

## APPLICATIONS OF ASSOCIATION RULE MINING IN DIFFERENT DATABASES

Dr. M. Renuka Devi <sup>\*1</sup> Mrs. A. Baby sarojini<sup>2</sup>

<sup>\*1</sup>Assistant Professor of MCA Department, Sree Saraswathi Thyagaraja College, Pollachi. Bharathiar University, Coimbatore, Tamil Nadu, India.  
renuga.srk@gmail.com

<sup>2</sup>Research Scholar of Computer Science, Sree Saraswathi Thyagaraja College, Pollachi. Bharathiar University, Coimbatore, Tamil Nadu, India.  
babysarojinia@gmail.com

**Abstract:** Data mining uses expert methods and techniques to recognize trends and profiles hidden in data. Mining is an iterative process in a sequence. Data are comes from different sources such as different databases. Mining depends on databases. This research is for various association rule mining applications of different databases. There are different databases like large database, distributed database, medical database, relational database, spatial database. They are mined using in data mining techniques. For making decision association rule is most important. They are associated with association rule mining techniques.

**Keywords:** Data Mining, Association Rule Mining, Spatial Data Mining, RDBMS, Medical Database, Large Database, Distributed Database.

### INTRODUCTION

Data mining having many techniques, methods, rules etc. to extract a particular data from large database. Association is mostly used for decision making with the measures such as support and confidence. Association used to finding patterns in many data. Association helps in business to make a decision in marketing and other fields. Decision making is most important in association rule mining. Association rule mining is to discover available association rules that fulfill the predefined least possible support and confidence from a supposed database.

Two types of association rule mining are used in large databases. One is positive association rule mining. The second one is negative association rule mining. Mining negative association rules acting as vigorous role in decision making consequence. Association rule mining seeks to discover associations among transactions encoded in a database. It can be used to improve decision making in a wide variety of applications such as: medical diagnosis, GIS, relational database, large database and distributed database etc. These databases are reviewed. This study

discuss about how the association rule mining used in different databases.

Figure 1 refers to the association rules data mining architecture. The alarms arrive at the Security Operations Center (SOC), there SOC Analyst to analysis data from association rule data mining; they are stored in the short term in a database on the monitoring database. From this temporary database extracted the set of all alarms generated in a single day for all databases and loaded them into a analytical data warehouse. It is on this warehouse that executed the data mining algorithms with the goal of producing new checking rules for installation in the ESM (Enterprise Security Management System). That checks the data from different databases.

Reference [5] proposed a method for intrusion detection in large database using association rule mining. Reference [10] introduced a new technique for medical large database for heterogeneous genome data using association rule mining. For mining remotely sensed images/data in association rules in spatial mining proposed by Dong et al (2000). Making decision in a business environment using association rule mining to sort a product assortment decisions proposed by [2], [3] and [12].

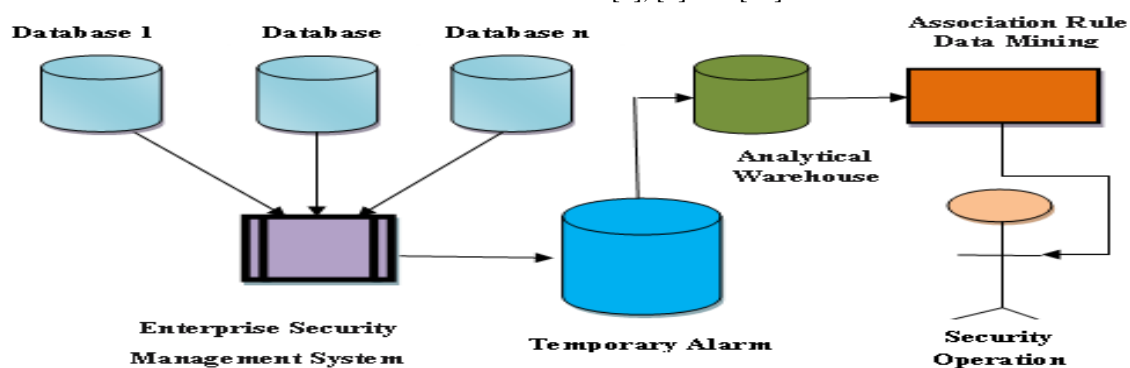


Figure 1. The Association Rules Data Mining Architectur

**Association Rule Mining in Large Database:**

Association rule mining used to mine the sales transactions between items in large database recognized as a most significant area of database research. Measuring a large database there are different techniques are used. Pruning strategy and interestingness is one of the measuring techniques for measuring large database. Large database consists of many fields. Each field consists of their own process. They different depends on their field of work. Suppose we consider a customer transaction of a large database each transaction consists of items purchased by a customer in a visit items purchased by a customer in a visit , time of purchase, category of payment, net amount etc. so it is a tedious process to maintain for huge amount of customer transaction.

An efficient algorithm implemented in association rule mining. Apriori algorithm is best for association rule mining in large database. This algorithm generates all significant association rules between items in the large database. Today, most research related work on data mining in association rules are encouraged by an wide range of application areas, such as financial transactions, engineering, health care, GIS, and broadcastings. Association rule mining used to originate interesting association or correlation relationships among a large set of items in the large database. In large database Application of association rule mining in market basket analysis are

- a. To analyses the point of sales transaction
- b. From uses information on what customers buy to provide insights into who they are and why they make certain purchases
- c. From which products are purchased together and which are most willing to support.

Agrawal & Srikant was proposed a well-known approach which is Apriori Algorithm. This approach is an iterative process. Each iteration has two steps.

**Step 1:** To generate a set of candidate item sets.

**Step 2:** To prune all the disqualified candidates (i.e. all infrequent item sets).

The iterations begin with size two item sets and each iteration size is incremented. The algorithm is depends on the closure property of frequent item sets: if a set of items is frequent, then all its proper subsets are also frequent. The faults of this algorithm are the generation of a large number of candidate item sets and the requirement to scan the database once at each iteration. To overcome the above weaknesses, Han and Pei & Yin were proposed a FP-tree and FP-growth algorithm. The idea of FP-tree is fetching all transactions from the database and inserting them into a compressed tree structure. Then, FP-growth algorithm reads from the structure FP-tree to mine frequent item sets.

Reference [9] established the effectiveness of their algorithm by put on it to sales data obtained from a large database. For this data set, the algorithm demonstrated outstanding performance. The assessment method demonstrated high precision and the pruning techniques were able to prune out a very large fraction of item sets without determining them.

**Association Rule Mining in Medical Database:**

Associative classification rule mining is a combination of association rule mining integrated with classification rule mining. It is used in medical database.

Based on a hospital physical examination database, [13] said in their article set up an association rules mining system, and through the establishment of the system of medical personnel in information management and analysis, in which application of association rule mining algorithm based on genetic algorithm for data mining. Look forward to the establishment and implementation of systems to help hospitals manage medical information. And also in the medical information from hospital, the Mining Association rule has practical significance; you can use these rules to guide the daily lives of medical staff, and makes recommendations for public health. Improved Pc and Pm Adaptive Algorithm is used to maintain weight control.

Association rules can expose biologically significant associations between different genes or between environmentally friendly belongings and gene expression. An association rule has the form  $LHS \Rightarrow RHS$ , where  $LHS$  and  $RHS$  are disjoint sets of items, the  $RHS$  set being likely to happen at whatever time the  $LHS$  set occurs. Items in gene expression data can include genes that are extremely articulated or inhibited, as well as related facts labeling the cellular atmosphere of the genes.

In many healthcare settings, patients visit healthcare specialists occasionally and report multiple medical illnesses, or symptoms, at each encounter. A statistical modeling technique, called the Hierarchical Association Rule Model (HARM) developed by [11], that forecasts a patient’s possible future symptoms given the patient’s present and past history of testified symptoms. The principal of this technique is a Bayesian hierarchical model for selecting predictive association rules (such as “symptom 1 and symptom 2  $\rightarrow$  symptom 3”) from a large set of candidate rules. Because this method “borrows strength” using the symptoms of many similar patients, it is able to provide predictions specialized to any given patient, even when little information about the patient’s history of symptoms is available.

Association rules relate disease data measures the patient risk factors and appearance of the disease in medical terms. Association rule medical consequence is estimated with the usual support and confidence metrics. Association rules are used to compare analytical rules mined with decision trees, a well-known machine learning method.

**Association Rule Mining in Distributed Database:**

Databases or data warehouses may store a large amount of data (large database) to be mined. Mining association rules in large databases may require extensive processing power. Distributed system is to solve this problem in large database mining. Many large databases are distributed .More feasible to use Distributed algorithms by distributed system. Distributed computing of large item sets encounters certain different complications. To solve this complications by using different distributed algorithms. Such as

- a. Distributed association rule learning

- b. Distributed hierarchical clustering
- c. Collective PCA and PCA-based clustering
- d. Collective decision tree learning
- e. Collective Bayesian network learning
- f. Other distributed clustering algorithms

The response time with the communication and calculation issues are measured to succeed a better response time, number of processors in a distributed environment by using the optimized distributed algorithm proposed by [8].

Centralized data mining to discover useful patterns in distributed databases isn't always feasible because merging data sets from different sites sustains huge network communication costs. An Improved algorithm based on good performance level for data mining is being proposed by [1]. Their research is to cultivate a distributed algorithm for geographically distributed data sets that decreases communication costs, superior running effectiveness, and stronger scalability than straight use of a sequential algorithm in distributed databases.

Distributed higher-order association rule mining algorithm is to determine propositional rules established on higher-order associations in a distributed surroundings and also detect a critical suppositions made in existing association rule mining algorithms that preclude them from scaling to complex distributed surroundings in which the complete global schema is indefinite, data is inconsistent in a hybrid non-vertical, non-horizontal form, and errors occur in record linkage.

An innovative application of deduction rules is introduced for distributed mining of association rules which achievements the derivability of item sets to moderate communication overhead and to develop response time. A new algorithm is proposed which mines derivable and non-derivable frequent item sets in a distributed database by [6].

**Association Rule Mining in Relational Database:**

An ever growing number of organizations are installing large data warehouses using relational database technology. There is a huge demand for mining bits of knowledge from these data warehouses. Association rule mining is used to make a decision to solve this problem.

Relational association rules and supervised learning methods help 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. A Fuzzy Association Rule Mining II for handling both relational and transactional data in relational database.

Many industrial databases applications make use of relational databases. It is used to store, manipulate and reclaim regulated data from large database. Through association rule mining from relational databases utilize database indexing and query optimization procedures applied in relational database management systems to develop performance and improve efficiency. Association Rule mining in the relational database is the process of recognizing the dependency of one item(s) with respect to the existence of other item(s). This helps to study the buying patterns of their customers.

The Algorithm SETM proposed by [7]. Association rule mining set-oriented algorithms suggest performing multiple joins and may appear to be fundamentally less effective than special-purpose algorithms. To solve this problem must develop innovative algorithms that can be expressed as SQL queries, and discuss optimization of these algorithms. Algorithm SETM uses only simple database primitives, viz., sorting and merge-scan join. Algorithm SETM is simple, fast, and stable over the range of parameter values. The set-oriented nature of Algorithm SETM simplifies the development of extensions in relational database mining.

Figure 2 explain about the multi relational database environment. The development of multi-database association rule mining is a fascinating and risky job. Since it wants information of all the data kept at different places and the capability to associate incomplete outcomes from specific RDBMS's into a particular outcome. The specific databases have to be analyzed to create rules to make local decisions. It would be easier for the organization to make decisions based on the rules created by the specific branches, rather than using the raw data.

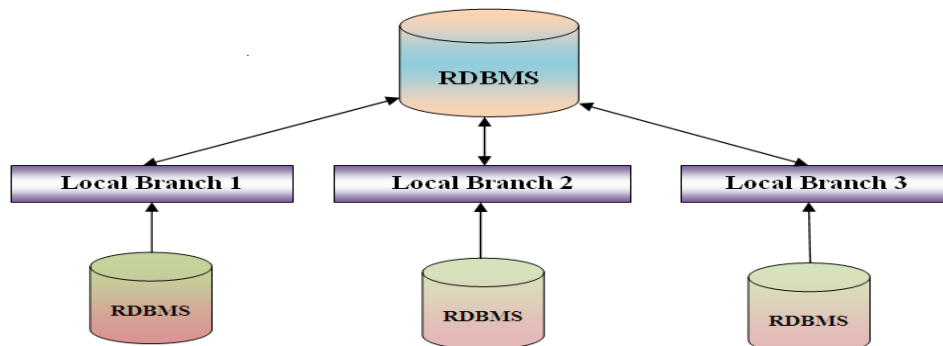


Figure 2. A Multi Relational Database Environment

**Association Rule Mining in Spatial Database:**

Spatial Data Mining is the discovery of fascinating patterns from large geospatial databases. It refers to the extraction of knowledge, spatial associations or other fascinating patterns

not clearly stored in spatial databases. In data mining association rules are encompassing spatial relations among spatial substances. Spatial database contains objects which are described by a spatial scene and/or extension as well as by several non-spatial attributes. The innovation process for

spatial data is more multifaceted than for relational data. Spatial data mining algorithms have to consider the neighbours of substances in order to mine useful knowledge. It is indispensable because the attributes of the neighbours of some substance of curiosity may have a momentous inspiration on the substance itself. Spatial data mining algorithms has several advantages. Similar to the relational standard language SQL,

- a. The use of standard primitives will speed-up the development of new data mining algorithms and will also make them more portable.
- b. Develop methods to powerfully support the proposed database primitives, thus speeding-up all data mining algorithms which are based on the database primitives. The basic operations for spatial data mining can be incorporated into commercial database management systems. This will offer bonus welfares for data mining applications such as
  - a) Efficient storage management,
  - b) Prevention of inconsistencies,
  - c) Index structures to support different types of database enquiries which may be part of the data mining algorithms.

The application of data mining techniques in spatial database to census data, and more generally, to official data, has great potential in supporting worthy public strategy and in sustaining the actual operational of an independent society. Spatial data mining approaches and procedures have been suggested for the mining of hidden knowledge, spatial relations, or other patterns not clearly stored in spatial databases.

Spatial data mining is used in

- a. NASA Earth Observing System (EOS) for Earth science data
- b. National Inst. of Justice for crime mapping
- c. Census Bureau, Dept. of Commerce for census data
- d. Dept. of Transportation (DOT) for traffic data
- National Inst. of Health (NIH) for cancer clusters

**Conclusion and Future Work:**

Association rule mining performing vital role in the essential area data mining. It poses many exciting problems for the enlargement of well-organized and effective techniques. After taking a closer look, we discover that the application of association rules needs much more investigations in order to aid in more unambiguous objectives. Applications of association rule mining are Large and Distributed database - Businesses, e.g. logistics, marketing and Government - almost all branches e.g. defense, public safety, Spatial database - GIS, Relational database - Industries, Medical database- Medical diagnosis, Hospital, Medical shops, scan centers... Future work is to find out the better support and confidence of different algorithms with association rule mining.

**REFERENCES**

[1]. J. Arokia Renjit and Dr.K.L.Shunmuganathan,” Mining the Data from Distributed Database Using an Improved Mining Algorithm”, Vol. 7, No. 3, March 2010 [(IJCSIS) International Journal of Computer Science and Information Security].

[2]. Brijs, T., Goethals, B., Swinnen, G., Vanhoof, K. and Wets, G, “A Data Mining Framework for Optimal Product Selection in Retail Supermarket Data: The Generalized PROFSET Model”, 2000 [SIGKDD].

[3]. Brijs, T., Swinnen, G., Vanhoof, K. and Wets, G, “Using Association Rules for Product Assortment Decisions: A Case Study”, 1999 [SIGKDD].

[4]. Dong, J., Perrizo, W., Ding, Q. and Zhou, J., “The application of association rule mining to remotely sensed data”, 2000 [Proceedings of the ACM symposium on applied computing].

[5]. Lee, W., Stolfo, S.J. and Mok, K.W., “A data mining framework for building intrusion detection models”, 1999 [IEEE Symposium on Security and Privacy].

[6]. M Deypir and M H Sadreddini, “Distributed Association Rules Mining Using Non- derivable Frequent Patterns”, Vol. 33, Issue B6, pp. 511-26, 2009 [Iranian Journal of Science & Technology, Transaction B: Engineering].

[7]. Maurice Houtsma and Arun Swami,” Set-oriented mining for association rules in relational databases”, Vol. 17, pp. 245-262, 1995 [Data and Knowledge Engineering].

[8]. Pallavi Dubey ,“Association Rule Mining on Distributed Data”,Vol. 3, Issue 1, ISSN : 2229-5518, 2012 [International Journal of Scientific & Engineering Research].

[9]. Rakesh Agrawal, Tomasz Imielinski and Arun Swami,”Mining Association Rules between Sets of Items in Large Databases “, May 1993 [Proceedings of the ACM SIGMOD Conference, Washington DC, USA].

[10]. Satou, K., Shibayama, G., Ono, T., Yamamura, Y., Furuichi, E., Kuhara, S. and Takagi, T., “Finding association rules on heterogeneous genome data”, 1997 [PSB].

[11]. Tyler H. McCormick, C Rudin and D Madigan, “A Hierarchical Model for Association Rule Mining of Sequential Events: An Approach to Automated Medical Symptom Prediction”, 2000 [Submitted to the Annals of Applied Statistics].

[12]. Wang, K. and Su, M.Y., “Item Selection by Hub-Authority Profit Ranking”, 2002 [SIGKDD].

[13]. Xinhang Xu, Qihong S, Hongtao Z, Lei W and Y Liu, “Study and Application on the Method of Association Rules Mining Based on Genetic Algorithm”, 2012 [The 2nd International Conference on Computer Application and System Modeling].

**Short Bio Data for the Author**



Dr. M.Renuka Devi has nearly 8 years of post-graduate teaching experience in Computer Science. She has indulged in training the post graduate students to complete real time projects and also guides research scholars in Computer Science. Currently she is working as Lecturer in the Department of MCA at Sree Saraswathi Thyagaraja College (Autonomous), and An ISO 9001 Certified / NAAC Accredited Institution, Pollachi, Coimbatore (Dt), Tamil Nadu, India.



Mrs. A. Babysarojini, has completed her Post Graduate in M.Sc (Software Engineering). Currently she is doing her research at Sree Saraswathi Thyagaraja College

(Autonomous), and An ISO 9001 Certified / NAAC Accredited Institution, Pollachi, Coimbatore (Dt), Tamil Nadu, India.