



## Association Rule Mining: An Overview

Kainaz B. Sherdiwala<sup>1st</sup>

Assistant Professor, Dept of Computer Application  
Z.F.Wadia Women's College,  
Surat, Gujarat (India)

Dr. Samrat O. Khanna<sup>2nd</sup>

Head and Professor, Dept of MSc. IT  
ISTAR  
V. V. Nagar, Gujarat (India)

**Abstract:** Data mining is the process of discovering previously unknown, undiscovered and hidden patterns from large information repositories like relational databases and warehouses. It is of great importance in the recent time where the amount of data has reached to gigabytes and even terabytes. Data mining makes use of various techniques such as Clustering, Classification, Association Rule Mining and Regression. Association Rule Mining finds association between the items in the database. This paper gives us a brief idea regarding Association Rule Mining and applications of association rule mining in different areas for effective decision making.

**Keywords:** Classification, Data Mining, Association Rule Mining.

### I. INTRODUCTION

Association rule mining is the most important technique in the field of data mining. Association rule mining helps in finding frequent patterns, associations, correlations, or causal structures among sets of items or objects in transaction databases, relational databases, and other information repositories. Association rule mining tries to find such relationships among the attributes of the database which may be helpful in the task of decision making. The most important application of association rule mining is in the field of market basket analysis. But it is not restricted to this field only. It finds the new useful rules in the sales transaction database, which reflects the customer purchasing behaviour patterns, such as the impact on the other goods after buying a certain kind of goods. These rules can be used in many fields, such as customer shopping analysis, additional sales, goods shelves design, storage planning and classifying the users according to the buying patterns, etc. These association rules can be easily interpreted and communicated.

Association rule mining finds interesting associations and/or correlation relationships among large set of data items. Association rules show attributes value conditions that occur frequently together in a given dataset. Association rules provide information of this type in the form of "if-then" statements. These rules are computed from the data and, unlike the if-then rules of logic, association rules are probabilistic in nature. In addition to the antecedent (the "if" part) and the consequent (the "then" part), an association rule has two numbers that express the degree of uncertainty about the rule. In association analysis the antecedent and consequent are sets of items.

There are three phases in the problem of data mining association rules.

1. **Support:** The support is simply the number of transactions that include all items in the antecedent and consequent parts of the rule. The support is sometimes expressed as a percentage of the total number of records in the database.
2. **Confidence:** Confidence is the ratio of the number of transactions that include all items in the consequent as well as the antecedent to the number of transactions that include all items in the antecedent
3. **Lift:** Lift is nothing but the ratio of confidence to expected confidence. Lift is a value that gives us information about the increase in probability of the "then" (consequent) given the "if" (antecedent) part.

### II. APPLICATIONS OF ASSOCIATION RULE MINING

The various areas in which association rule mining can be applied are:

#### 1. Market Basket Analysis:

A typical and widely-used example of association rule mining is market basket analysis. For example, data are collected using bar-code scanners in supermarkets. Such 'market basket' databases consist of a large number of transaction records. Each record lists all items bought by a customer on a single purchase transaction. Managers would be interested to know if certain groups of items are consistently purchased together. They could use this data for adjusting store layouts, for cross-selling, for promotions, for catalogue design and to identify customer segments based on buying patterns [18].

#### Association rules are of the form if X then Y. For example:

In an online book store there are always some tips after you purchase some books, for instance, once you bought the book Data Mining Concepts and Techniques, a list of related books such as: Database System 40%, Data Warehouse 25%, will be presented to you as recommendation for further purchasing.

In the above example, the association rules are: when the book Data Mining Concepts and Techniques is brought, 40% of the time the book Database System is brought together, and 25% of the time the book Data Warehouse is brought together. Those



rules discovered from the transaction database of the book store can be used to rearrange the way of how to place those related books, which can further make those rules more strong. Those rules can also be used to help the store to make his market strategies such as: by promotion of the book Data Mining Concepts and Techniques, it can blow up the sales of the other two books mentioned in the example.

## 2. Medical diagnosis:

Applying association rules in medical diagnosis can be used for assisting physicians to cure patients. The general problem of the induction of reliable diagnostic rules is hard because theoretically no induction process by itself can guarantee the correctness of induced hypotheses.

Practically diagnosis is not an easy process as it involves unreliable diagnosis tests and the presence of noise in training examples. This may result in hypotheses with unsatisfactory prediction accuracy which is too unreliable for critical medical applications.

Serban has proposed a technique based on relational association rules and supervised learning methods. It helps to identify the probability of illness in a certain disease. This interface can be simply extended by adding new symptoms types for the given disease, and by defining new relations between these symptoms.

## 3. Census Data:

Censuses make a huge variety of general statistical information on society available to both researchers and the general public. The information related to population and economic census can be forecasted in planning public services (education, health, transport, funds) as well as in public business (for setup new factories, shopping malls or banks and even marketing particular products).

The application of data mining techniques to census data and more generally to official data, has great potential in supporting good public policy and in underpinning the effective functioning of a democratic society. On the other hand, it is not undemanding and requires exigent methodological study, which is still in the preliminary stages.

## 4. CRM of Credit Card Business:

Customer Relationship Management (CRM), through which, banks hope to identify the preference of different customer groups, products and services tailored to their liking to enhance the cohesion between credit card customers and the bank, has become a topic of great interest [10]. Shaw [8] mainly describes how to incorporate data mining into the framework of marketing knowledge management.

The collective application of association rule techniques reinforces the knowledge management process and allows marketing personnel to know their customers well to provide better quality services. Song [6] proposed a method to illustrate change of customer behaviour at different time snapshots from customer profiles and sales data. The basic idea is to discover changes from two datasets and generate rules from each dataset to carry out rule matching.

### III. OTHER APPLICATIONS

Some of the areas in which association rules have been used include:

- **Credit card transactions:** items purchased by credit card give insight into other products the customer is likely to purchase.
- **Supermarket purchases:** common combinations of products can be used to inform product placement on supermarket shelves.
- **Telecommunication product purchases:** commonly associated options (call waiting, caller display, etc) help determine how to structure product bundles which maximise revenue.
- **Banking services:** the patterns of services used by retail customers are used to identify other services they may wish to purchase.
- **Insurance claims:** unusual combinations of insurance claims can be a sign of fraud.
- **Medical patient histories:** certain combinations of conditions can indicate increased risk of various complications.

### IV. LIMITATIONS OF ASSOCIATION RULE MINING

The main difficulty in applying association rules mining is the setting of support threshold. Many approaches assume that all items in the database are of the same kind and have similar frequencies in the database but this assumption is not applicable in reality. In real life applications, some items appear very frequently in the database, while others hardly ever appear and the frequent item sets are alone not interesting, but the rare items also. If a high support threshold is specified, it misses some interesting rules and if a low support threshold is used, it generates numerous unnecessary rules.

In short, we can say that some of the major drawbacks of Association Rule Mining are:

1. Obtaining non interesting rules
2. Obtaining huge number of rules



3. Low algorithm performance
4. Cannot incorporate domain/ user defined knowledge
5. Not suitable for supervised learning

## V. SOLUTIONS FOR INCREASING EFFICIENCY OF ALGORITHM

The efficiency of association rule mining algorithm can be increased in the following ways:

- By reducing the number of passes over the database
- By sampling the database
- By adding extra constraints on the structure of patterns
- Through parallelization.

## VI. FUTURE WORK

In this paper a revision on the main application areas of association rules has been focused. It is all about to find some kind of pattern or relationship between various datasets. The outcome is association rules, and it is an iterative refinement process. End users of association rule mining tools encounter several well known problems in practice. Solutions to these problems are also given here.

## REFERENCES

1. Akash Rajak and Mahendra Kumar Gupta, Association Rule Mining: Applications in Various Areas, International Conference on Data Management. Agrawal, R., Imielinski, T., and Swami, A. N. 1993. Mining association rules between sets of items in large databases. In Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, 207-216.
2. Agrawal, R. and Srikant, R. 1994. Fast algorithms for mining association rules in Proc. 20th Int. Conf. Very Large Data Bases, 487-499.
3. Agarwal, R. Aggarwal, C. and Prasad V., A tree projection algorithm for generation of frequent itemsets in J. Parallel and Distributed Computing, 2000.
4. Klossgen, W., & Zytkow, J.: Handbook of data mining and knowledge discovery. Oxford University Press, New York (2002).
5. Sotiris Kotsiantis, Dimitris Kanellopoulos, Association Rules Mining: A Recent Overview, GESTS International Transactions on Computer Science and Engineering, Vol.32 (1), 2006, pp. 71-82.
6. Ashrafi, M., Taniar, D. Smith, K. A New Approach of Eliminating Redundant Association Rules, Lecture Notes in Computer Science, Volume 3180, 2004, Pages 465 – 474.
7. Ashrafi, M., Taniar, D., Smith, K., Redundant Association Rules Reduction Techniques, Lecture Notes in Computer Science, Volume 3809, 2005, pp. 254 – 263.
8. Baralis, E., Psaila, G., Designing templates for mining association rules. Journal of Intelligent Information Systems, 9(1):7-32, July 1997.
9. Brin, S., Motwani, R., Ullman, J. D., and Tsur, S. 1997. Dynamic itemset counting and implication rules for market basket data. In SIGMOD 1997, Proceedings ACM SIGMOD International Conference on Management of Data, May 13-15, 1997, 255-264.
10. Brin, S., Motwani, R. and Silverstein, C., "Beyond Market Baskets: Generalizing Association Rules to Correlations," Proc. ACM SIGMOD Conf., pp. 265-276, May 1997.
11. Cheung, D., Han, J., Ng, V., Fu, A. and Fu, Y. (1996), A fast distributed algorithm for mining association rules, in 'Proc. of 1996 Int'l. Conf. on Parallel and Distributed Information Systems', Miami Beach, Florida, Pages 31 - 44.
12. Cheung, D., Xiao, Y., Effect of data skewness in parallel mining of association rules, Lecture Notes in Computer Science, Volume 1394, Aug 1998, Pages 48 – 60.
13. Chuang, K., Chen, M., Yang, W., Progressive Sampling for Association Rules Based on Sampling Error Estimation, Lecture Notes in Computer Science, Volume 3518, Jun 2005, Pages 505 – 515.
14. Li, Y., Gopalan, R., Effective Sampling for Mining Association Rules, Lecture Notes in Computer Science, Volume 3339, Jan 2004, Pages 391 – 401.
15. Liu, B. Hsu, W., Ma, Y., "Mining Association Rules with Multiple Minimum Supports," Proc. Knowledge Discovery and Data Mining Conf., pp. 337-341, Aug. 1999.
16. Manning, A., Keane, J., Data Allocation Algorithm for Parallel Association Rule Discovery, Lecture Notes in Computer Science, Volume 2035, Page 413-420.
17. [http://www.resample.com/xlminer/help/Assocrules/associationrules\\_intro.htm](http://www.resample.com/xlminer/help/Assocrules/associationrules_intro.htm)
18. N. Gupta, N. Mangal, K. Tiwari and P. Mitra, "Mining Quantitative Association Rules in Protein Sequences", In Proceedings of Australasian Conference on Knowledge Discovery and Data Mining – AUSDM, 2006.
19. R. S. Chen, R. C. Wu and J. Y. Chen, "Data Mining Application in Customer Relationship Management Of Credit Card Business", In Proceedings of 29th Annual International Computer Software and Applications Conference (COMPSAC'05), Volume 2, pages 39-40.
20. C.S.Kanimozhi Selvi, and A.Tamilarasi, An Automated Association Rule Mining Technique With Cumulative Support Thresholds, Int. J. Open Problems in Compt. Math, Vol. 2, No. 3, September 2009, ISSN 1998-6262, 2009.