

Precising Multi-Dimensional Data Cube through Association Rule Mining Algorithm

Dr. K.Kavitha

Assistant Professor, Department of Computer Science

Mother Teresa Women's University, Kodaikanal

kavitha.urc@gmail.com

Abstract

Huge range of data mining techniques has made significant improvements in the field of knowledge discovery towards various domains. This paper deals with the works related to the association rule in data mining and Multi-dimensional schema. Various frameworks have suggested exploring a knowledge discovery process which integrates the concept of machine learning, statistical schema and association rules for finding interesting rules. Multidimensional schema and data cubes are constructed for acquiring the rules. This paper highlights the recent works related with Multidimensional data cubes and presented the limitations. Finally performance is evaluated and clearly proven that the proposed work provides better accuracy than existing systems.

Keywords: Association Rules, Datacubes, Data Mining, Multidimensional Schema, Information Gain

INTRODUCTION

A huge range of data mining techniques has made significant improvements to the field of knowledge discovery in various domains. For example in the banking sector, these methodologies are used for loan payment prevision, classification of customers for targeted marketing, customer credit policy analysis, detection of money laundering schemes and other financial crimes. The banking database should control credit management thoroughly. Loan sanctioning requires the usage of huge data and significant processing time. To sanction the loan to customers, the bank needs to take some kind of precautions such as performance of the customer firm by analyzing the previous year's financial statements. The major tool used in multidimensional analysis in a data warehousing is the use of data aggregation and exploratory techniques that forms a part of (OLAP). The traditional OLAP technique is limited to detect hidden association between the items side in a data warehouse. So a lot of research undergoes to extend the OLAP technique to anticipate the future events. In this paper the capability of OLAP is extended to detect the hidden association and forecast the future events based on the historical data driven from the multidimensional schema.

The rest of the paper is organized as follows. Section II presents a description about the previous research

which is relevant to the design and analysis of multidimensional schema. Section III describes about the existing model DDAR, DCAR, CMDC and its limitations. Section IV concludes this research work.

RELATED WORK:

Herawan and Deris presented an alternative approach for mining regular association rules and maximal association rules from transactional datasets. The transactional dataset was transformed into a Boolean-valued information system. The notions of regular and maximal association rules between two sets were defined. Also, support, confidence, maximal support, maximal confidences were defined using soft set theory [6]. *Kumar and A. Chadha* presented a case study of a university that hopes to improve the quality of education. So that association rule discovery was used to predict the accurate results [7]. *Romero, et al* explored the extraction of rare association rule from a Moodle system. Some relevant results were obtained and compared with the rare association rule mining algorithm [8]. *Zhu and Li* presented an association rule mining algorithm Ex- Apriori which was based on the predicate path graph. The algorithm can produce the predicate path graph by scanning database only once and dig out the frequent pattern based on the frequent predicate path graph. It avoids the shortcoming of scanning database many times [9].

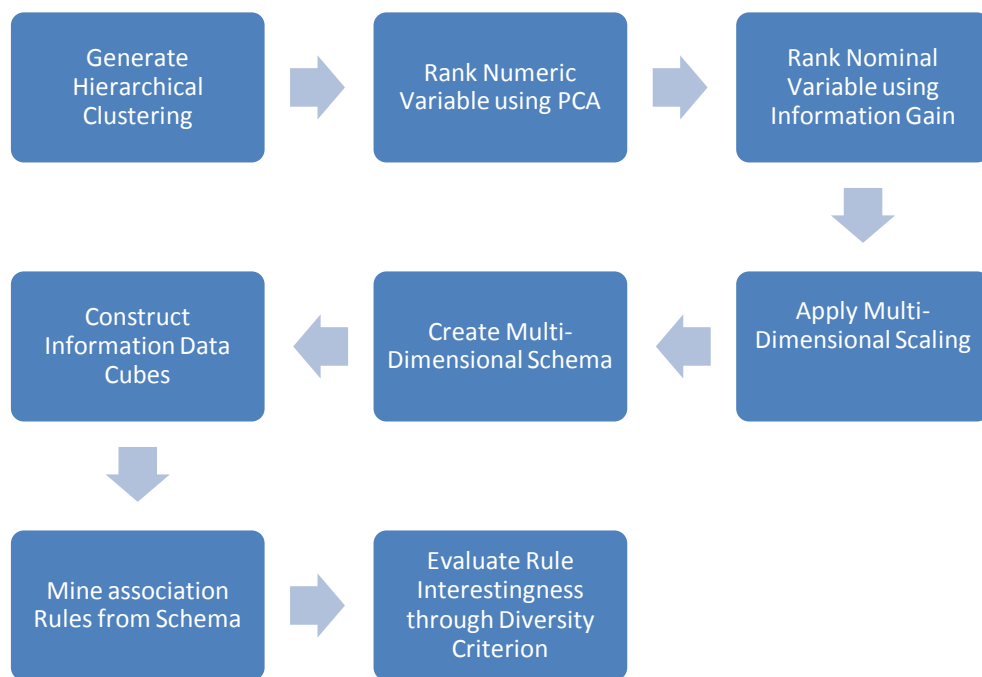
Manda et al proposed an approach called Multi-ontology data mining at All Levels (MOAL) which uses the structure and relationships of the Gene Ontology (GO) to mine multi-ontology multi-level association rules. In this paper, two interesting measures were introduced i.e. Multi-ontology Support (MOSupport) and Multi-ontology Confidence (MOConfidence) customized to evaluate multi-ontology multi-level association rules [18].

OVERVIEW OF MULTI-DIMENSIONAL DATA CUBES:

DDAR

Discovery diverse association rule inferred as DDAR. This framework explores knowledge discovery

process which takes the integration of machine learning, statistical schema and association rules. To find out the interested rules in large dimensional data cubes. Initially, hierarchical clustering strategy is applied to the dataset to classify the dataset into smaller fragments. Later, Numerical and Nominal attributes undergo the ranking procedure. Hence the Principal Component Analysis is used to rank the attributes. Based on facts and dimensions, Multi-dimensional schema is generated with fact and dimensions. The data cubes are constructed based on the association rules. Information gain strategy is used to extract Information Datacubes[1].



Graph 1:

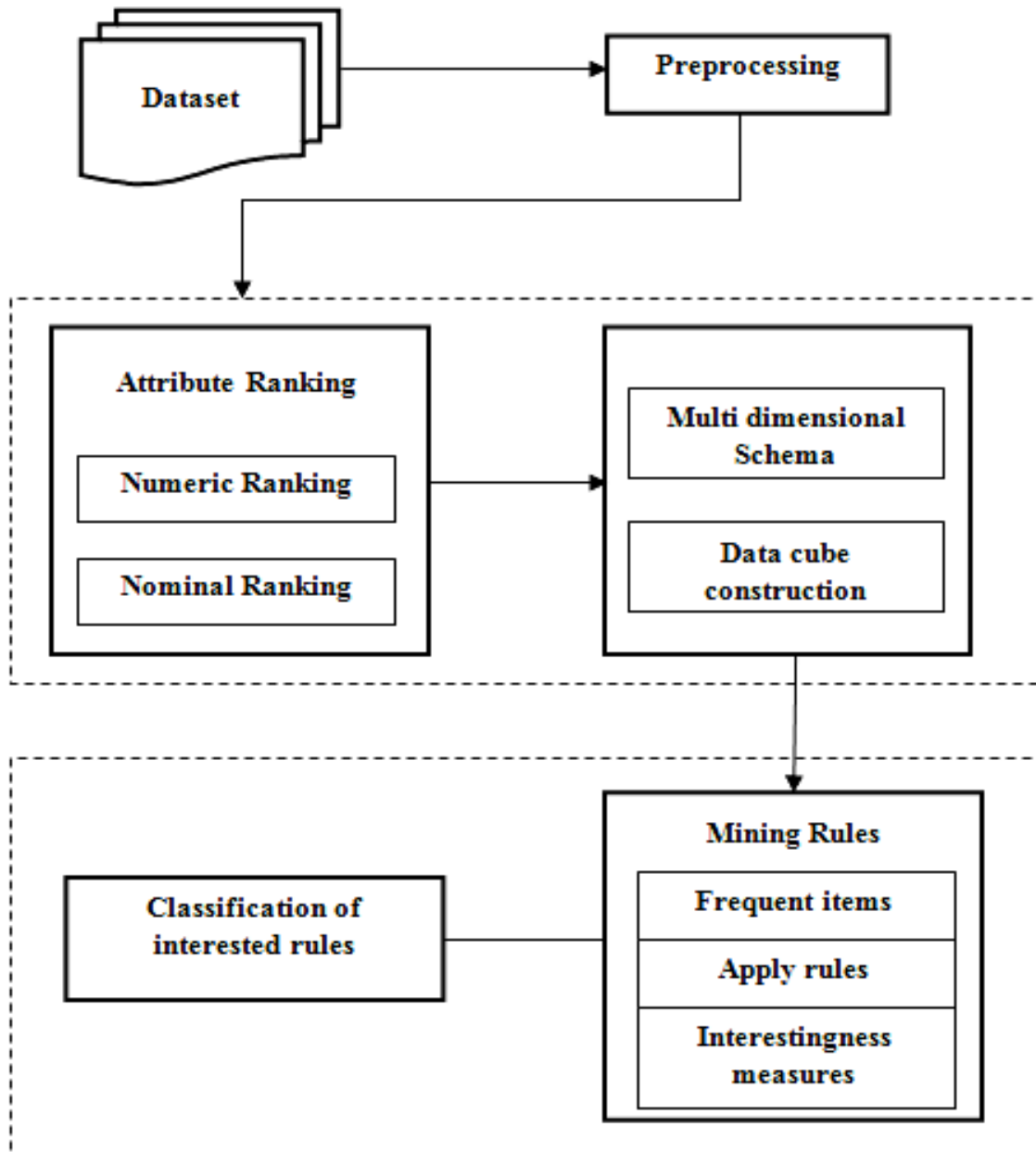
Limitations

Takes more time to complete the datacube construction
Discovered rules are not categorized based on interesting measure.

DCAR

Data cube Association Rule is termed on DCAR. Data mining and statistical techniques are occupied with PCA to rank the facts and dimensions. Attributes are

ranked based on nominal and numerical attributes. Information gain is utilised to rank the numerical attributes. Highly ranked dimensions and facts discover interesting information nested in multi-dimensional cubes. Datacube is constructed using highest ranked dimensions and facts. Present in multi-dimensional scheme. The mine the association rules based on the importance among the rules form the schema [22].



Graph 2:

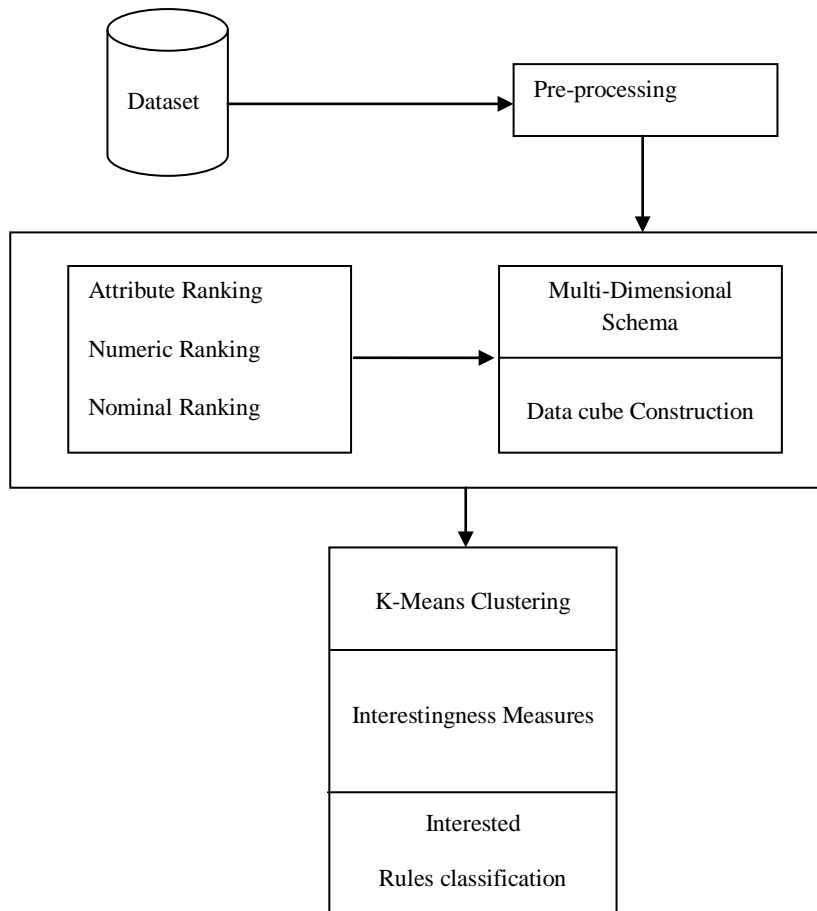
Limitations

- Discovered rules are categorized but it leads complexity
- Time complexity increases
- Prediction accuracy is lesser
- Objective diverse measures are lesser

CMDC

This section presents an overview of Multi-dimensional schema formation which suggests the discovery of Association Rules. Exploring knowledge discovery process by integrating machine learning and statistical scheme. Highly ranked dimensions and

fact results in the discovery of interesting information nested in multi-dimensional cubes. The following figure shows the methodology for classifying Multi-Dimensional Data cube. The real World Bank loan dataset is used to diverse association rules. The dataset is initially performing pre-processing stages to attain the highest quality of the dataset. It removes the unwanted data in the dataset. The proposed system does not depend on hierarchical structure. It applies k-means clustering techniques for grouping the data which provides better results than other existing methods.



Graph 3:

Limitations

- Time complexity still increases
- Clustering leads memory complexity

CONCLUSION

Risk prediction and classification is the crucial task in real world datasets. This paper presents the related work towards multidimensional data cube and Association Rule classification. Finally, experiments is carried out for related works and proven that proposed work is better than existing methods. In future, the proposed method is implemented in real time dataset and prove the accuracy efficiently.

REFERENCES

1. M. Usman, R. Pears, and A. Fong, "Discovering diverse association rules from multidimensional schema," 2013.
2. R. Pears, M. Usman, and A. Fong, "Data guided approach to generate multi-dimensional schema for targeted knowledge discovery," 2012.

3. G. Liu, H. Jiang, R. Geng, and H. Li, "Application of multidimensional association rules in personal financial services," in Computer Design and Applications (ICCD), 2010 International Conference on, 2010, pp. V5- 500-V5-503.
4. W.-Y. Chiang, "To mine association rules of customer values via a data mining procedure with improved model: An empirical case study," Expert Systems with Applications, vol. 38, pp. 1716-1722, 2011.
5. M. A. Domingues and S. O. Rezende, "Using taxonomies to facilitate the analysis of the association rules," arXiv preprint arXiv:1112.1734, 2011.
6. T. Herawan and M. M. Deris, "A soft set approach for association rules mining," Knowledge-Based Systems, vol. 24, pp. 186-195, 2011.
7. V. Kumar and A. Chadha, "Mining Association Rules in Student's Assessment Data," International Journal of Computer Science Issues, vol. 9, pp. 211-216, 2012.

8. C. Romero, J. R. Romero, J. M. Luna, and S. Ventura, "Mining Rare Association Rules from e-Learning Data," in EDM, 2010, pp. 171-180.
9. H. Zhu and Q. Li, "An Algorithm Based on Predicate Path Graph for Mining Multidimensional Association Rules," in Proceedings of the 2012 International Conference on Information Technology and Software Engineering, 2013, pp. 783-791.
10. C.-A. Wu, W.-Y. Lin, C.-L. Jiang, and C.-C. Wu, "Toward intelligent data warehouse mining: An ontology-integrated approach for multi-dimensional association mining," Expert Systems with Applications, vol. 38, pp. 11011-11023, 2011.
11. J. K. Chiang and H. Sheng-Yin, "Multidimensional data mining for healthcare service portfolio management," in Computer Medical Applications (ICCM), 2013 International Conference on, 2013, pp. 1-8.
12. P. Allard, S. Ferré, and O. Ridoux, "Discovering Functional Dependencies and Association Rules by Navigating in a Lattice of OLAP Views," in CLA, 2010, pp. 199-210.
13. W. Moudani, M. Hussein, M. Moukhtar, and F. Mora-Camino, "An intelligent approach to improve the performance of a data warehouse cache based on association rules," Journal of Information and Optimization Sciences, vol. 33, pp. 601-621, 2012.
14. M. Usman and S. Asghar, "An Architecture for Integrated Online Analytical Mining," Journal of Emerging Technologies in Web Intelligence, vol. 3, pp. 74-99, 2011.
15. H. Uğuz, "A two-stage feature selection method for text categorization by using information gain, principal component analysis and genetic algorithm," Knowledge-Based Systems, vol. 24, pp. 1024-1032, 2011.
16. W. Abdelbaki, S. B. Yahia, and R. B. Messaoud, "NAP-SC: A Neural Approach for Prediction over Sparse Cubes," in Advanced Data Mining and Applications, ed: Springer, 2012, pp. 340-352.
17. J. Nahar, T. Imam, K. S. Tickle, and Y.-P. P. Chen, "Association rule mining to detect factors which contribute to heart disease in males and females," Expert Systems with Applications, 2012.
18. P. Manda, F. McCarthy, and S. M. Bridges, "Interestingness measures and strategies for mining multi-ontology multi-level association rules from gene ontology annotations for the discovery of new GO relationships," Journal of biomedical informatics, 2013.
19. H. R. Qodmanan, M. Nasiri, and B. Minaei-Bidgoli, "Multi objective association rule mining with genetic algorithm without specifying minimum support and minimum confidence," Expert Systems with applications, vol. 38, pp. 288-298, 2011.
20. C.-H. Weng and Y.-L. Chen, "Mining fuzzy association rules from uncertain data," Knowledge and Information Systems, vol. 23, pp. 129-152, 2010.
21. N. Zbidi, S. Faiz, and M. Limam, "On mining summaries by objective measures of interestingness," Machine learning, vol. 62, pp. 175-198, 2006.
22. K Kala Dr. E.Ramaraj "DCAR: A Novel Approach for Datacubes Association Rule Algorithm in Multidimensional Schema", CiiT International Journal of Data Mining and Knowledge Engineering, Vol 6, No 07, August 2014