

## Study of various Improved Apriori Algorithms

Deepali Bhende<sup>1</sup>, Usha kosarker<sup>2</sup>, Mnisha Gedam<sup>3</sup>

<sup>1</sup>(deepalibhende80@gmail.com, Computer Science, RTM Nagpur University, India)

<sup>2</sup>(usha.kosarker@raisoni.net, Computer Science, RTM Nagpur University, India)

<sup>3</sup>(manisha.gedam@raisoni.net, Computer Science, RTM Nagpur University, India)

---

**Abstract:** The Apriori algorithm is a popular and a classical algorithm in data mining. The main idea of this approach is to find a useful pattern in various sets of data. The algorithm suffers from many drawbacks. This paper deals with the apriori algorithm, and various techniques that were proposed to improve the apriori algorithm. The paper discusses about various approaches use to overcome the drawback of the apriori algorithm as to improve its efficiency.

**Keywords** - Ariori algorithm ,frequent pattern, association rule mining.

---

### I. INTRODUCTION

Association rules problems were first brought out by Agrawal and others in 1993, which were researched by many other researchers after that. They optimized the original algorithm, such as bringing in random sampling, parallel thoughts, adding reference point, declining rules, changing storing framework, etc. Those works were aimed at improving the efficiency of algorithm rules, spreading the applications of association rules from initial business direction to other fields, such as education, scientific research, medicine, etc. [1] Association rules mining is to discover the associations and relations among item sets of large data. Association rules mining is an important branch of data mining research, and association rules is the most typical style of data mining. Presently, association rules mining problems are highly valued by the researchers in database, artificial intelligence, statistic, information retrieval, visible, information science, and many other fields. Many incredible results have been found out. What can efficiently catch the important relationships among data are simple forms of association rules and easily to explanation and understanding. Mining association rules problems from large database has become the most mature, important, and active research contents. Association rules mining is to discover the associations and relations among item sets of large data. Association rules mining is an important branch of data mining research, and association rules is the most typical style of data mining. Presently, association rules mining problems are highly valued by the researchers in database, artificial intelligence, statistic, information retrieval, visible, information science, and many other fields. Many incredible results have been found out. What can efficiently catch the important relationships among data are simple forms of association rules and easily to explanation and understanding. Mining association rules problems from large database has become the most mature, important, and active research contents.

### II. ASSOCIATION RULE MINING

Before discussing apriori algorithm, it is necessary to have a look on association rule mining. Data mining has so many techniques, among all association rule is considered as most important and useful technique. It is used to discover the frequently occurring patterns in the database. It helps in discovering the important correlations in the database [4]. Association rule mining has many applications and is best known for decision making and constructive marketing. Association rule can be best explained by this example. If customer buys a shampoo then he may also buy a conditioner. It will help in suspecting the buying behavior of the customers. This can be used as a information which will be helpful in taking important decisions for marketing purposes. Association rule are considered interesting if they are able to satisfy both minimum support threshold and minimum confidence threshold. There are many application domains, in which association rules are used. Some of them are:[5][6].

- Knowledge extraction from software engineering metrics.
- Telecommunication networks
- Supermarket data management.
- Finding of patterns in biological fields.
- Market basket analysis

Consider the following example:

Sample Database

Tid	Items Purchased
1	Shampoo
2	Shampoo, Soap, Paste, Face wash
3	Conditioner, Soap, Paste, Oil
4	Shampoo, Conditioner, Soap, Paste
5	Shampoo, Conditioner, Soap, Oil

Table : 1

The association rules for the above data:

- {Soap}->Paste
- {Shampoo, conditioner}->{Face Wash, Oil}
- {Shampoo, paste}-> {Conditioner}

Thus interesting patterns can be revealed, which are very beneficial, using association rules.

Some common terminologies which are used in algorithm are:

- Itemset- It is the collection of itemsets in the database.
- Transaction- It is database entry which contains collection of items. It is denoted by T.
- Minimum set- This condition should be satisfied by the items. It helps in removing the in-frequent items. Candidate set- The only items which are considered for processing.
- Frequent Itemsets- The items which are frequently occurring, satisfies minimum support condition.
- Support- Suppose we are having two items X and Y, then support is a transaction that contains both X and Y. Confidence- Measures how often items in Y appear in transactions that contain X [77][88].

### III. APRIORI ALGORITHM

Apriori algorithm is given by Agrawal. It is used to generate frequent itemsets from the database. The Apriori algorithm uses the Apriori principle, which says that the item set I containing item set (say) X is never large if item set X is not large [1][7] or All the non empty subset of frequent item set must be frequent also.

Notations Being Used In Apriori Algorithm

K	itemset Any itemset which consist of k items.
Ck	Set of Candidate k itemsets
Lk	Set of large k itemsets (frequent k itemsets).

Table : 2

These itemsets are derived for the candidate itemsets in each pass. Based on this principle, the Apriori algorithm generates a set of candidate item sets whose lengths are (k+1) from the large k item sets and prune those candidates, which does not contain large subset. Then, for the rest candidates, only those candidates that satisfy the minimum support threshold (decided previously by the user) are taken to be large (k+1)- item sets. The Apriori generate item sets by using only the large item sets found in the previous pass, without considering the transactions.

Steps involved are:

1. Generate the candidate 1-itemsets (C1) and write their support counts during the first scan.
2. Find the large 1-itemsets (L1) from C1 by eliminating all those candidates which does not satisfy the support criteria.
3. Join the L1 to form C2 and use Apriori principle and repeat until no frequent itemset is found.

#### **Drawbacks of Apriori algorithm :**

No doubt apriori algorithm is considered as the most beneficial and best algorithm for generating association rules, it too has some drawbacks. Some of these are listed below:

1. It takes time to scan the database.

2. There is a need of several iterations for mining of data.
3. Large numbers of in-frequent itemsets are generated and thus increase the space complexity.
4. More search space is required and I/O cost will be increased.

#### IV. IMPROVEMENTS IN APRIORI ALGORITHM

Association Rule Mining has attracted a lot of intention in research area of Data Mining and generation of association rules is completely dependent on finding Frequent Item sets. Various algorithms are available for this purpose.

#### Comparison Of Improved Versions Of Apriori Algorithm

Authors	Technique	Benefit
Suhani Nagpal	-Temporary Tables for scanning., - Logarithmic Decoding	-Low system overhead and good operating performance [4]. -Efficiency higher than Apriori Algorithm.
Jaishree Singh, Hari Ram	Variable Size Of Transaction on the basis of which Transactions are reduced.	- Reduces the I/O cost. - Reduce the size of Candidate Item sets (Ck) [9].
Jaio Yabing	-Double Pruning method is used. -States that before Ck come out, Prune Lk-1	For large datasets, it saves time and cost and increases the efficiency [8].
Sunil Kumar	-Probability Matrix has been used. -Uses Bottom Up approach.	Reduced Execution time than Apriori Algorithm [5].
Kiran R. U., and Reddy P. K	-Specify multiple minimum supports to reflect the natures of the items and their varied frequencies in the database called as minimum item supports (MIS)	Different support equirements for different rules can be expresses effectively[21].
J. S. Park, M. S. Chen, and P. S. Y	Combines the Apriori algorithm and FP tree structure of FP-growth algorithm	-It doesn't generate conditional & sub conditional patterns of the tree recursively - It works faster than Apriori and almost as fast as FP-growth. [22]
Sujatha Dandu, B.L.Deekshatulu & Priti Chandra	Modify the APFT to include correlated items & trim the non correlated itemsets	-optimizes the FP-tree & removes loosely associated items from the frequent itemsets[24].

Table : 3

#### V CONCLUSION

Association rule mining is used to discover the frequently occurring patterns in the database. Apriori algorithm can be considered as one of the oldest algorithm in the field of association rule mining. This paper includes a brief overview of apriori algorithm and recent improvements done in the area of apriori algorithm. With the survey on various improved algorithms, it is concluded that the main focus is to generate less candidate sets which contains frequent items within a reasonable amount of time. Also, in future some more algorithms can be developed that requires only single scan for the database and are efficient for large databases.

#### REFERENCES

- [1] Ms. Rina Raval, Prof. Indr Jeet Rajput , Prof. Vinitkumar Gupta, "Survey on several improved Apriori algorithms", IOSR Journal of Computer Engineering (IOSR-JCE), Volume 9, Issue 4 (Mar. - Apr. 2013), PP 57-61
- [2] Reeti Trikha, Jasmeet Singh, "improving the efficiency of apriori algorithm by adding new parameters", International Journal for Multi Disciplinary Engineering and Business Management, Volume-2, Issue-2, June-2014
- [3] Mohammed Al-Maolegi, Bassam Arkok, "An improved apriori algorithm for association rules", International Journal on Natural Language Computing (IJNLC) Vol. 3, No.1, February 2014
- [4] Arti Rathod, Mr. Ajaysingh Dhabariya, Chintan Thacker, "A Survey on Association Rule Mining for Market Basket Analysis and Apriori Algorithm", International Journal of Research in Advent Technology, Vol.2, No.3, March 2014

- [5] Pratibha Mandave, "Data mining using Association rule based on APRIORI algorithm and improved approach with illustration" International Journal of Latest Trends in Engineering and Technology (IJLTET), ISSN: 2278-621X, Vol. 3 Issue2 November 2013
- [6] Ila Chandrakar, A. Mari Kirthima, "A Survey on Association Rule Mining Algorithms", international journal of mathematics and computer research, vol 1, issue 10, nov 2013
- [7] Charanjeet Kaur, "Association Rule Mining using Apriori Algorithm: A Survey", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), Volume 2, Issue 6, June 2013
- [8] Pranay bhandari, K.Rajeswari, Swati Tonge, Mahadev Shindalkar, "improved apriori algorithms- A survey", International Journal of Advanced Computational Engineering and Networking, ISSN (p): 2320-2106, Volume-1, Issue-2, April-2013
- [9] Sheila A. Abaya, "Association Rule Mining based on Apriori Algorithm in Minimizing Candidate Generation", :International Journal of Scientific & Engineering Research Volume 3, Issue 7, July-2012
- [10] Yanfei Zhou, Wanggen Wan, Junwei Liu, Long Cai, "Mining Association Rules Based on an Improved Apriori Algorithm", 978- 1-4244-585 8- 5/10/\$26.00 ©2010 IEEE.
- [11] Mamta Dhanda, Sonali Guglani , Gaurav Gupta, "Mining Efficient Association Rules Through Apriori Algorithm Using Attributes", International Journal of Computer Science and Technology Vol 2, Issue 3, September 2011, ISSN:0976-8491
- [12] Shuo Yang, "Research and Application of Improved Apriori Algorithm to Electronic Commerce" 2012 11th International Symposium on Distributed Computing and Applications to Business, Engineering & Science, 978-0-7695-4818-0/12, IEEE DOI 10.1109/DCABES.2012.51
- [13] 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
- [14] Jiao Yabing "Research of an Improved Apriori Algorithm in Data Mining Association Rules", International Journal of Computer and Communication Engineering, Vol. 2, No. 1, January 2013(1)
- [15] J. Han and M. Kamber, *Conception and Technology of Data Mining*, Beijing: China Machine Press, 2007.
- [16] J. N. Wong, translated, *Tutorials of Data Mining*. Beijing. TsinghuaUniversity Press, 2003.
- [17] Y. Yuan, C. Yang, Y. Huang, and D. Mining, *And the OptimizationTechnology nd Its Application*. Beijing. Science Press, 2007.
- [18] Y. S. Kon and N. Rounteren, "Rare association rule mining andknowledge discovery:technologies for frequent and critical event detection" PA: *Information Science Reference*, 2010
- [19] W. Sun, M. Pan, and Y. Qiang, "Improved association rule mining method based on t statistical," *Application Research of Computers*. vol.28, no. 6, pp. 2073-2076, June, 2011.
- [20] C. Wang, R. Li, and M. Fan, "Mining Positively Correlated FrequentItemsets," *Computer Applications*, vol. 27, pp. 108-109, 2007
- [21] Kiran R. U., and Reddy P. K.: An Improved Multiple Minimum Support Based Approach to Mine Rare Association Rules.
- [22] J. Pei, J. Han, and H. Lu. Hmine: Hyper-structure mining of frequent patterns in large databases. In ICDM, 2001, pp441-448.
- [23] J. S. Park, M. S. Chen, and P. S. Yu. An effective hashbased algorithm for mining association rules. Proceedings of ACM SIGMOD International Conference on Management of Data, San Jose, CA, 1995, pp175-186
- [24] Sujatha Dandu, B.L.Deekshatulu & Priti Chandra "Improved Algorithm for Frequent Item sets Mining Based on Apriori and FP-Tree" Global Journal of Computer Science and Technology Software & Data Engineering Volume 13 Issue 2 Version 1.0 2013 ISSN: 0975-4172