Global Journal of Engineering Science and Research Management

PREDICTING BEST TEACHERS BASED ON CLASSIFYING INTERESTING RULES IN MULTIDIMENSIONAL SCHEMA FORMATION

Dr. K.Kavitha*
*Assistant Professor, Department of Computer Science, Mother Teresa Women's University, Kodaikanal.

KEYWORDS: Association Rule, Interesting Measures, Cluster, Multi-Dimensional Data, DataCube..

ABSTRACT
Data mining offers many technologies to analyze and detect the hidden pattern and also convert raw data into useful information. Clustering multi-dimensional data cube is an important technique used to group the related elements without advance knowledge. Identifying the best teachers is an important task of teacher recruitment. Quality education depends largely on teachers performance evaluation and prediction is done based on clustering and decision making approaches. It allows the institution to generate the interested rule classification and determine whether teacher can be recruited or not. This paper concentrates the concept of multi dimensional Association Rule clustering for data prediction.

INTRODUCTION
Data mining plays a major role in knowledge discovery using various domains. Decision making and clustering is an important concept to group the related information and make the better decision accordingly. In education field, quality education is a tedious process to decide. Every institution has its own criteria. Based on the criteria, best teachers has to be selected but it is very difficult to identify the qualified teachers. Information's are collected along with their experiences. The key problem is to extract the information and classification, rule prediction and interesting rule classification to perform the required task. Clustering and multi-dimensional data cube classification provides some additional tools to reduce the outcome uncertainty for enhancing the educational quality. The objective of best teachers recruitment is to evaluate the ranks so that the process makes better decision making. It helps to improve the speed and consistency of recruitment application.

The rest of the paper is organized as follows. follows. Related work is explained in Section 1.The proposed work is presented in Section 2. Section 3 contains the algorithm explanation. Finally, the conclusion is drawn at section4.

RELATED WORK
This section deals with the work related to risk evaluation and association rules. Usman.et.al proposed a method to select a subset of informative dimension and fact identifiers from initial candidate sets. Knowledge Discovered from standard approach for mining original data[1]. Pears.et.al presented a methodology to incorporate semi-automated extraction methods. Binary tree of hierarchical clusters are constructed for each node. This paper presented new method to generates rank for nominal attributes and generate candidate multidimensional schema[2].

Liu.et.al recommended a system to generate association rules for personal financial schemas. The data cubes generated based on financial information and multidimensional rules are generated[3]. Chiang.et.al proposed a model to mine association rules of customer value. Ward's method initiated to partition the online shopping market into three markets. Here Supervised learning is employed to create association rules[4].

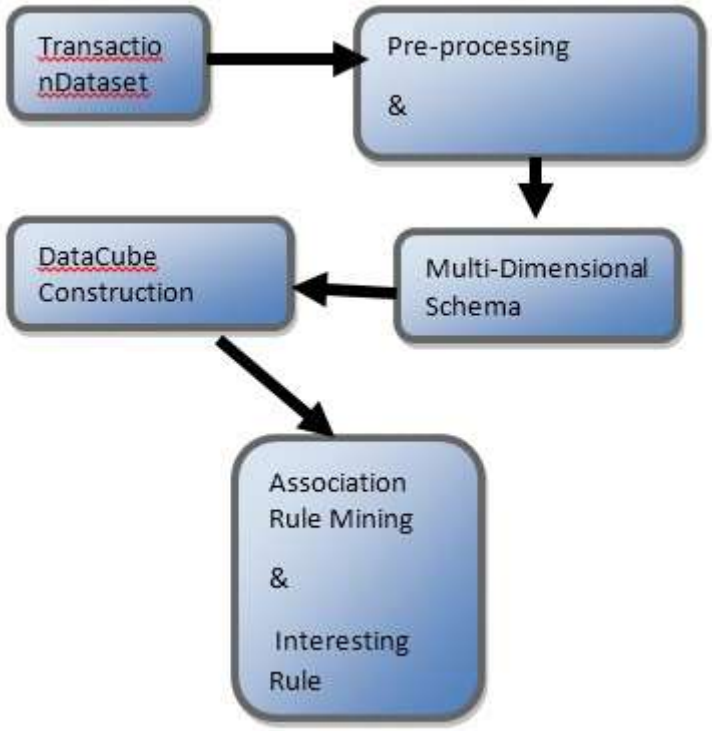
Herawan.et.al presented an approach for regular and maximal association rules from transactional datasets based on soft set theory. Here Transactional datas are transformed into Boolean values based on information system[5].

MULTI-DIMENSIONAL CLASSIFICATION AND ASSOCIATION RULES PREDICTION
Multidimensional schema is generated by merging the benefits of hierarchical and multi-dimensional scaling techniques based on the neighbour cluster, nominal variables and numeric variables are ranked. It symbolizes the relationship between higher dimensional and performance criteria. Highly ranked dimensions and facts can be used to construct a data cube. It helps to provide interesting rule classification techniques.

ISSN 2349-4506
Impact Factor: 2.265

Global Journal of Engineering Science and Research Management

The following figure shows the main steps involved in proposed work. The real world teachers recruitment dataset is used to diversify the association rule. The dataset is initially pre-processed to attain the highest quality of the dataset. Initial step removes the unnecessary portions from the dataset. Existing methods applies clustering techniques to categories the rules. But in the proposed method provides better performance than the clustering technique.



ATTRIBUTE RANKING

To reduce the size of the dataset data mining uses some technologies. It can be classified in to two such as

a. Nominal Attributes
b. Numerical Attributes.

RANKING NUMERIC VARIABLES

Numeric variables are ranked by Principal Components Analysis (PCA) in term of the degree of variance. It transforms a given set of attributes into smaller set of attributes. Here dataset can be reduced with numeric variables. Here teacher's experience and age is considered and represent rank for further processing.

RANKING NOMINAL ATTRIBUTE

Nominal Attributes are ranked by using Information gain. Entropy and information gain measures the correlation. Entropy measures the uncertainty among random variables in the dataset. Information gain is used to select the best attributes splitting criteria. Here teacher's marital_status, academic qualification, category and sex are considered and predicted the ranks.

MULTI-DIMENSIONAL SCALING

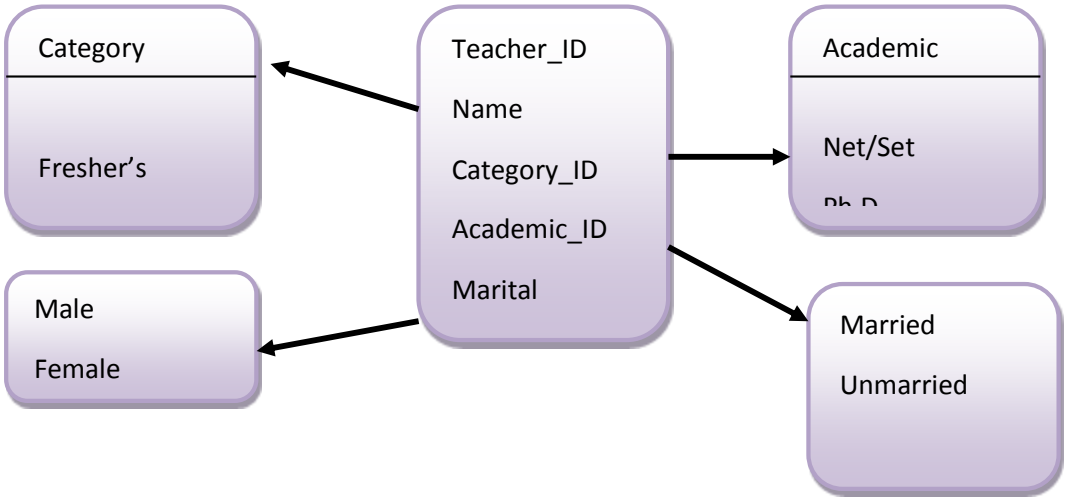
To achieve natural grouping for each of the nominal attributes, Multi dimensional scaling is applied to find the semantic relationship between the calculated values for each nominal attribute. Relationship can be easily visualised by parallel co-ordinates. Based on the similar semantic values, Assigned values are grouped together.

MULTI-DIMENSIONAL SCHEMA FORMATION

Multi-Dimensional schema is constructed by facts(numeric variables) and dimension values(nominal attributes). Grouping defines the hierarchical dimensions. Each dimension in a cluster has group and value level. Teacher's Recruitment dataset is shown in the following figure

ISSN 2349-4506
Impact Factor: 2.265

Global Journal of Engineering Science and Research Management



DATA CUBE CONSTRUCTION

Highly ranked facts and dimensions are used to construct data cube. It helps the user to show the rank positions. The key point in a data cube stores combined measures of the dimension values in a multi-dimensional space. OLAP operations are used to explore the data into meaningful pattern. Data cubes can be constructed by providing any number of dimensions and particular for the Information cubes. Here Facts and Dimensions are separated based on Multidimensional Schema formation. In teacher's recruitment database, Category,academic, sex and Marital_status are Dimension tables and personal information kept in fact table.

ASSOCIATION RULE DISCOVERY SCHEMA

The association rule is defined as $P \rightarrow Q$ where P and Q are disjoint and frequent items in the dataset respectively. Most frequent rules are extracted based on user specified constraints such as threshold limit. It finds the interesting relationships and association among the attribute sets present in the dataset. Interested rules are predicted based on Academic, Category and Marital Status datas.

INTERESTED RULES CLASSIFICATION

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

The proposed algorithm classifies and predicts both certain and uncertain date. Accordingly it combines the uncertain data into rule based mining.The interested rule are classified as low interested (LI), medium(MI) and highly interested rule(HI). The mean value M is calculated based on the importance among the rules. Weight denotes the maximum importance value of the rule.

Input: Calculate_Rule CR, i=1,2,.....,n
Output: LI, MI, HI
Find maximum importance (weight) for CR
Calculate M for CR
M = weight / N
For each $r_i \in CR$
Get $weight_i$ for r_i
if ($weight_i \leq \mu$)
Add r_i into LI
if ($weight_i > \mu$ && $r_i \leq 2\mu$)
Add r_i into MI
if ($weight_i > 2\mu$)
Add r_i into HI
End

The above algorithm categorise the given rule (CR) into three sets LI, MI and HI.

CONCLUSION

Classifying association rules in multi-dimensional data cube is an important task in research Industry. This paper proposed a framework for Association Rule generation where bulk amount of data are engendered and clustering is done based on data mining technique. The proposed method classifiers the interesting rules such as highly ranked, medium ranked and low ranked Interesting Rules. Based on these rule, institution can easily identity the best teachers for improvising educational quality.

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 Multi0Dimensional 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 ,pp V5-500-503
4. 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
5. T.Herawan and M.M.Deris, "A soft set approach for association rule mining Knowledge based Systems", vol 24, pp 186-195,2011.
6. 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
7. D.Zakrzewska, "On integrating unsupervised and supervised classification for credit risk evaluation", Information Technology and Control, Vol. 36, pp. 98-102, 2007.
8. M.L.Bhasin, "Data Mining: A Competitive tool in the Banking and Retail Industries" Banking and Finance, vol 588,2006
9. 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.
10. R. Raghavan, "Risk Management in Banks," *CHARTERED ACCOUNTANT-NEW DELHI-*, vol. 51, pp. 841-851, 2003.
11. 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.
12. M. Anbarasi, E. Anupriya, and N. Iyengar, "Enhanced prediction of heart disease with feature subset selection using genetic algorithm," *International Journal of Engineering Science and Technology*, vol. 2, pp. 5370-5376, 2010.
13. M. Du, S. M. Wang, and G. Gong, "Research on decision tree algorithm based on information entropy," *Advanced Materials Research*,vol. 267, pp. 732-737, 2011.
14. X. Liu and X. Zhu, "Study on the Evaluation System of Individual Credit Risk in commercial banks based on data mining," in *Communication Systems, Networks and Applications (ICCSNA), 2010 Second International Conference on*, 2010, pp. 308-311.
15. B. Azhagusundari and A. S. Thanamani, "Feature selection based on information gain," *International Journal of InnovativeTechnology and Exploring Engineering (IJITEE) ISSN*, pp. 2278-3075.
16. M. Lopez, J. Luna, C. Romero, and S. Ventura, "Classification via clustering for predicting final marks based on student participation in forums," *Educational Data Mining Proceedings*, 2012.