International Journal of Engineering Sciences Paradigms and Researches Vol. 06, Issue 01, July 2013
ISSN (Online): 2319-6564
www.ijesonline.com

Multi Relational Classifications based on Association Rule (MCAR) using RBF Neural Network

Swati Khare[1], Anju Singh[2]

[1]Pursuing M. Tech., Computer Science and Engineering, Barkatullah University
Bhopal, Madhya Pradesh, India
Swati_soni12@rediffmail.com

[2]Assistant Professor, Information and Technology, Barkatullah Univercity
Bhopal, Madhya Pradesh, India
Asingh0123@rediffmail.com

Abstract

Data classification is a demanding task in the area of machine learning. Now a days in every field such as online data dispensation, pattern recognition, pattern classification, medical diagnosis are required classification of data. MCAR employs a novel data structure, association rule, to compactly store and efficiently retrieve a large number of rules for classification. Association rule is a prefix rule structure to explore the sharing among rules, which achieves substantial compactness. We use classification using association rules not only to solve classification problems, but also to compare the quality of different association rule mining approaches. In this context we show that the quality of rule sets from the standard algorithm for association rule mining can be improved by using a different association rule mining strategy. In this paper we used wine data set provided by UCI machine learning website. Proposed approach implement in mat lab 7.8.0. Matlab is high computational and algorithm based computational software.

Keywords*: Associative Classification, MRDM, MCAR, RBF.*

1. Introduction

Classification is an important subject in data mining and machine learning, which has been studied extensively and has a wide range of applications. Classification based on association rules, also called associative classification, is a technique that uses association rules to build classifier. Generally it contains two steps: first it finds all the class association rules (CARs) whose right-hand side is a class label, and then selects strong rules from the CARs to build a classifier. In this way, associative classification can generate rules with higher confidence and better understand-ability comparing with traditional approaches. Thus associative classification has been studied widely in both academic world and industrial world, and several Effective algorithms [3,6] have been proposed successively. However, all the above algorithms only focus on processing data organized in a single relational table. In practical application, data is often stored dispersedly in multiple tables in a relational database, this leads to the evolution of multi-relational data mining (MRDM)[20].

Multi-relational data mining learns the interesting patterns directly from multiple interrelated tables with the support of primary key /foreign keys. Simply converting multi-relational data into a single flat table may lead to the high time and space cost, moreover, some essential semantic information carried by the multi-relational data may be lost. Thus the existing associative classification algorithms cannot be applied in a relational database directly.

We propose a novel algorithm, MCAR (based on MRDM), for associative classification which can be applied in multi-relational data environment. The main idea of MCAR is to mine relevant features of each class label in each table respectively, and generate strong classification rules. By relevant features, we mean two kinds of frequent close item sets: single table item sets in the target table and cross table item sets in non-target tables. Experiment results show that the above two kinds of item sets have contained sufficient relevant features of class labels. Then we breadth-firstly generate strong classification rules from these item sets with a pruning strategy used in this step. After that, a classifier can be easily built to predict unseen objects' class

labels. Multi-Relational Classification (MRC/RC)[9], which focuses on classification from relational databases comprising multiple tables, is one important task in multirelational data mining (MRDM/RDM)[4,7] and widely uses in many disciplines. That is to say, RC need not to transform multi-tables into a single data table, which effectively avoid these problems [4,7] of relational information loss, statistical skew and efficiency reducing that often happen in propositional or attribute-value classification approaches. Representation is a fundamental as well as a critical aspect in data mining. According to the differences in knowledge representation, the paper divides RC into three main categories that are ILP-based MRC (LBRC), graph based MRC (GBRC) and relational database-based MRC (RBRC). LBRC is a traditional MRC technology. It can state quite complicated relational patterns and is easy to use valid background (domain) knowledge for inductive inference. GBRC uses graphs to provide a more natural means for expressing real-world data. RBRC mainly includes selection graphs based RC and RC by tuple ID propagation, where the first can directly RC through database operation and need not to transform into other knowledge form and the second is being RC through virtually joins among relational tables. CRBF models are creating for data training for minority and majority class data sample for processing of associative classification. The input processing of training phase is combination of SMOTE and CMTNN sampling technique for classifier. While single-layer RBF networks can potentially learn virtually any input output relationship, RBF networks with single layers might learn complex relationships more quickly. The function neCrf creates cascade-forward networks. For example, a cascaded layer network has connections from layer 1 to layer 2, layer 2 to layer 3, and layer 1 to layer 3. The cascade layer network also has connections from the input to all cascaded layers. The additional connections might improve the speed at which the network learns the desired relationship. CRBF artificial intelligence model is similar to feed-forward back-propagation neural network in using the back-propagation algorithm for weights updating, but the main symptom of this network is that each layer of neurons related to all previous layer of neurons. The rest of paper is organized as follows. In Section 2 discuss related work of associative classification. The Section 3 proposed METHOD for classification. The section 4 discusses experimental result and finally followed section 5 conclusion and future scope.

2. Related Work

In this section we discuss some related work for associative classification using rule mining technique using apriori algorithm as [10] describe in the field of data classification as Associative classification (AC) has shown great promise over many other classification techniques on static dataset. However, the increasing prominence of data streams arising in a wide range of advanced application has posed a new challenge for it. The author describes and evaluates AC-DS, a new associative classification algorithm for data streams which is based on the estimation mechanism of the Lossy Counting (LC) and landmark window model. They apply AC-DS to mining several datasets obtained from the UCI Machine Learning Repository and the result shows that the algorithm is effective and efficient. An associative classification approach based on association rules for mining data streams. Empirical studies show its effectiveness in taking advantage of massive numbers of examples. AC-DS's application to a high-speed stream is under way. Describe in the field of data classification [13] as weighted association rule mining reflects semantic significance of item by considering its weight. Classification extracts set of rules and constructs a classifier to predict the new data instance. The author proposes compact weighted associative classification method, which integrates weighted association rule mining and classification for constructing an efficient weighted associative classifier. Compact weighted associative classification algorithm randomly chooses one non class attribute from dataset and all the weighted class association rules are generated based on that attribute. The weight of the item is considered as one of the parameter in generating the weighted class association rules. Weight of item is computed by considering quality of the transaction using link based model. Experimental results show that the proposed system generates less number of high quality rules. The aim of integrating classification and weighted Association rule mining is to address some important requirements arising from modern data mining

processes. The development of weighted associative classification using compact weighted associative classification (CWAC) algorithm, greatly reduce the number of rules in the classifier. The author shows how to generate Compact Weighted Class Association Rules, which may greatly improve the classification accuracy. This plays a vital role in market basket analysis, medical diagnosis and in many other applications. Experimental results show that the proposed Compact Weighted Associative Classification (CWAC) method outperformed the CBA method. This work can be further applied for more number of benchmark datasets. The author proposed a CMAR (Classification based on Multiple Class-Association Rules)[14] and Adriano Veloso proposed Lazy Associative Classifier algorithm for Small Disjunction mining. In addition, we collocate with a new weight calculation method in our algorithm to solve weight bias problem of CMAR. The author uses UCI 26 data set for experiment on our proposed algorithm. The finally results convincingly demonstrated that our proposed algorithm is high accuracy. The author proposed a CMAR algorithm, and successfully integrates support calculation method of LAC algorithm. The experiment proves that using support calculation method of LAC algorithm to obtain Small disjunction rule. After combining with multiple associative-rule method, we can effectively raise the accuracy. On the other hand, the author proposed a new weight calculation method not only greater than CMAR and LAC algorithm but also greater than L3 algorithm 1.01%. In addition, the author proposed an algorithm usually has high accuracy whether the number of rules or number of candidates are high limitation or low limitation. Describe in the field of data classification[15] as Spatial Co-location patterns are similar to association rules but explore more relying spatial auto-correlation. They represent subsets of Boolean spatial features whose instances are often located in close geographic proximity. Existing co-location patterns mining researches only concern the spatial attributes, and few of them can handle the huge amount of non-spatial attributes in spatial datasets. Also, they use distance threshold to define spatial neighborhood. While many methods to facilitate this task have been proposed, few can derive classification rules that involve ranges (numerical intervals)[16]. We consider how range-based classification rules may be derived from numerical data and propose a new method inspired by classification association rule mining. This method searches for associated ranges in a similar way to how associated item sets are searched in categorical attributes in association rule mining, but uses class values to guide the search, so that only those ranges that are relevant to the derivation of classification rules are found. That is, we search for associations among numerical ranges, but our search is guided by class values. This allows accurate range-based classification rules to be derived efficiently. Our preliminary experimental results have shown that the new method is promising. There are a number of issues still to be addressed in future research. First, there is a need to study how good classifiers may be constructed from the discovered classification rules. Second, better data structures and alternative measures may be developed to speed up the computation of associated ranges[11]. The author proposed an association rules mining to the software of examination paper evaluation system, obtaining the useful information which is hidden in the database. It's concluded that the algorithm provides a valuable analysis of information to the examination paper evaluation system. Keywords-association rules. Association rules is the one most important theory in data mining, which have a wide range of applications in the various fields, but, applied to the Evaluation of reliable, it can be said that has just begun, with the mining association rules theoretical the constant further research and using, the rational, efficient and objective analysis of The examination, all from the Association Rules theoretical support. Cascading of neural network model play an important role in data classification and data pattern generation. cascaded model of RBF network for associative classification. The great advantage of RBF network is single layer processing unit and target output independent with input data. In the process of cascading input data passes through train sample, majority function decomposed data into layers such as lower and higher in data space domain. The part of lower used as input in cascaded model.

3. Proposed Method for Classification

By generating rules (and frequent patterns) from association Rules (Apriori algoritm) classification method is applied and in the case

proposed in paper the classification of data which are stored in multi relational tables are classified using RBF neural network and is called Radial basis function CRBF.
CRBF models are creating for data training for minority and majority class data sample for processing of associative classification. The input processing of training phase is combination of SMOTE and CMTNN sampling technique for classifier. While single-layer RBF networks can potentially learn virtually any input output relationship, RBF networks with single layers might learn complex relationships more quickly. The function neCrf creates cascade-forward networks. For example, a cascaded layer network has connections from layer 1 to layer 2, layer 2 to layer 3, and layer 1 to layer 3. The cascade-layer network also has connections from the input to all cascaded layers. The additional connections might improve the speed at which the network learns the desired relationship. CRBF artificial intelligence model is similar to feed-forward back-propagation neural network in using the back-propagation algorithm for weights updating, but the main symptom of this network is that each layer of neurons related to all previous layer of neurons. Tan-sigmoid transfer function, log - sigmoid transfer function and pure linear threshold functions were used to reach the optimized status.

3.1 Steps of associative Classification using RBF neural network*:*

1. Perform Association Rule mining to given data.
2. By using Apriori Algorithm find frequent items from the given dataset.
3. This frequent item generation is done with the help of user defined support value.
4. After frequent item generation create rules with the help of given confidence threashold.
5. Now after rule generation classification using RBF is performed.
6. For performing CRBF first data is devided into training and test.
7. The training phase data are passes through SMOTE and CMTNN sampler.
8. The sampling of data passes through CRBF and balanced the data for minority and majority ratio of class.
9. The sampled data assigned to k-type binary class.
10. Binary class data are coded in bit form.
11. If code bit value is single assigned the class value.
12. Else data goes to training phase.
13. Balanced part of training is updated
14. Data are classified.
15. Find accuracy and relative mean Error
16. Exit

Functioning of CRBF model is shown by forllowing flowchart:

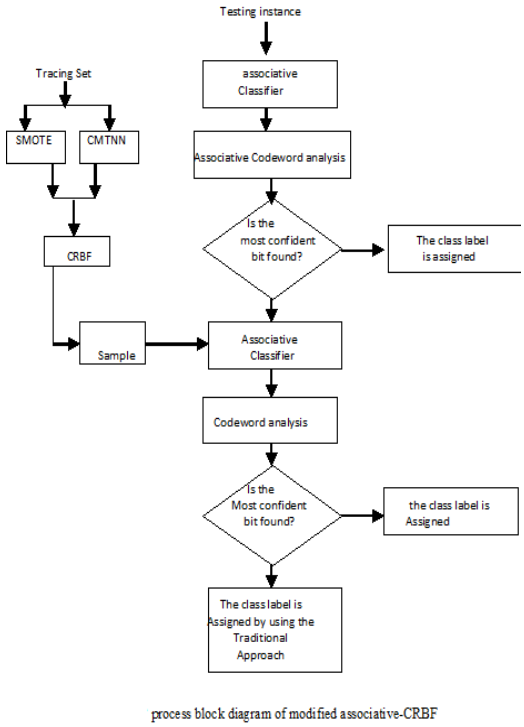

Figure 1 Process block diagram of modified associative – CRBF.

4. Experimental Result

It is simulating on mat lab 7.8.0 and for this work Intel 1.4 GHz Machine is used. MATLAB is a high-level technical computing language and interactive environment for algorithm development, data visualization, data analysis, and numeric computation Matlab is a software program that allows you to do data manipulation and visualization, calculations, math and programming. It can be used to do very simple as well as very sophisticated tasks. Image Processing Toolbox provides a comprehensive set of reference-standard algorithms and

graphical tools for image processing, analysis, visualization, and algorithm development. You can perform image enhancement, feature detection, noise reduction, image segmentation, spatial transformations. As by table , Figure 2 and Figure 3, performance of Multi Relational Classification algorithm shows that when we used MCAR on wine data set then value of runtime is 7.4375 Sec and classification rate accuracy is 84.570%.Performance of multiple relational classification algorithm using Radial function shows that when we used MCAR Using RBF on wine data set then value of runtime is 5.7812 Sec and classification rate accuracy is 91.789%, which is showing that classification rate accuracy increased above 90%.

Table 1 showing the difference of output in MCAR and MCAR-RBF (Proposed method)

dataset	*method*	*support count*	*confidence*	*accuracy (in %)*	*Run Time (in Sec.)*
Wine data set	*MCAR*	*0.3*	*0.5*	*84.570%*	*7.4375 Sec*
	MCAR-RBF (proposed method)	*0.3*	*0.5*	*91.789%*	*5.7812 Sec*

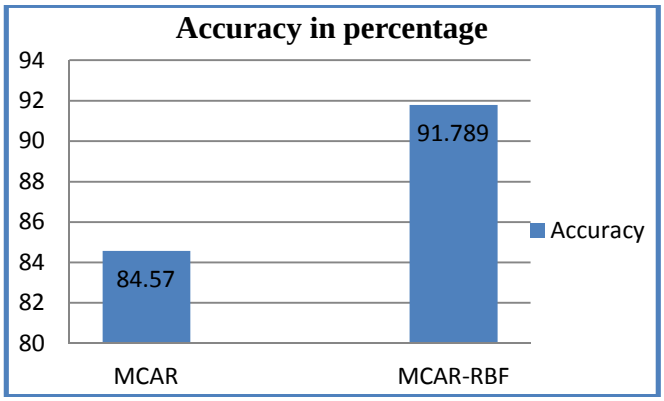

Figure 2 Simulation results with MCAR & MCAR-RBF for Classification Accuracy.

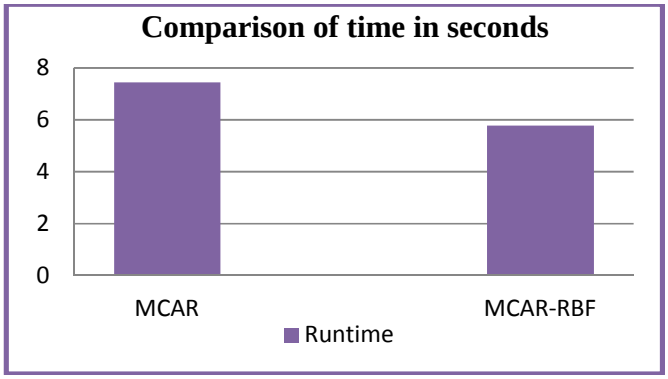

Figure 3 Simulation results with MCAR & MCA -RBF for runtime.

5. Conclusions and Future work

Multi-relational classification is an important subject in data mining and machine learning and it can be widely used in many fields. Novel associative classification algorithm, MCAR, which is the first one as far as we know in the literature to apply associative classification in multi-relational environment. Experimental results show that MCAR gets higher accuracy comparing with the existing multi-relational algorithm. Furthermore, rules discovered by MCAR have a more comprehensive characterization of databases. There are several possible extensions to MCAR. Currently, MCAR uses a support-confidence framework to discover frequent item sets and generate classification rules. It may discover more relevant features of each class label by using related measures extending current framework. Also the current algorithm could be improved in terms of

efficiency by using the optimization technique. Multiple relational classification algorithm modified by RBF so improved rate of classification in comparison of MCAR. In the process of RBF the calculation complexity are increases, the complexity of time are also increases. Our proposed algorithm test wine data set. In this data set the rate of classification is 92%.We also use another data set (abalone data set) and estimate some little bit difference of rate of classification is 91%. The rate of classification increases in previous method on the consideration of time complexity. In future we minimize the complexity of time and also increase the rate of classification using Meta heuristic function such as ant colony optimization, power of swarm (pos) and dendrites cell algorithm.

References

[1] Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth, "From Data Mining to KDD in Databases" pp. 0738-4602, 1996.

[2] Han J. and M. Kamber,"Data Mining: Concepts and Techniques", Academic Press, San Diego, CA, 2000.

[3] W. Li, J. Han, and J. Pei, "CMAR: Accurate and efficient Classification Based on Multiple Class-Association Rules", Proceedings of the ICDM, IEEE Computer Society, San Jose Californi, pp. 369-376, 2001.

[4] S. Dzeroski, N. Lavrac eds. "Relational data mining",Berlin: Springer, 2001.

[5] M.J. Zaki, and C.j. Hsiao, "CHARM: An Efficient Algorithm for Closed Item set Mining", Proceedings of SIAMOD International Conference on Data Mining,, pp. 457-473, 2002.

[6] X. Yin, and J. Han, "CPAR: Classification based on Predictive Association Rules", Proceedings of the SDM, SIAM, Francisco California, 2003.

[7] He J, Liu HY, Du XY."Mining of multi-relational association rules" Journal of Software, 18(11): 2752-2765, 2007.

[8] March 18 - 20, Thair Nu Phyu, "Survey of Classification Techniques in Data Mining" Vol I Imecs 2009, , Hong Kong, 2009.

[9] Zhen Peng, Lifeng and Wu Xiaoju Wang, "Research on Multi-Relational Classification Approaches" pp.3-9 IEEE, 2009.

[10] Zhen- Hui Song and Yi Li, "Associative classification over Data Streams", IEEE, PP.2-10, 2010.

[11] XING Xue, CHEN Yao. WANG Yan-en, "Study on Mining Theories of Association Rules and Its Application"IEEE PP.2-10, 2010.

[12] Xun Zhu1, Hongtao Deng and Zheng Chen "A Brief Review On Frequent Pattern Mining" PP-4-11 IEEE, 2011.

[13] S.P.Syed Ibrahim, K. R. Chandran and M. S. Abinaya "Compact Weighted Associative Classification" IEEE pp.8-11, 2011.

[14] Pei-yi hao, yu-de Chen "a novel associative classification algorithm: a combination of LAC and CMAR with new measure of Weighted effect of each rule group" IEEE pp.9-11, 2011.

[15] You Wan , and Chenghu Zhou " QuCOM: k nearest features neighborhood based qualitative spatial co-location patterns mining algorithm" IEEE pp.8-11, 2011.

[16] Achilleas Tziatzios and Jianhua Shao, Grigorios Loukides "A Heuristic Method for Deriving Range-Based Classification Rules" IEEE pp.6-11, 2011.

[17] March 3 Rupali haldulakar, prof. Jitendra agrawal "Optimization of Association Rule Mining through Genetic Algorithm" (IJCSE) Vol. 3 No, 2011.

[18] Yingqin Gu1,2, Hongyan Liu3, Jun He1,2, Bo Hu1,2 and Xiaoyong Du1,2 "A Multi-relational Classification Algorithm based on Association Rules" pp.4-9 IEEE, 2009.

[19] February 17 & 18, B.N. Lakshmi and G.H. Raghunandhan, "A onceptual Overview of Data Mining" Proceedings of the National Conference on Innovations in EmergingTechnology-2011 Kongu Engineering College, Perundurai, Erode, Tamilnadu, pp.27-32. India, 2011.

[20] March, M. Thangaraj, C.R.Vijayalakshmi, "Performance Study on Rule-based Classification Techniques across Multiple Database Relations", International Journal of Applied Information Systems (IJAIS) – ISSN: 2249-0868 Foundation of Computer Science FCS, New York, USA Volume 5, 2013.