITHM:** It uses both Apriori and AprioriTID. In earlier it passes Apriori algorithm and later it passes AprioriTID algorithm.<sup>[4]</sup>

**ENHANCED APRIORI ALGORITHM:** This algorithm is an efficient and scalable method for mining the complete set of frequent patterns. It takes less time for scanning database. It can be achieved by eliminating redundant generation of sub items. It also reduces the size of the database. Transaction size is used to represent the number of items in individual transaction<sup>[5]</sup>.

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor & UGC Approved Journal)

Website: [www.ijirset.com](http://www.ijirset.com)

Vol. 6, Issue 8, August 2017

**IMPROVED APRIORI ALGORITHM:** This algorithm can save time, cost and increase the efficiency of data mining. It decline the number of candidate itemsets. For dense database and in longer forms its performance is good<sup>[6]</sup>.

**DHP ALGORITHM (DIRECT HACHING AND PRUNING):**It is an extension of apriori algorithm. It uses hashing technique. It is used to generate efficient large itemset. It reduces the transaction database size. Any transactions does not contain frequent itemset, that transactions are removed<sup>[7]</sup>.

**GSP (GENERALIZED SEQUENTIAL PATTERNS)**It is representative of Apriori based sequential pattern mining algorithm. It uses the downward closure property of sequential patterns and it also adapts multi pass of candidate generate-and-test approach<sup>[7]</sup>.

**REVERSE APRIORI:** This algorithm is used to generate large frequent itemsets only if it satisfies user specified minimum item support. It decrease the number of items in the itemsets until it get the largest frequent itemsets. It is used for frequent pattern production<sup>[7]</sup>.

**SCALING APRIORI :**This approach is used to reduce the number of items that are generated. It also reduce the overall execution time of the algorithm. It mainly used to minimize the number of candidate set. Evaluating quantitative information associated with each itemset that occurs in transaction<sup>[7]</sup>.

**APRIORI WITH WEKA FOR FREQUENT PATTERN MINING:** This may demonstrate the use of WEKA tool for association rule mining using apriori algorithm. WEKA for referring the process of association rule mining<sup>[8]</sup>.

**A NEW APPROACH FOR MINING FUZZY ASSOCIATION RULES USING APRIORI ALGORITHM:** Apriori algorithm specify the minimum support in fully automated manner. It uses fuzzy logic to distribute data in different clusters. In that the introduction of appropriate threshold automatically. It is simply called averaging technique. In this approach, they attempt to transfer crisp data to a fuzzy environment. It automatically determines the threshold for the minimum support by the algorithm itself<sup>[9]</sup>.

**IMPROVING EFFICIENCY OF APRIORI ALGORITHM USING TRANSACTION REDUCTION:**Improved Apriori algorithm reduces the scanning time by cutting down unnecessary transaction records. It also reduce the redundant generation of sub-items during pruning the candidate itemset. It eliminate the candidate having a subset that is not frequent. It reduce I/O Spending<sup>[10]</sup>.

**EFFICIENTLY MINING FREQUENT ITEMSET ALGORITHMS(COUNT TABLE):**Count table is one of the important facility to employ subsets property for compressing the transaction database to lower representation of occurrences items. It avoid costly candidate generate-and-test, expensive and repeated database scans. CountTableFI and BinaryCountTableF are extended from Apriori. It represent all transaction in binary numbers and decimal numbers. It is simple and fast to use. Does not need to generate candidate itemset<sup>[11]</sup>.

**MR-APRIORI:ASSOCIATION RULE MINING BASED ON MAPREDUCE:** One of the major problem is power shortage in dealing with massive datasets. In order to overcome this problem ,a distributed association rule algorithm based on Mapreduce programming model is used. It make sharp increase in efficiency<sup>[11]</sup>.

**APRIORI BASED ALGORITHM FOR MINING FREQUENT SUBSTRUCTURES FROM GRAPH DATA:** It is a novel approach named AGM. It efficiently mine the frequently appearing substructure in a given graph data set. Frequent patterns appear in matrices<sup>[12]</sup>.

**FPGA WITH APRIORI ALGORITHM:** Utilizes minimum of hardware and minimum of time. It also insures that utilization of the hardware comparator is near 100%.Candidate generation and pruning operations are more complicated. So it store only needed data to reduce memory and size<sup>[13]</sup>.

**BOOSTED APRIORI:** It is an effective data mining association rule for heart disease prediction system. The dataset those are preprocessed and normalized. This algorithm performance is good to find heart disease and also make some predictions also. The classification accuracy can be improved by reduction in features<sup>[16]</sup>.

**DP-APRIORI:**A differentially private frequent itemset mining algorithm based on transaction splitting. Truncate of long transactions may reduce performance and also opportunity to lose dat. To avoid this problem the long transactions are splitted not truncated. Small weighted technique is proposed to divide long transaction into sub transaction<sup>[17]</sup>.

**BIT-APRIORI:** An Apriori based frequent itemsets mining using Bit streams.A new method using bit stream to improve their speed. A map of all attributes is created. After that each attribute presented by 1 bit, that are processed by logic operations<sup>[19,24]</sup>.

**SEMI-APRIORI ALGORITHM FOR FREQUENT ITEMSET:** A new binary based semi-approach technique that efficiently discovers the frequent itemset. It reduces execution time<sup>[19,23]</sup>.

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor & UGC Approved Journal)

Website: [www.ijirset.com](http://www.ijirset.com)

Vol. 6, Issue 8, August 2017

**C-APRIORI:CONVICTION BASED APRIORI ALGORITHM FOR DISCOVERING FREQUENT DETERMINANT PATTERNS FROM HIGH DIMENSIONAL DATASETS:** This CApriori is used to prune the generate frequent determinant sets effectively. Frequent determinant set is compared with conviction threshold value and then compared with support threshold. So, it makes double comparison to eliminate redundancy and generate strong association rules. It also minimizes main memory utilization. This algorithm may access high dimensional databases<sup>[19, 22]</sup>.

**ADAPTIVE APRIORI ALGORITHM FOR FREQUENT ITEMSET MINING:**The size of the database for defining candidates has great effect on running time and memory need.So, it is outperform Apriori<sup>[19,20]</sup>.

**META-APRIORI:A NEW ALGOTITHM FOR FREQUENT PATTERN DETECTION:**A new recursive algorithm based on Apriori. It aim to reduce substantially the run time, so as to increase its efficiency while preserve its effectiveness. In this a most valuable technique Divide-and-Conquer is used to partition the whole database and then apply Meta-Apriori for the part of database. So, this way may find the frequent item set in an effective manner<sup>[19, 21]</sup>.

**K-APRIORI ALGORITHM:** Binary data is clustered based on a linear data transformation technique, then frequent itemsets and association rules are mined. K-Apriori is a multi-pass algorithm. It is mainly used in market basket transactions. This algorithm is the advanced version of Apriori algorithm. It may used in large set of data transactions with frequent itemset retrieval. Performance is also high compared to apriori<sup>[18]</sup>.

## VI. CONCLUSION

Data mining plays a vital role in many of the sectors like industry, institutions, companys,Finance, sports, medical, aerospace etc.,The data can be accessed much more depend upon our needs. Association rule mining algorithms will give frequently related data that are needed to the user in a correct way.So,there is no chance to lose any related data.Frequent itemset can be generated to see the related data of what we required also.In this paper we discussed the Apriori algorithm and their types to find frequent itemset.This algorithm may give relevant need with less execution time.So. apriori algorithm plays a major role to find frequent item set in effective manner.In future ,we may concentrate on other association rule mining algorithms such as fp-growth, genetic algorithm for finding frequent item set in very large set of database transactions.

## REFERENCES

1. Anand H.S., Vinodchandra S.S.:" Applying Correlation Threshold on Apriori Algorithm" IEEE International Conference on Emerging Trends in Computing, Communication and Nanotechnology (ICECCN 2013).
2. Hongjian Qiu, Rong Gu, Chunfeng Yuan, Yihua Huang:"YAFIM: A Parallel Frequent Itemset Mining Algorithm with Spark", 2014 IEEE 28th International Parallel & Distributed Processing Symposium Workshops
3. Ke Wang, Yanjun Qi, Jeffrey J. Fox, Mircea R. Stanand Kevin Skadro:"Association Rule Mining with the Micron Automata Processor
4. Trupti A. Kumbhare, Prof. Santosh V. Chobe:" An Overview of Association Rule Mining Algorithms" Trupti A. Kumbhare et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (1), 927-930, issn no:0975-9646, 2014
5. Logeswari T, Ph.D, Valarmathi N, Sangeetha A, Masilamani M:" Analysis of Traditional and Enhanced Apriori Algorithms in Association Rule Mining" International Journal of Computer Applications (0975 – 8887) Volume 87 – No.19, February 2014
6. Jiao Yabing:" Research of an Improved Apriori Algorithm in DataMining Association Rules" International Journal of Computer and Communication Engineering, Vol. 2, No. 1, January 2013
7. Thabet Slimani, Amor Lazzez:" Efficient Analysis of Pattern and Association Rule Mining Approaches" Efficient Analysis of Pattern and Association Rule Mining Approaches
8. Paresh Tanna#1, Dr. Yogesh Ghodasara\*2:" Using Apriori with WEKA for Frequent Pattern Mining" International Journal of Engineering Trends and Technology (IJETT) – Volume 12 Number 3 - Jun 2014
9. Heydar Jafarzadeh, Rouhollah Rahmati Torkashvand, Chamran Asgari , and Amir Amiry3:" Provide A New Approach for Mining Fuzzy Association Rules using Apriori Algorithm" Indian Journal of Science and Technology, Vol 8(S7), 127–134, April 2015 ISSN (Online) : 0974-5645 DOI: 10.17485/ijst/2015/v8iS7/71227
- 10.



# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor & UGC Approved Journal)

Website: [www.ijirset.com](http://www.ijirset.com)

Vol. 6, Issue 8, August 2017

10. Jaishree Singh, Hari Ram, Dr. J.S. Sodhi: "improving efficiency of Apriori algorithm using transaction reduction" International Journal of Scientific and Research Publications, Volume 3, Issue 1, January 2013 1 ISSN 2250-3153
11. Marghny H. Mohamed • Mohammed M. Darwieesh: "Efficient mining frequent itemsets algorithms" DOI 10.1007/s13042-013-0172-6 Xueyan Lin: "MR-Apriori: Association Rules Algorithm Based on MapReduce" 978-1-4799-3279-5 /14/\$31.00 ©2014 IEEE
12. Akihiro Inokuchi, Takashi Washio and Hiroshi Motoda: "An Apriori-based Algorithm for Mining Frequent Substructures from Graph Data". 13-16, 2000.
13. Zachary K. Baker and Viktor K. Prasanna: "Efficient Hardware Data Mining with the Apriori Algorithm on FPGAs" 1 Supported by the United States National Science Foundation/ITR under award No. ACI-0325409 and in part by an equipment grant from the Xilinx and HP Corporations
14. [https://www.cs.indiana.edu/~predrag/classes/2010springi211/week\\_13](https://www.cs.indiana.edu/~predrag/classes/2010springi211/week_13).
15. <http://blog.hackerearth.com/beginners-tutorial-apriori-algorithm-data-mining-r-implementation>
16. R.Thanigaivel and K.Ramesh Kumar: "Boosted Apriori: an effective data mining association rules for heart disease prediction system" Middle east journal of scientific research: 192-200, ISSN 1990-9233, 2016.
17. Xiang cheng, sensu, shen zhi xu, zhengyi li: "DP-Apriori: a differentially private frequent itemset mining algorithm based on transaction splitting"
18. D. Ashok Kumar, M. C. Loraine Charlet Annie: "Frequent Itemset Mining of Market Basket Data using K-Apriori Algorithm" International Journal of Computational Intelligence and Informatics, Vol. 1 : No. 1, April – June 2011.
19. [ieeexplore.ieee.org](http://ieeexplore.ieee.org)
20. Shubhangi, D.patil, Ratnadeep, R.Deshmukh, D.k.Kirange: "Adaptive Apriori algorithm for frequent itemset mining" International conference on System modeling and advancement in research trends (SMART), 16808058.
21. Neyla cherifa benhamouda, Habiba Drias, Celia Hierche: "Meta-Apriori: A new algorithm for frequent pattern detection" Asian conference on intelligent information and database systems.
22. K.Prassana, M.seetha, A.P.SivaKumar: "CApriori: Conviction based Apriori algorithm for discovering frequent determinant patterns from high dimensional datasets".
23. Sallam osman fageeri, Rhiza ahmad, baharum B. Baharudin: "A semi-Apriori algorithm for discovering the frequent itemsets".
24. Thi Thanh Nhan Le, Thi Thanh Thuy Nguyen, Tae4 chong Chung: "BitApriori: An Apriori-Based frequent itemsets mining using bit streams".