ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

Performance Analysis for DCAR and CMDC Using Shopping Complex Dataset

Dr. K.Kavitha

Assistant Professor, Dept. of Computer Science, Mother Teresa Women's University, Kodaikanal, India

ABSTRACT: Data Mining extracts knowledge automatically based on the abilities of data analysis, decision making and prediction. The main aim of data mining is to provide support for decision making by detecting useful patterns from large voluminous datasets. Large numbers of association rules are often produced by association rule mining algorithm. Sometimes, it can be very difficult for decision makers to understand the rules and find useful knowledge source apply to the Business process. Predicting financial information is very difficult task. To improvise the shopping complex dataset strategy, DCAR and CMDC algorithms are implemented in shopping complex dataset and finally experimental results are carried out. It is clearly proven that the proposed CMDC provides better accuracy than existing algorithm.

KEYWORDS: Association Rule, Multi-Dimensional Datacube, OLAP, Clustering, Confidence, Threshold

I. INTRODUCTION

Data mining facilitates the discovery of unrevealed trends from large voluminous data sets. Data warehousing provides an interactive analysis of data through the use of different data aggregation methods. Data warehousing contributed key technology for complex data analysis, automatic extraction of knowledge from wide data repositories and decision support. Recently, there has been an increased research going on to integrate those two technologies. Moreover this work is concentrated on applying the data mining technique as a front end technology to a data warehouse to extract trends and rules from the data repository in data warehouses.

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.

In this paper, a novel approach called datacubes association rule algorithm is proposed across multidimensional datacubes. This method is based on information gain to detect and rank the most informative dimensions among the nominal variables. The application of Principal Component Analysis extracts the informative datacubes at different level of data abstraction. With the help of objective measures interesting rules are discovered.

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 DCAR, CMDC and its limitations. Section IV involves the detailed description about the proposed method with shopping complex dataset. Section V presents the performance analysis. This paper concludes in Section VI.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

II. RELATED WORK

This section deals with the works related to the association rules in data mining and multidimensional schema. *Usman, et al* proposed a methodology that selects a subset of informative dimension and fact variables from an initial candidate set. The experimental results of this method were conducted on three real world datasets, which was extracted from the UCI machine learning repository. The knowledge discovered from the schema was more diverse and informative than the standard approach of the original data [1]. *Pears, et al* presented a generic methodology which incorporated semi-automated knowledge extraction methods to provide data-driven assistance towards knowledge discovery. A binary tree of hierarchical clusters were constructed and numeric variables were annotated. Three case studies were performed three real-world datasets from the UCI machine learning repository in order to validate the generality and applicability [2].

III. PROPOSED METHODOLOGY

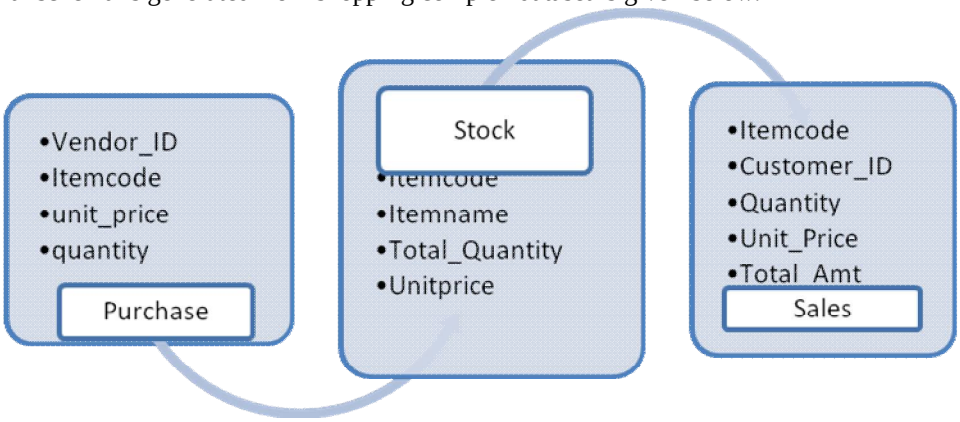
Application of Data mining techniques further yields the value of data warehouse by converting valuable data into assets for future tactics and strategic business enhancement. MIS provides advanced capability which gives the user to arise sophisticated and pertinent queries. It empowers the right people by providing specific needed information. To improvise shopping complex business strategy, association Rule Mining is discovered due to fashion trends in the world. In this research work, informations are collected from Customer, Vendor and Consumer database. To identify the similarity of Purchase and Sales points, Association Rule Mining is applied.

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 discovers 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].

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 perform 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[23]. Multidimensional schema is generated from shopping complex dataset is given below.



Facts table represent stock details and Dimensions represents Sales, Purchase, Vendor and Customer details.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

The above multi-dimensional schema contains all the dimension and facts in the shopping complex dataset. It is used to construct informative datacubes. Highly ranked facts and dimensions are used to construct data cube. It helps the Consumer to show the rank positions. It stores combined measures of the dimension values in a multi-dimensional space. OLAP operations are used to explore the data in to meaningful information. Here facts and dimensions are separated based on the schema formation. In shopping complex database, Sales, purchase, customer and vendor are dimensional tables and stock kept in fact table.

Most frequent rules are extracted based on the user specified constraint. It finds the interesting relationships and associations among the attribute set effectively. Interested rules are predicted based on the constraint Weight which is to evaluated by support and confidence.

Discovered Rules are classified into highly interested, medium interested and low interested rules. To classify the rules, mean values are fixed based on highly ranked rules dimensions.

Discovered Rules are

Itemset	**Weight**
Shampoo	0.98
Conditioner	0.80
Oil	0.88
Detergent	0.70
Lotion	0.76
Shampoo, Lotion	0.86
Lotion, Oil	0.76
Soap, Oil	0.70
Shampoo, Conditioner	0.88
Lotion, Conditioner	0.80
Conditioner, Lotion, Soap	0.85
Lotion, Oil, Detergent	0.60
Soap, Oil, Lotion	0.56
Shampoo, Conditioner, Comb	0.55

Based on the constraint value such as weight, Highly Ranked Rules are classified separately

IV. EXPERIMENTAL RESULTS

To validate Prediction accuracy and classification of association Rules, experimental dataset is generated from shopping complex dataset. Initially dataset is selected and undergone preprocessing steps to filter irrelevant data. Customers and frequently purchased items are classified based on weight such as interestingness measures. Experimental results are validated based on the metrics such as prediction accuracy and time consumption.

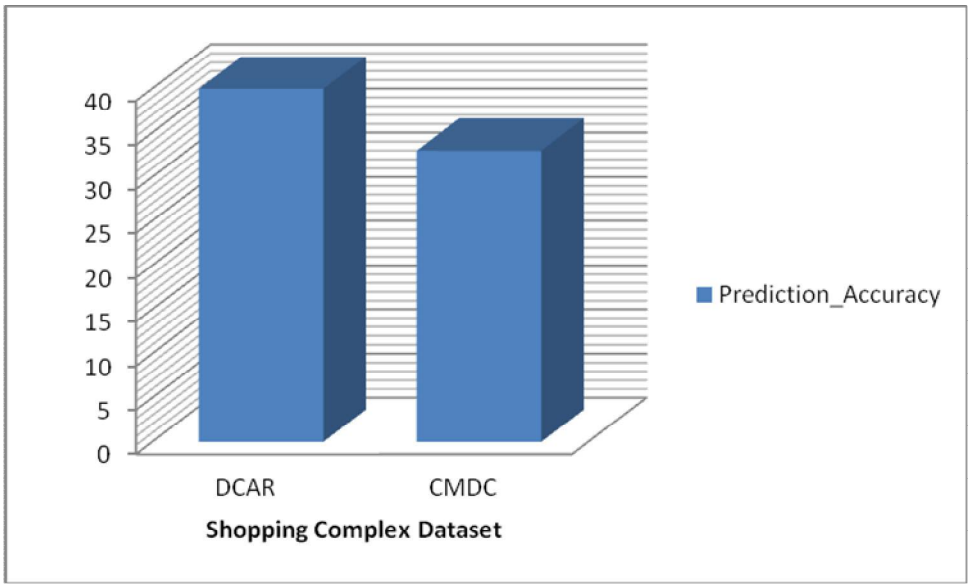
To assess the quality of learned model, Prediction accuracy is observed by the model. Figure shows the prediction accuracy of the rule generated through DCAR and CMDC algorithm which is higher than existing DCAR.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

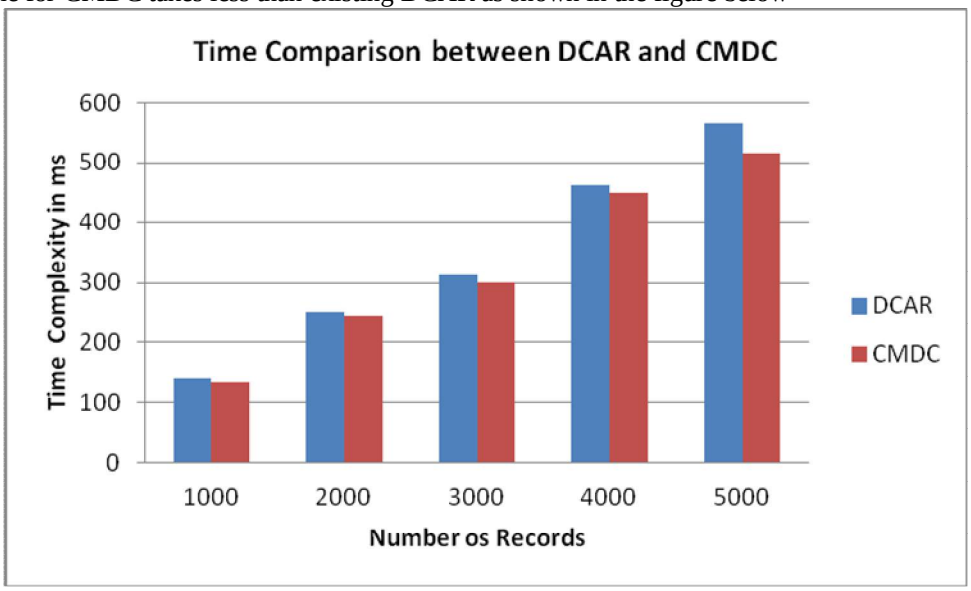
International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016



Computing time for CMDC takes less than existing DCAR as shown in the figure below



V. CONCLUSION

In this paper, Shopping complex dataset is applied in the existing algorithms DCAR and CMDC for evaluating the experimental results to predict accuracy and time complexity. It is clearly stated that CMDC algorithm provides better accuracy than existing algorithm.

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 (ICCDA), 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.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

[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
CiiT International Journal of Data Mining and Knowledge Engineering, Vol 6, No 07, August 2014 281 0974-9683/CIIT–IJ-5341/09/$20/$100 © 2014 CiiT Published by the Coimbatore Institute of Information Technology Applications (ICCMA), 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
[23] Dr. K.Kavitha, "Modified Approach for Classifying Multi-Dimensional Data-Cube Through Association Rule Mining for Granting Loans in Bank",International Journal of Computer Trends and Technology (IJCTT), ISSN: 2231-2803 , Volume 33 Number 1 - March 2016, Page No. 9-13.