

Association Rule Mining: A Survey

Jagmeet Kaur¹, Neena Madan²

¹*Department of Computer Science, Guru Nanak Dev University, Regional Campus
Jalandhar, Punjab, India*

²*Lecturer, Department of Computer Science, Guru Nanak Dev University,
Regional Campus
Jalandhar, Punjab, India*

Abstract

Last few years have seen a rapid growth in the area of data mining. Association rule mining is one of the most important techniques of data mining. It is a technique which is used to discover the frequently occurring patterns in the database. A lot of research has been done in the area of association rule mining. This paper gives a brief overview of association rule mining, some of its applications and algorithms. The limit of discussion is to the areas where association rule is used to a high extent.

Keywords- *Data mining, association rule, CRM, FP-tree, apriori algorithm, FP-growth algorithm, census data*

1. Introduction

Data mining is a process of retrieving useful information from the databases. It is used to deal with large amount of data. Data mining is used to predict the future trends and behaviour. These days data mining is becoming a useful technique as it has applications in various areas. Every day huge amount of data is generated in different fields. It is difficult to analyze such a large amount of data, thus data mining is used. It is also known as knowledge discovery in database (KDD). Data mining tools are used to predict the pattern and behaviour of the data, thus helping the organizations to make decisions [1]. Data mining is becoming popular because of its success in different areas. It has its applications in areas such as healthcare, finance, telecommunication, business, education and many more [2]. Data mining has two categories- descriptive mining and prescriptive mining. In descriptive mining, we identify and summarize the already existing data. In prescriptive mining, historical data is used to make predictions [3]. In data mining, we have different techniques and these are classification, clustering, association rule mining.

1.1. Association Rule Mining

Association rule mining is one of the most important techniques of data mining. It is used to discover the frequently occurring patterns in the database. The main aim of association rule mining is to find out the interesting relationships and correlations among the different items of the database [3]. Association rules are used in many fields to find out the patterns in the data. With the help of patterns, we can find out that how many combinations of events occur at the same time. Association rules are applied on large amount of data. In business field, association rules are used to discover the frequently occurring pattern which will help in marketing and decision making. Also association rules are applicable to many other fields, and these include medical, market basket analysis, library etc [4]. For example, if customer buys a shampoo then he may also buy a conditioner. These types of relationships can be used as a information which will be helpful in taking decisions for marketing purposes [5]. Generally association rules are

expressed in the form of $X \Rightarrow Y$. Here X and Y are the itemsets in the database. X is called as an antecedent and Y is called as consequent. Association rule mining consists of two basic measures and these are: support (s) and confidence (c).

Support: It is the probability that both X and Y will occur together in a transaction.

Confidence: It is also the probability, but it follows a condition. In confidence, if a transaction has X then it will also contains Y [6-8].

Rules are generated in association rule mining using following two steps:

1. All the frequent itemsets are found using minimum support.
2. Using these frequent itemsets strong association rules are generated, having confidence c above a predefined threshold value.

The first step requires more concentration, as it is difficult to find out all the frequently occurring itemsets [9].

2. Applications of Association Rule Mining

Association rule mining is used in different areas, as it is very advantageous. Some of the fields that have adopted association rule mining are discussed below.

2.1. Market Basket Analysis

One of the best and typical examples of association rule mining is market basket analysis, for example in supermarket stores. Managers of these stores want to increase the interest of the existing customers and to attract the new ones. Such stores consist of large no of databases and there is huge number of transactional records. So the managers may be interested to know whether some items are consistently purchased. The main aim is to analyze the buying behavior of the customers. So, association rule mining is used to generate the rules to find out the frequently occurring itemsets in a store. For example, if customers are buying bread and maximum number of customers are buying milk along with bread. Thus it will be beneficial for managers to place milk near the bread. Thus, association rule is helpful to design a layout for the store. So, market basket can be defined as the combination of different items purchased together by a customer in a single transaction [10][8][4].

2.2. CRM of the Credit Card Business

CRM i.e. Customer Relationship Management is a system that is used by a bank to manage its relationships with the customers. Banks usually identify the behavior of their customers to find out their likings and interests. In this way banks can increase the coherence of credit card customers and the bank. Here, association rules are helpful for the banks to know their customers in a better way and provide them good quality services [4][10].

2.3. Medical Diagnosis

Association rules are also used in the field of medical sciences. It is helpful for the physicians to cure the patients. Association rules are used to find out the probability of illness in a disease. There is a need for proper explanation of a disease. Thus diagnosis in itself is not an easy task. Adding some new symptoms to an existing disease, and then finding out the relationships between the symptoms will help the physicians. For example a physician is examining a patient, so it is obvious that he will require all the information of the patient, only then he can take a better decision for the patient. Here, association rule helps out as it is one of the most important research technique of data mining [4, 10-11].

2.4. Census Data

Association rule mining has a great possibility in census data. Association rules helps in supporting good public policy and in business development. Huge statistical information is generated by census. The information is related to the economic and population census and thus can be used in planning public services and business. In services, it includes health, education, transport etc. In business, it has constructing new malls, factories and banks [4, 10].

2.5. Protein Sequences

Proteins are the basic constituents of any cells of organism. There are many DNA technologies are available with different tools used for the fast determination of DNA sequences. Basically, proteins are the sequences which are made up of almost 20 types of amino acids. Every protein has a three dimensional structure, depending upon the sequence of amino acids. Thus there exists a too much dependency of protein functioning on amino acids. Association rules are generated between different amino acids of a protein. It will help to understand the protein composition [4, 10].

3. Different Methods of Association Rule Mining

There exists number of methods and algorithms which are used to generate association rules in databases. These include:

3.1. Apriori Algorithm

The most popular and classical algorithm used for association rule mining is Apriori algorithm. It is used for generating frequent itemsets for the database. The basic idea behind this algorithm is that multiple passes are made over the database. Apriori algorithm is based on a property, "all the nonempty subsets of a frequent itemsets must be frequent". It is based on breadth first approach. First of all, by scanning the database, the set of frequent 1-itemset is found and items that satisfy the minimum support are collected. The resulting set will be denoted as L1. Next, we will find out L2 using L1 and the process goes on until no more frequent itemsets can be found. This whole process requires a one full scan of the database.

Advantages of apriori algorithm:

1. It is easy to implement.
2. It implements level-wise search.

Disadvantages of apriori algorithm:

1. There is a need to scan database so many times.
2. There is need of more search space and cost of I/O is increased [7, 10].

3.2. FP- Growth Algorithm

This algorithm was used to overcome the drawbacks of apriori algorithm. Here, there is need to scan the database only two times. In this algorithm, there is a need to construct a FP- tree, thus two passes are used.

Pass 1:

1. The database is scanned to find out support and frequent 1-itemsets.
2. The infrequent items are discarded.
3. The items are arranged in the decreasing order of support.

This order is used to construct the FP-tree.

Pass 2:

Frequent itemsets having lower support are found out. FP-tree is used to extract the frequent itemsets [5][12].

3.3. Genetic Algorithm

This algorithm is basically implemented in the areas of optimization, biological evaluation and machine learning problems. In this, new population strings are generated using the old ones in an iteration manner. The strings are encoded as binary, real etc. In this algorithm, genetic operators are applied and these are selection, mutation and crossover. These are applied on random population to compute new strings. Selection is the probability of survival of fittest. Chromosomes are chosen to survive; fitness is comparable that how well a chromosome can solve a problem. In crossover, individual chromosomes are found out and these are combined to form new ones. In mutations, new solutions are altered in order to find the better ones [5].

4. Conclusion

Association rule mining is a very interesting rule and has number of applications in data mining. The applications and algorithms of association rule mining are reviewed in this paper. It is used to find out the interesting patterns and correlations among different items of database. It has applications in a number of areas; some of them are discussed above. Apriori, FP-growth and genetic are some important algorithms of association rule mining. Every algorithm has some advantages and disadvantages too. Enhancement can be done in the above discussed algorithms to reduce execution time and improve accuracy. A number of new algorithms can be developed to generate association rules, which can remove the disadvantages of existing algorithms.

References

- [1] N. Padhy, P. Mishra and R. Panigrahi, "The Survey of Data mining Applications and Feature Scope", International Journal of Computer Science, Engineering and Information Technology (IJCSSEIT), vol. 2, no. 3, (2012) June.
- [2] A. N. Paidi, "Data Mining: Future Trends and its Applications", International Journal of Modern Engineering Research (IJMER), vol. 2, no. 6, November-December (2012), pp. 4657-4663.
- [3] T. Karthikeyan and N. Ravikumar, "A Survey on Association Rule Mining", International Journal of Advanced Research in Computer and Communication Engineering, vol. 3, no. 1, (2014) January.
- [4] A. Rajak and M. K. Gupta, "Association Rule Mining: Applications in Various Areas", International Conference on Data Management.
- [5] I. Chandrakar and A. M. Kirithima, "A Survey on Association Rule Mining Algorithms", International Journal of Mathematics and Computer Research", vol. 1, no. 10, (2013), November.
- [6] S. Kotsiantis and D. Kanellopoulos, "Association Rules Mining: A Recent Overview", GESTS International Transactions on Computer Science and Engineering, vol. 32, no. 1, (2006), pp. 71-82.
- [7] C. Kaur, "Association Rule Mining Using Apriori Algorithm: A Survey", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), vol. 2, no. 6, (2013) June.
- [8] R. Trikha and J. Singh, "Improving the Efficiency of Apriori Algorithm by adding new Parameters" International Journal for Multi Disciplinary Engineering and Business Management, vol. 2, no. 2, (2014) June.
- [9] S. Maitrey and C. K. Jha, "Association Rule Mining: A Technique for Revolution in Requirement Analysis", International Journal of Scientific and Research Publications, vol. 4, no. 8, (2014) August.
- [10] J. K. Jain, N. Tiwari and M. Ramaiya, "A Survey: On Association Rule Mining", International Journal of Engineering Research and Applications (IJERA), vol. 3, no. 1, (2013) January-February, pp. 2065-2069
- [11] S. Soni and O. P. Vyas, "Using Associative Classifiers for Predictive Analysis in Health Care Data Mining" International Journal of Computer Applications, vol. 4, no. 5, (2010) July.
- [12] T. A. Kumbhare and S. V. Chobe, "An Overview of Association Rule Mining Algorithms", International Journal of Computer Science and Information Technologies, vol. 5, no. 1, (2014), pp. 927-930.