# Association Rule Classification based on Multi-Dimensional Data Cubes

### Dr. K.Kavitha
Assistant Professor, Department of Computer Science,
Mother Teresa Women's University, Kodaikanal
kavitha.urc@gmail.com

## Abstract

*Data mining facilitates the discovery of unrevealed trends from huge datasets. Data warehouse is a key technology for complex data analysis automatic knowledge extraction and decision making. Multi-dimensional database permits the data for efficient and appropriate storage. The dimensional table holds all different attributes and dimensions. Detecting the hidden association between the items are limited in OLAP. Researchers have proposed many ideas to reduce the limitations. This paper presents an approach called Association rule classification for Multi-Dimensional dataset. This proposed work detects the hidden association form OLAP and also categorize the rule sets effectively.*

**Keyword:** Association Rule, Interesting Measures, Cluster, Multi-Dimensional Data, DataCube

## 1. Introduction

Data mining facilitates the discovery of unrevealed trends from large dataset. Data warehousing provides primary and interactive analysis of data through the use of different data aggregation methods. Data mining and warehousing are the key technologies which made important contributions in the field of knowledge discovery. A huge range of data mining techniques has made significant improvements in knowledge discovery from various domain. OLAP tool used in multidimensional analysis and organizers multidimensional cubes based on this dimensional model. OLAP achieves different operations on the data which stored in multi-dimensional database. Association rule is a data mining method used to determine interesting patterns and correlations between there it is a descriptive techniques to discover data which identity relationships between values in a database. The association rule extraction process is divided into two phases such as

- Determine frequent itemsets
- Build the association rules using obtained frequent itemsets.

This approach defines an interesting measure that is used to prepare the rule set. The association is found between elements in the similar dimension using additional dimension to compute the support called intra-dimension association.

The association hold between elements in different dimensions called Inter dimension association. Hybrid association rules inter and intra dimension association. It generates frequent predicates. Data cube structure describes relations at the lowest abstraction level. Interesting measure used to dimension the certain problems of confident while measuring the quality of rules. However, the different abstraction levels do not control the redundant association.

Rest of this paper is structured as follows. Research works related to Risk Assessment are discussed in Section2.

Detailed explanations of implementing Prediction Algorithm presented in section 3. finally section 4 concludes this research work.
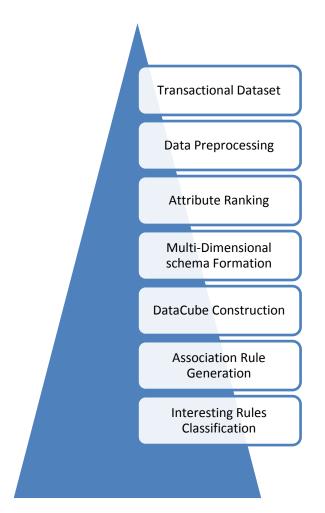
## 2. Related Work

Credit Risk evaluation is an interesting management problem in financial structure. Francesca.et.al proposed hazard model for predicting loan population which involves different probability of risk factors. Probability is modeled into two groups such as good and bad borrowers [1].

Zakrzewska.et.al proposed a technique allows building of different rules for different type of customers; each applicant is assigned to similar group of concerned part [2].Bhasin.et.al presented a model to extract efficient information from existing data and enables the system to make better decisions [3].

## 3. Association Rule for classifying Multi-Dimensional data

Multi-dimensional schema is generated by combining hierarchical and multi dimensional scaling techniques. It needs to identity the significant variables which defines the data splitting. Additionally, ranking nominal and numerical variables in each cluster by comparing the neighbouring clusters. It creates a candidate schema with high ranked dimensions and measures. This algorithm presents the multi-dimensional schema design for diversing association rules. Highest ranked dimensions and facts present in Multi-dimensional schema is to be used to built data cube which mines the effective association rule based on the importance among the rules. The following figure shows the main steps involved in proposed work.

The proposed approaches generates multi-dimensional schema and AR algorithm for classifying Interesting rules, Multi-dimensional schema is formulated with Informative data cubes. It filters unnecessary dimensional and facts from initial database. This method extracts more interesting Association Rules. Prediction accuracy of interesting rules is higher than the rules generated from existing methods.

## References

[1] G.Francesca, "A Discrete-Time Hazard Model for Loans: Some evidence from Italian Banking system", American Journal of Applied Sciences, vol. 9, p. 1337, 2012

[2]D.Zakrzewska, "On integrating unsupervised and supervised classification for credit risk evaluation", Information Technology and Control, Vol. 36, pp. 98-102, 2007.

[3]M.L.Bhasin, "Data Mining: A Competitive tool in the Banking and Retail Industries" Banking and Finance, vol 588,2006

[4] M.Usman, R.Pears and A.Fong, "Discovering diverse Association Rules from Multidimensional Schema,",2013

[5] R.Pears,M.Usman and A.Fong, "Data guided approach to generate Multi0Dimensional schema for targeted Knowledge discovery", 2012.

[6] G.Liu, H.Jiang, R.Geng and H.Li, "Application of Multidimensional association rules in personal financial services" in Computer design and Applications(ICCDA), 2010 International Conference ,pp V5-500-503

[7] W.Y.Chiang "To mine association rules of Customer values via data mining procedure with improved model:An empirical case study'Expert Systems with Applications, Vol 38, pp 1716-1722,2011

[8] T.Herawan and M.M.Deris, "A soft set approach for association rule mining Knowledge based Systems", vol 24, pp 186-195,2011.

9]A. Ghatge and P. Halkarnikar, "Ensemble Neural Network Strategy for Predicting Credit Default Evaluation."

[10]S. Purohit and A. Kulkarni, "Credit evaluation model of loan proposals for Indian Banks," in Information and Communication Technologies (WICT), 2011 World Congress on, 2011, pp. 868-873.

[11]D. Zakrzewska, "On integrating unsupervised and supervised classification for credit risk evaluation," Information Technology and Control, vol. 36, pp. 98-102, 2007.

[12]M. L. Bhasin, "Data Mining: A Competitive Tool in the Banking and Retail Industries," Banking and finance, vol. 588, 2006.

[13]N. İkizler and H. A. Guvenir, "Mining interesting rules in bank loans data," in Proceedings of the Tenth Turkish Symposium on Artificial Intelligence and Neural Networks, 2001.

[14]J. Nassali, "A Loan Assessment System for Centenary Rural Development Bank," 2005.

[15]T. Jacobson and K. Roszbach, "Bank lending policy, credit scoring and value-at-risk," Journal of banking & finance, vol. 27, pp. 615-633, 2003.

[16]G. Kabir, I. Jahan, M. H. Chisty, and M. A. A. Hasin, "Credit Risk Assessment and Evaluation System for Industrial Project."

Dataset is initially pre-processed to attain the highest quality. It is used to remove the unnecessary dates in the dataset.

## Algorithm

Input: Sample dataset
Output: Interesting Rules
Step 1: Ranking Attributes
  Attributes Categorization (Numeric and nominal)
  Rank Numeric attribute by Eigen value
  Rank Nominal attribute by Information given
Step 2: Multi Dimensional schema
  Grouping N & No attributes
  MD Schema Formation
Step 3: Categorise Data cubes based on Ranking
  Fix threshold
  Apply OLAP operation
Step 4: Association Rule generating
  Find support, confidence weight Identify Interesting Measures.
Step 5: Classification of Interested Rules
  Calculate mean for rule set
  Categorize the rules based on mean value and weight into high, medium and low.

## Conclusion

[17]B. Bodla and R. Verma, "Credit Risk Management Framework at Banks in India," ICFAI Journal of Bank Management, Feb2009, vol. 8, pp. 47-72, 2009.

[18]R. Raghavan, "Risk Management in Banks," CHARTERED ACCOUNTANT-NEW DELHI-, vol. 51, pp. 841-851, 2003.

[19]M. A. Karaolis, J. A. Moutiris, D. Hadjipanayi, and C. S. Pattichis, "Assessment of the risk factors of coronary heart events based on data mining with decision trees," Information Technology in Biomedicine, IEEE Transactions on, vol. 14, pp. 559-566, 2010.