

# Review on Association Rule Mining: A Survey

Vintee Chaudhary

M.Tech(CS&E)

School of Computing Science and Engineering  
Galgotias University, Greater Noida, U.P .

Aparna Choudhary

M.Tech(CS&E)

School of Computing Science and Engineering  
Galgotias University, Greater Noida, U.P.

Nawagata Nilambari

M.Tech(CS&E)

School of Computing Science and Engineering  
Galgotias University, Greater Noida, U.P.

Manika Tyagi

M.Tech(CS&E)

School of Computing Science and Engineering  
Galgotias University, Greater Noida, U.P.

**Abstract:** Association rule mining plays important role in the field of data mining. Association rule mining is a technique that helps to prepare the way to improve the mining technique. It is a method to discover relationships among variables in the database. Association rule basically divided into two different parts (a) an antecedent and (b) a consequent. In association rule mining different types of approaches and algorithms have been designed but it is very important to know which approach is best and suitable for association rule. So In this paper, we present a complete survey on different algorithms and approaches used in association rule mining in different domain.

**Keywords—** Association rules, Apriori, Confidence, Support, Frequent itemsets, Minimum support.

## 1. INTRODUCTION

Data mining is used to extract potentially useful, meaningful and valid information from large databases. Data mining plays important role in analyzing large database and extract relevant information from a data repository. Database, data warehouse, the other information repository is a data source .Association rule is used for mining data for frequent itemsets. Association can also define as finding frequent patterns. Frequent pattern is item sets which occur frequently in the data.

There are various data mining technology which is link mining, clustering, statistical learning, association rule mining and classification .Using data mining technology data analyzing and pattern extraction is accomplished. All of these have their importance in development and data research. GUHA procedure, FP-growth algorithm Eclat algorithm, ASSOC, OPUS search are algorithm which is used for association rule mining. But APRIORI is the best algorithm to generate association rule[1].

### 1.1 Association Rule Mining

Association rule can be defined as the relationship between items with confidence and support. Association rule can be represented as  $A \rightarrow C$  in which 'A' is an antecedent and 'C' is consequent .It follows if and then statement where the antecedent is if and consequent is then [2].

Before discussing association rule it is necessary to define APRIORI algorithm. There are some terms which help to understand the concept of APRIORI algorithm. These terms describe below:

#### Itemsets:

Item sets are a group of items together in a single transaction. It is the accumulation of item sets. If database contains 'n' items than there are  $2^n$  item sets

#### Support:

Support is defined as no. of transaction containing that item. Support of rule can be defined as no. of transaction containing both antecedent and consequent. Support of ' $A \rightarrow C$ ' is defined as No. of transaction that contain both A and C.

#### Confidence:

Confidence of rule is defined as no. of transaction containing both antecedent and consequent divided by No .of transaction containing antecedent. Support of ' $A \rightarrow C$ ' is No. of transaction that contain both A and C divided by No. of transaction containing A.

#### Frequent itemsets:

If support of itemsets is greater than or equal to minimum support then it is called frequent item sets otherwise item sets are infrequent.

### 1.2 APRIORI Algorithm

APRIORI algorithm is a classical algorithm. APRIORI algorithm is used to find frequent itemsets from large databases. APRIORI algorithm has two steps which are

- Join step
- Prune step

#### Join step

To find  $L_k$ , joining  $L_{k-1}$  with itself to generate  $C_k$ .  $C_k$  is collection of candidate item set of size  $k$ .  $L_k$  is collection of frequent item sets of size  $k$ .

#### Prune step

In this step find all frequent item sets.  $L_k$  is the subset of  $C_k$  but it is not necessary that all items in  $C_k$  are frequent. To determine the support count of each candidate in  $C_k$  scans the database. If support count of item sets is greater than or equal to minimum support then item sets are frequent which belongs to  $L_k$ . And if support count of item sets is not greater than or equal to minimum support then it is not frequent. After this item set which are not frequent eliminates from  $C_k$ . This process continues until all frequent items occur in  $L_k$ [3].

Example of candidate itemset and frequent itemsets generation with minimum support is 2. The transaction database is given below.

TID	Items
T1	ABC
T2	BD
T3	BC
T4	ABD
T5	AC
T6	BC
T7	AC
T8	ABCE
T9	ABC
T10	F

#### Candidate 1 itemsets ( $C_1$ )

Itemsets	Support Count
A	6
B	7
C	7
D	2
E	1
F	1

#### Frequent 1 itemsets ( $L_1$ )

Itemsets	Support Count
A	6
B	7
C	7
D	2

#### Candidate 2 itemsets ( $C_2$ )

Itemsets	Support Count
AB	4
AC	5
AD	1
BC	5
BD	2
CD	0

#### Frequent 2 itemsets ( $L_2$ )

Itemsets	Support Count
AB	4
AC	5
BC	5
BD	2

#### Candidate 3 itemsets ( $C_3$ )

Itemsets	Support Count
ABC	3
ABD	1
BCD	1

#### Frequent 3 itemsets ( $L_3$ )

Itemsets	Support Count
ABC	3

ABC is frequent itemsets. These frequent itemsets is used to generate association rules that satisfied minimum support and minimum confidence criteria.

## 2. LITRETURE SURVEY

In this section, we discussed on some approaches and algorithms used in association rule mining.

*2.1. An Approach to Improve Apriori Algorithm Based On Association rule Mining* [1]. proposed a new algorithm, improve APRIORI algorithm which is used to reduce the limitation of APRIORI algorithm. Task is divided into two parts. In a first part eliminates bad or duplicate data from database to improve database consistency and in second part proposed algorithm applied to filtered datasets. Proposed algorithm computes support count of candidate itemsets like a APRIORI algorithm compute. It generates frequent itemsets, candidate itemsets. Improve algorithm applied to the dataset and eliminate infrequent itemsets from each transaction. Horizontal partitioning also applied to datasets. It takes less time to generate association rule and generate fewer rule as compared to an original APRIORI algorithm.

*2.2. Mining Association Rule from Data with Missing Values by Database Partition and Boolean Matrix*[4]. elaborates a Database Partitioning & Boolean Matrix. This algorithm applied on data which have missing values for mining association rule. To generate frequent itemsets very quickly, proposed algorithm used Boolean matrix and conjunction operation. For mining association rule from large dataset it takes less memory space and time.

*2.3. An Improved Tree Algorithm for Association Rule Mining Using Transaction Reduction* [5]. presents a new algorithm which is used to overcome the limitation of APRIORI algorithm. This proposed algorithm is used to increase the execution time of processing and reduce candidate item sets. Hash table used to reduce multiple time scanning of the database.

*2.4. An Integrated Approach to Derive Effective Rules from Association Rule Mining Using Genetic algorithm* [6]. proposed an algorithm which is the combination of APRIORI algorithm and Genetic algorithm. The Proposed algorithm is used to overcome the drawback of classical algorithm. Using the classical algorithm, find frequent item sets with single minimum support but using proposed algorithm, multiple minimum support is consider to find frequent item sets. APRIORI algorithm with multiple minimum supports is used to find frequent item set and generate association rules. To find reduced set of association rule, genetic algorithm is used. Lift factor is also used to find a strong association rule. It generates less association rule.

*2.5. A Parallel Algorithm of Association Rules Based on Cloud Computing* [7]. new method parallel association rule mining based on cloud computing is proposed. To reduce inter processor communication and I/P overhead. Proposed algorithm takes up the separation strategy which is used to visit a local database only once. A framework for a cloud computing platform system are proposed. With the cloud environment this framework helps to use cloud nodes

to protect data privacy. A PFP(Parallel Frequent Pattern) growth model is used to mine maximal frequent itemsets and frequent closed itemsets. For cloud computing platform, a cloud based PFP growth model is combined with Map Reduce Model. Meanwhile Map Reduce Model is used to mine data from the database and build local FP tree to improve efficiency of data mining and reduce communication overhead and offer good speed up.

*2.6. An Improved Algorithm for Mining Association Rules in Large Database*[8]. elaborates a novel approach for association rule mining which can be used for generating association rule effectively for large database. To reduce the execution time of algorithm, features of items and weight of candidate item sets are used. Transaction database is transformed into features matrix. Transformation is used to reduce the no. of I/P access and speeding up mining process. In this proposed algorithm leverage measures is used to find interestingness of rule. Leverage measure is used to reduce the candidate item sets generation and reduce the memory space requirement which is used to store useless candidate item sets.

*2.7. Hybrid Association-Classification Algorithm for Anomaly Extraction* [9]. A hybrid algorithm which is the combination of fuzzy algorithm and a classification algorithm that is used to generate an association rule. Frequent item sets are generated using a fuzzy APRIORI algorithm and then CART algorithm is applied on frequent item sets to generate lesser rules and for better network anomalies. It takes less time and give less error rate and also provide higher accuracy.

*2.8. Association Rule Mining based on APRIORI Algorithm in Minimizing Candidate Generation* [10]. To improve the APRIORI algorithm, two factors which are set size and set frequency are added in APRIORI algorithm where set size is no. of items per transaction and set frequency is the no. of transaction that contain at least a set size items. Set size and set frequency are used to remove useless candidate generation. Using these factors, the proposed algorithm minimized candidate generation in an efficient and effective way. These factors helped to generate frequent item sets very quickly. It takes less execution time as compared to APRIORI algorithm.

*2.9. MRA for Association Rule Mining used for Cooperative Learning* [11]. APRIORI algorithm with multiple level used to generate frequent item sets. The Proposed algorithm is called Multilevel Relationship Algorithm(MRA). Multilevel APRIORI algorithm and Bayesian Probability is combined together to generate an association rule. The new algorithm is applied to various datasets and find frequent item sets and external

dependencies. In this proposed work generate association rule more efficiently.

*2.10. Programming Parallel Apriori Algorithm for Mining Association Rules*[12]. There are various parallel programming languages that provide low level construct but some difficulties occur when implement, design, debug, and maintain these programming languages. A parallel programming using a sequential programming language helps the programmer to debug, design and implement. The parallel paradigm gives high level language construct to implement distributed and parallel algorithm. To avoid that processor's speed is not slow down when generating frequent item sets, dataset and item set is partitioned into different parts and give to the processor. This proposed algorithm improves the efficiency of original APRIORI algorithms.

*2.11. Optimizing network traffic by generating association rules using Hybrid Apriori-Genetic algorithm* [2]. APRIORI and Genetic algorithm are used This algorithm applied on network traffic data sets. It generates less frequent item sets and takes less computational time. Then it generates less rules for the network traffic as compared to an original APRIORI algorithm.

*2.12. An Efficient Approach Using Rule Induction and Association Rule Mining Algorithm in Data Mining* [13].

Rule induction algorithm helps to find the best result. To cover large data from the dataset, it used Decision List Induction algorithm. It used Induction Rule algorithm to generate association rule with less error rate. Then find a reduced set of rule.

*2.13. An Improved APRIORI Algorithm for Association Rules* [14]. Improve APRIORI algorithm is used to reduce the limitation of the original APRIORI algorithm. The proposed algorithm reduces the wasting time which is used to scan the whole database for generating frequent itemsets. It scans only some transaction. This algorithm generates association rules efficiently with less time.

*2.14. A Study on Various Data Mining Approaches of Association Rules*[15]. To remove the replicas from data sets, preprocessing is applied on datasets to find desired data set. From desired data set, candidate item sets are found. For finding frequent item sets, candidate item sets is scanned. Strong Association rules are generated from frequent item sets. For time complexity and space complexity, Hash tree and Hash Mapping table is used.

### 3. CONCLUSION AND FUTURE WORK

Association Rule Mining is a good topic for the research in data mining. In this paper we present the complete survey of Association Rule Mining algorithms

and approaches. From this survey we found that there are various new and improved algorithms which are used for generating association rule. But there are some problem occurs for generating association rules that are solved by data mining researchers in the upcoming times.

### REFERENCES

- [1] Chanchal Yadav, Shuliang Wang and Manoj Kumar, "An Approach to Improve Apriori Algorithm Based On Association rule Mining". 4th ICCNT - 2013 July 4-6, 2013, Tiruchengode, India
- [2] Surendra Kumar Chadokar, Divakar Singh and Anju Singh "Optimizing network traffic by generating association rules using Hybrid Apriori-Genetic algorithm" 2013 IEEE.
- [3] Mrs. R. Sumithra and Sujni Paul. "Using distributed apriori association rule and classical apriori mining algorithms for grid based knowledge discovery" 2010 Second International conference on Computing, Communication and Networking Technologies, IEEE, 2010.
- [4] Jayanti Dansana, Shailender Kumar, Abhilash Kumar Srivastava, "Mining Association Rule from Data with Missing Values by Database Partition and Boolean Matrix", IEEE - International Conference on Research and Development Prospects on Engineering and Technology (ICRDPET 2013) Vol.5 March 29, 30 - 2013 .
- [5] Krishna Balan, Karthiga and Sakthi Priya, "An Improved Tree Algorithm for Association Rule Mining Using Transaction Reduction", International Journal of Computer Application Technology and Research, Volume-2, Issue2, 166-169, 2013.
- [6] Kannika Nirai M and E Ramaraj, "An Integrated Approach to Derive effective rules from Association Rule Mining using Genetic Algorithm", 2013 International Confrence on Pattern Recognition, Informatics and Mobile Engineering (PRIME) February 21-22.
- [7] Wang Yong, Zhang Zhe and Wang Fang, "A Parallel Algorithm of Association Rules Based on Cloud Computing", 2013 8<sup>th</sup> International Conference on Communication and Networking in China (CHINACOM).
- [8] Farah Hanna AL-Zawaidah, Yosef Hasan Jabara and Marwan AL-Abed Abu-Zanona, "An Improved Algorithm for Mining Association Rules in Large Database", World of Computer Science and Information Technology Journal(WCSIT) ISSN: 2221-0741 Vol.1, No. 7, 311-316, 2011.
- [9] Gaurav Shelke, Anurag Jain and Shubha Dubey, "Hybrid Association-Classification Algorithm for Anomaly Extraction", 4<sup>th</sup> ICCNT-2013 July 4-6 2013 Tiruchengode, India.
- [10] 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 ISSN 2229-5518.
- [11] Deepak A Vidhate and Parag Kulkarni, "Multilevel Relationship Algorithm for Association Rule Mining used for Cooperative Learning", International Journal of Computer Applicatio(0975-8887), Volume 86, No. 4, January 2014.
- [12] Chia -Chu Chiang, "Programming Parallel Apriori Algorithm for Mining Association Rules", 2010 International Conference on System Science And Engineering.
- [13] Kapil Sharma and Sheveta Vashisht, "An Efficient Approach Using Rule Induction and Association Rule Mining Algorithms in Data Mining", Graduate Research in Engineering and Technology: An International Journal ISSN 2320-6632, Volume-1, Issue-2, 2013.
- [14] Hassan M.Najadat, Mohammed AL-Maolegi, Bassam Arkok, "An Improved Apriori Algorithm For Association Rules", International Research Journal of Computer Science and Application Vol.1, No.1, June 2013, PP: 01-08.
- [15] Rachna Somkunwar, "A Study on Various Data Mining Approaches of Association Rules", International Journal of Advanced Research in Computer Science and Software Engineering Volume.2, Issue9, September 2012, pp.141-144.