



International Journal of Emerging Technologies in Computational and Applied Sciences (IJETCAS)

www.iasir.net

Review of Apriori Algorithm and its Recent Improvements

Jagmeet Kaur¹, Neena Madan²

¹Student Department of Computer Science, ²Lecturer Department of Computer Science
Guru Nanak Dev University, Regional Campus
Ladhewali Road, Jalandhar, INDIA

Abstract: Apriori algorithm is one of the most important and widely used algorithms in the association rule mining. This algorithm is used to find out frequent occurring patterns in a dataset. It follows a level-wise search. Till now apriori has been considered as the best algorithm for association rule generation. No doubt this algorithm is considered as efficient algorithm, but still it has many drawbacks. To improve efficiency of apriori algorithm, many researchers have done improvements in apriori. This paper gives a theoretical survey on apriori algorithm. A brief overview of the recent improvements in apriori algorithm by many researchers is also presented.

Keywords: Association Rule Mining, Minimum Support, Confidence, Item set, frequent item, Transaction.

I. Introduction

With the growth of science and technology, data volume has also increased to a greater extent. To manage such large data is a difficult task. Data mining, with its efficient rules has made it easy to extract such large data. Thus, data mining is a process of extracting useful information from large dataset. It is also called as knowledge discovery process. Processing of data is slower than its speed at which it is generated, so often it remains concealed. Availability of huge data is useless until it is managed properly. So to handle this large set of data in any form is a difficult task. Hence there is a need of technique that can provide some useful information or knowledge from raw data. In such a competitive world every field needs to be up to date and for that they require a technique which helps them to store their data in such a way, that it can be easily retrieved and updated. Data mining is that technique which helps to retrieve knowledge from such abundant data. Data mining includes many rules for extracting data and these include association rule, clustering etc. Data mining has its applications in many fields. Some of them are science, engineering, medical, education, banking telecommunication etc [1][2][3].

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.
- Supermarket data management.
- Finding of patterns in biological fields.
- Telecommunication networks
- Market basket analysis

Consider the following example:

Table 1 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

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 [7][8].

III. Apriori Algorithm

Association rule mining involves many algorithms and methods for generating rules. These include apriori algorithm, FP- growth algorithm, genetic algorithm etc. Till now apriori has been considered as the best algorithm for association rule generation. It was proposed by R.Agrawal and R.Srikant in 1994. Apriori is helpful in discovering frequently occurring itemsets in the database. The basic idea behind apriori is to make multiple passes over the database. Apriori algorithm follows a property, "All nonempty subsets of a frequent itemsets must be frequent".

The Apriori algorithm involves two steps to find out frequent itemsets.

a) Self join and b) Pruning

In apriori, an iterative approach breadth-first approach is used. First of all the set of frequent 1-itemsets is found.

The main purpose is to traverse the database completely, so that frequent itemsets can be found.

Steps of apriori algorithm are as follows:

- a) There will be a support threshold value 's'. In the very first pass, the items that satisfy the given threshold are found. These are named as L1, frequent item set.
- b) The pairs of this item set will be considered as candidate pair C2 for second pass. The pairs in C2 which support s are frequent pairs, L2.
- c) In the third pass, count the occurrence of triples in C3 and find the frequent triples L3. The candidate triples, C3 are those sets which have their sets in L2.
- d) The steps continues until we find only frequent item sset [2][8].

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

Many researchers have done the modifications to improve the efficiency of apriori algorithm. Some of the recent improvements in apriori algorithm are described below.

A. Improvement based on set size frequency

A modified algorithm introduces issues such as size and set size frequency to remove non significant candidate keys. It will help to remove candidate keys in a better way. The improved algorithm for Apriori takes for the set size which is the number of items per transaction and set size frequency which is the number of transactions that have at least set size items. Database is given with set size initially and set size frequency of the initial database is for second database. The items with less frequency then minimum support value initially are removed and initial set size is determined to get the highest set size whose frequency is greater than or equal to minimum support of set size. And thus, set sizes which are not equal to greater than minimum set size support eliminated [9].

B. Improvement by reducing redundant operation

An improved algorithm removes some defaults such as taking more time to generate candidate itemsets, repeatedly scanning of database. The new improved apriori algorithm consists of three segments: first of all, number of judgements during the time of generating frequent candidate itemsets is decreased. Secondly, there is pruning of frequent itemsets. And at the last the database is optimized. It was found out that improved apriori algorithm increases efficiency, improves performance and reduction of redundant operation when frequent itemsets are produced [10].

C. Utilization of resources

Algorithm treats all the itemsets equally whether present or absent, importance of a item to user or business perspective is not considered. This new improved apriori algorithm illustrates that such disadvantages can be removed by using attributes like quantity, profit, frequency of items and thus helps business and user [11].

D. Improvements based on customer habits

More time is consumed by traditional apriori algorithm to scan the database. The proposed algorithm helps to reduce database access by taking in account customers. Relative theorems are used to find frequent itemsets [12].

E. Improvement by reducing transactions and memory utilizations

The classical algorithm scans the data so many times. And if the database is large, then scan time is relatively high. The improved apriori algorithm is able to reduce the scan time, by reducing the number of transactions, those transactions that have unnecessary records.

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.

VI. 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", 978-1-4244-5858-5/10/\$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