
“STUDY OF ASSOCIATION RULE MINING TECHNIQUES AND ITS APPLICATION”

¹TRIVENI MANKAR

M. Tech Scholar, Bhabha Engineering Research Institute, Bhopal, India
triveni_mankar09@yahoo.com

²DR. TRIPTI ARJARIYA

Bhabha Engineering Research Institute, Bhopal, India
tripti.beri@gmail.com

ABSTRACT: *With the development of computer network technology, the demand for information storage expansion, construction and security access to the index database gradually into people's vision, the database access is based on mining association rules in large data on the characteristics through the association rules mining user interest value, and according to the large database optimization index association rules features, improve data retrieval ability. Data mining association rules are adopted to extract useful features in data stream information, key information data in the estimation of parameters, data acquisition estimation, the process of mining association rules is parallel scheduling and grid multi thread processing data in a distributed cloud computing environment, reflect the correlation dimension feature by feature data extraction, realize the integration of data analysis and data processing [1]. This paper aims at discussing the process carried out in the association rule mining techniques also the applications of the algorithm.*

Keywords: association rules, Data mining, optimization, apriori, data acquisition, distributed cloud computing

1. INTRODUCTION

Data Mining is the process of discovering knowledge. It is the process of extracting information from available raw data. The data are stored in databases. There are various kinds of data that can be used in data mining which includes transactional data, statistical data etc. Data mining includes various techniques for each purpose. Techniques include Association rule mining, classification and prediction, regression etc. Association rule mining techniques are widely used in discovering hidden correlations and relationship between set of items in a transaction. It includes every transaction in the database during the discovery process. It also reveals the set of strategies that can be followed or neglected in the field for respective development.

Association rule mining is a technique has been widely used in data mining and it is doing well in the industry. But as the amount of data increase, new problems arise in the implementation of this method. For example: The available rule miners can only mine rules from one file or table. While it is often required to combine data from more than one data source; therefore; it is important to develop miner having the ability to connect many files or tables to generate a suitable data set for mining [3].

The available rule miners do not split the two regular phases: the frequent item generation and the rule generation and make one run for the mining. The separation of these two phases can reduce work time significantly by working at dead times in many cases. Because of the pruning process during the rule generation, there is no chance to find weak association rules instead of strong ones. Weak association rules are helpful in many applications.

2. ASSOCIATION RULE MINING

Association rule mining is a technique is to find frequent patterns, correlations, associations, or causal structures from data sets found in various kinds of large set of databases such as relational databases, transactional databases, and other forms of data repositories. Given a set of transactions, association rule mining aims to find the rules which enable us to predict the occurrence of a specific item based on the occurrences of the other items in the transaction.

Association rules are created by thoroughly analyzing data and looking for frequent if/then patterns. Then, depending on the following two parameters, the important relationships are observed

Support: Support indicates how frequently the if/then relationship appears in the database.

Confidence: Confidence tells about the number of times these relationships have been found to be true.

So, in a given transaction with multiple items, Association Rule Mining primarily tries to find the rules that govern how or why such products/items are often bought together. For example, sugar, milk and bread are frequently purchased together because a lot of people like to make tea.

3. LITERATURE REVIEW

Hao Feng et. al. [1] studied the association rules data mining algorithm, propose an association rule mining algorithm based on particle swarm algorithm, attribute object template data using the constrained concept lattice structure model set waiting for mining association rules, construction template data information flow time series analysis model, data

structure analysis, with frequent item index list search, data clustering using particle swarm algorithm for association rule features, to achieve large data clustering, a given minimum support and confidence threshold of two, in order to find the valuable association rules. The results show that this method of mining frequent itemsets can accurately reflect the fusion clustering characteristics of big data association rules of convergence of the mining process is good, can effectively extract the user interest all constrained association rules, it has the very good Application value.

S. Neelima et. al. [2] presenting an algorithm called hybridization of ABC with BAT algorithm is proposed which is used for optimization of association rules. Instead of onlooker bee phase of ABC, random walk of BAT is used in order to increase the exploration. Hybridized ABC with BAT algorithm is applied on the rules generated from apriori algorithm, for optimizing association rules. The experiments are performed on datasets taken from UCI repository which show the proposed work performance and proposed methodology can effectively optimize association rules when compared to the existing algorithms. In the paper, we also proved that the rules generated using proposed work are simple and comprehensible.

Polla A. Fatah et. al. [3] introduces an optimization approach for association rule mining in the time-memory domain. The approach splits the running mode of the traditional data mining algorithm into two phases. The first phase is designed to calculate all item sets in every transaction together with their frequencies (without pruning) and indexes their accumulation in a database. This procedure needs the fetch cycle of each transaction only once which reduces fetching transactions' I/O reasonably. In the second phase, the item sets and their frequencies are used in rule producing. a new algorithm has been designed, implemented, coded, verified and tested on real data. The approach enables users to change their queries and criteria using the second phase only which reduces the cost effectively.

4. ASSOCIATION MINING TECHNIQUES

4.1 APRIORI

Apriori is an association rule mining technique which when given the input of transactional databases it mines all frequently occurring items in the transaction. Here when given the Electronic Medical Record as the input to Apriori it then generates a set of risk factors that occur frequently and indicates those to be factors for developing diabetes.

4.2 ELCAT

Elcat procedure is similar to that of the Apriori algorithm which functions in a recursive manner. It makes use of tree like structure known as the Tid set. Further it processes the given EMR to generate the frequently occurring risk

factors of diabetes. The tid set starts with all one time occurring risk factor in the database.

The algorithm perform search in Depth first search manner to determine all the frequently occurring risk factor. The algorithm functions recursively and makes use of join operation() in combining all possible set of risk factors to the tid set and to generate (n+1) risk factors by considering the nth itemset. Eclat helps in lowering the memory being used during processing. As it is similar to that of apriori algorithm, elcat also generates a very large set of rule set but there is no candidate generation. The limitation is that it can discover only frequent item pattern rather rule set is not generated.

4.3 OPUS

Opus is an efficient technique that functions recursively with respect to the parameters in the Left hand side and the right hand side. The algorithm considers the current left hand side, compares with available left hand side then updates the current left hand side. The parameters can be constrained for left hand side and the right hand side.

Unlike other association algorithms it monotonic doesn't require any parameter like support or confidence for Association rule mining. User can specify the maximum number of associations to be generated.

5. FOLLOWING ARE THE SOME APPLICATIONS OF ASSOCIATION RULE MINING

Basket data analysis: is to analyze the association of purchased items in a single basket or single purchase as per the examples given above.

Cross marketing: is to work with other businesses that complement your own, not competitors. For example, vehicle dealerships and manufacturers have cross marketing campaigns with oil and gas companies for obvious reasons.

Catalog design: The selection of items in a business catalog is often designed to complement each other so that buying one item will lead to buying of another. So these items are often complements or very related.

Medical Diagnosis: Association rules in medical diagnosis can be useful for assisting physicians for curing patients. Diagnosis is not an easy process and has a scope of errors which may result in unreliable end-results. Using relational association rule mining, we can identify the probability of the occurrence of an illness concerning various factors and symptoms. Further, using learning techniques, this interface can be extended by adding new symptoms and defining relationships between the new signs and the corresponding diseases.

Protein Sequence: Proteins are sequences made up of twenty types of amino acids. Each protein bears a unique 3D structure

which depends on the sequence of these amino acids. A slight change in the sequence can cause a change in structure which might change the functioning of the protein. This dependency of the protein functioning on its amino acid sequence has been a subject of great research.

6. CONCLUSION

This paper presents the extensive of study of various Association Rule Mining algorithms in data mining which are really useful and very much needed to obtain useful facts or associations among data items in large data sets to take some important decision making in any kind of problems. The algorithmic aspects of association rule mining are dealt with. From a broad variety of efficient algorithms, the most important ones are compared. The algorithms are systemized and their performance is analysed based on runtime and theoretical considerations. Despite the identified fundamental differences concerning employed strategies, runtime shown by algorithms is almost similar.

7. REFERENCES

- [1] Hao Feng, Rongtao Liao, Fen Liu, Yixi Wang, Zheng Yu, Xiaojun Zhu, "Optimization Algorithm Improvement of Association Rule Mining Based on Particle Swarm Optimization", 10th IEEE, International Conference on Measuring Technology and Mechatronics Automation, 2018.
- [2] S. Neelima, Dr. N. Satyanarayana, Dr. P. Krishna Murthy, "Optimization of Association Rule Mining Using Hybridized Artificial Bee Colony (ABC) with BAT Algorithm", IEEE 7th International Advance Computing Conference, 2017.
- [3] Polla A. Fatah, Ibrahim Hamarash, "Optimization of Association Rule Mining A Two Step Breakdown Variation of Apriori Algorithm", 2015.
- [4] R. Agrawal, R. Srikant. "Fast algorithms for mining association rules in large databases". Proc. of 20th Int'l conf. on VLDB: 487-499, 1994.
- [5] J. Han, J. Pei, Y. Yin. "Mining Frequent Patterns without Candidate Generation". Proc. of ACM-SIGMOD, 2000.
- [6] Foto Afrati, Aristides Gionis, and Heikki Mannila. Approximating a collection of frequent sets. In ACM International Conference on Knowledge Discovery and Data Mining (KDD), 2004.
- [7] Association Rule Mining: A Survey. Gurneet Kaur. International journal of computer science and information technologies.
- [8] Comparative Analysis of Association Rule Mining Algorithms Based on Performance Survey. K.Vani. International journal of Computer Science and Information Technologies, 2015.
- [9] A Survey: On Association Rule Mining. Jeetesh Kumar Jain, Nirupama Tiwari and Manoj Ramaiya. International Journal of Engineering and Research and Application. February 2013.