ISSN : 2278 – 1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 1, Issue 7, September 2012

Multi-relational Bayesian Classification through Genetic Approach

Nitin Kumar Choudhary, Gaurav Shrivastava, Mahesh Malviya

Dept.of CSE, RKDF Institute of Science & Technology, Bhopal, India

ABSTRACT: 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. CMAR 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. To speed up the mining of complete set of rules, CMAR adopts a variant of recently developed FP-growth method. FP-growth is much faster than Apriori-like methods used in previous association-based classification, such as especially when there exist a huge number of rules, large training data sets, and long pattern rules. 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 Above classification rate is 80%(MAX) hence the 20% data are unclassified. This is a challenge in the field of data classification. In this paper, we used multiple relational Bayesian classification algorithm based on genetic algorithm used for optimization of classification rate, generated by association rule.

Keywords: Classification, Genetic algorithms, association rule

I. INTRODUCTION

Classification [1] is a significant technique for studies extensively and has wide range of applications in the field of data mining and machine learning. Classification has aim to discover a set of Association mining rules in the database that that satisfy some minimum support and minimum confidence constraints and forms an accurate classifier. Associative classification based on association rules is a procedure that uses association rules to build classifier. Usually it includes 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 fashion, associative classification can generate rules with higher confidence and better support with conventional approaches. Thus associative classification has been deliberate widely in both academic world and industrial world, and numerous useful algorithms [2, 3] have been proposed sequentially. Whereas the proposed methodology till now focusing on organization of data on a single relational table. But in real life problem, data is distributed in multiple tables in a relational database. However transformation procedure of multi-relational data into a single flat table having high time and space complexity, moreover, some crucial

Semantic information carried by the multi-relational data may be lost. problem in the field of data classification: Initially used are simple classifier KNN, the rate of classification of KNN Is only 78% so it is called lazy classifier. Naive bayes Classification replaces a KNN classification technique, the rate of classification is just 80%.in this method data are Unclassified 20%, so this method is not up to the mark for the use of classification. In place of KNN and Naïve bayes some other classifier are available but the rate of classification varies 80 to 90% but not ahead 90%.these are decision tree, SVM, CBA and also association classification. These are limitation and problem of classification algorithm. This paper proposed a methodology based on genetic algorithm for the optimization of classification rate of association classification and results are improved because heuristic functionality of genetic algorithm. Proposed methodology gives an optimal result and adopts optimization of classification of association rule with the help of genetic algorithm and Bayesian approach .

II. RELATED WORK

Since long time before research in field of classification association rules mining from relational data is carried out. By R. Agrawal, T. Imielinski, and A. Swami [4] in 1993 Association rule mining was first introduced. After that many algorithms have been proposed and developed. In 2010, Zhen- Hui Song &Yi Li [5] introduce associative classification approach based on association rules for mining data streams and introduce AC-DS, a new associative classification algorithm for data streams ,based on the evaluation method of the Lossy Counting (LC) and landmark window model. In 2011 S.P. Syed Ibrahim,K. R. Chandran,M. S. Abinaya [6] proposed Compact Weighted

ISSN : 2278 – 1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 1, Issue 7, September 2012

Class Association Rules (CWAC), which may greatly improve the classification accuracy by greatly reduce the number of rules in the classifier and generates less number of high quality rule . In 2011 Pei-Yi Hao, Yu-De Chen [7] proposed a CMAR algorithm that successfully integrates support calculation method of LAC algorithm by obtain Small disjunction rule. After combining with multiple associative-rule method the accuracy is effectively raised.In 2009 Jun He1,Bo Hu1,and Xiaoyong Du introduce an a Multi-Relational Classification Algorithm based on Association Rule[8]: Describe in the field of data classification as MCAR has higher accuracy and better understand ability comparing with a typical existing multiple relational classification algorithm. MCAR uses a support-confidence framework rules. The rate of classification is only 88%. This algorithm will not output an accurate result.

III. APPROACH USED

There are some limitation and problem of classification algorithm. now we choose association classification for classification algorithm and we used genetic algorithm along with probabilistic Bayesian approach for the optimization of classification rate of association classification. In our case the results are improved because the genetic algorithm is a heuristic function. The heuristic function gives an optimal result. Whereas Bayesian approach apply over historical data and give better result. Now we adopted optimization of classification of association rule with the help of genetic algorithm and Bayesian approach

IV. PROPOSED ARCHITECTURE

In this paper we have proposed thee tier architecture. As shown in figure 1 we apply the aprori algorithm in order to generate the candidate set. After that we use the minimum threshold of confidence and support. This minimum threshold used in genetic algorithm . with the use of genetic algorithm we will generated over the population of candidate set in order to calculate the association rules. These association rules may have the ambiguity. Here we need to optimize the association rule. So that the Bayesian approach use to optimized the rules.

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, CMAR, 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 CMAR gets higher accuracy comparing with the existing multi-relational algorithm. Furthermore, rules discovered by CMAR have a more comprehensive characterization of databases. There are several possible extensions to CMAR. Currently, CMAR 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 techniqueProposed Methodology build a hybrid model through classification association and genetic algorithm to increase the data classification rate in order to modify Multi-Relational association rule classification. Higher degree of classification rate leads better classification. So our proposed algorithm gives higher classification rate. Our proposed work is dividing into two parts-:

- For finding frequent item set and candidate key – we used Aprior algorithm
- For Rule generation and optimization- we used genetic algorithm along with Bayesian approach.

Our experiment result shows that our approach makes a significant improvement in classification

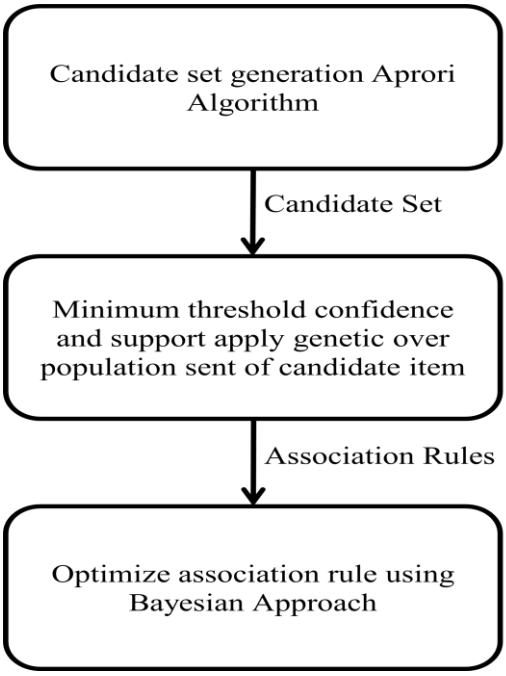

Figure 1: thee tier architecture

The above architecture provide us the three execution environment so that we called it the thee tier architecture.

ISSN : 2278 – 1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 1, Issue 7, September 2012

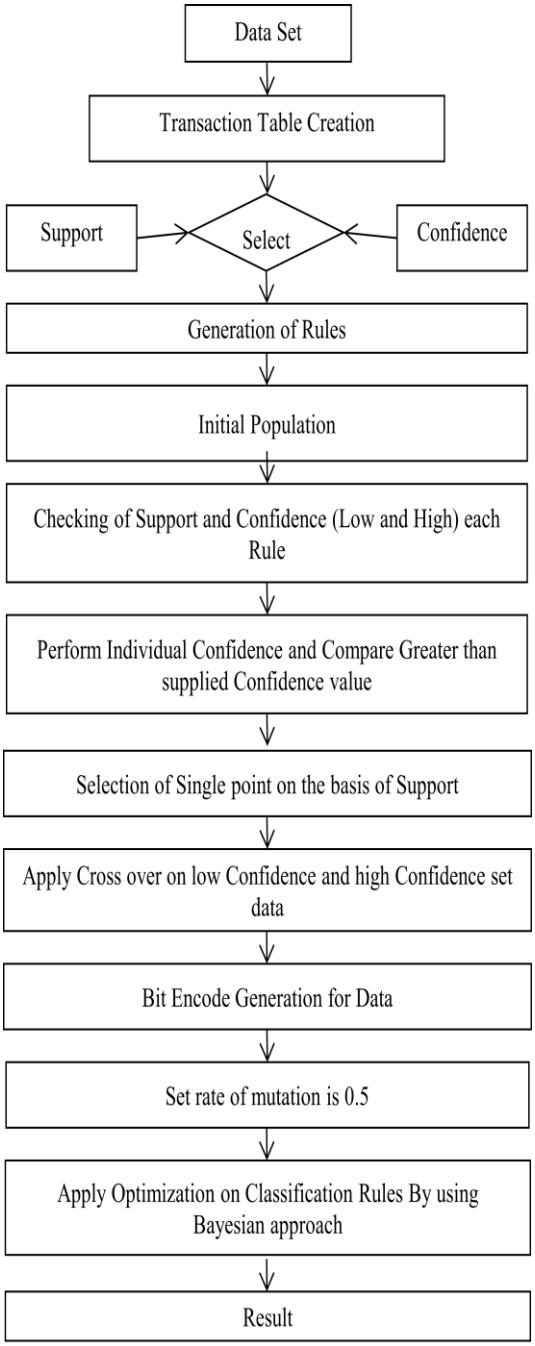

Figure 2: Proposed Architecture

Performance of multiple relational classification algorithm shows that when we used CMAR on wine data set then value of runtime is 2.30881 Sec and classification rate accuracy is 82.53220%.Performance of multiple relational classification algorithm using Genetic algorithm shows that when we used CMAR Using GA/Bayesian on wine data set then value of runtime is 3.02642 Sec and classification rate accuracy is 90.2046%, which is showing that classification rate accuracy increased above 90%.

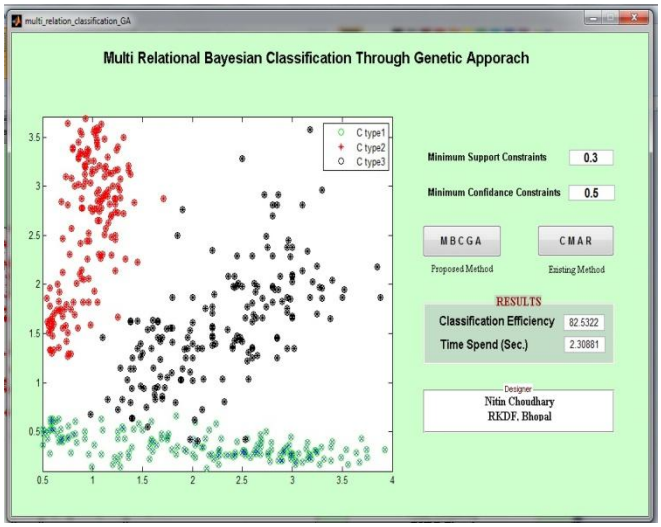

Figure 4: Result of CMAR

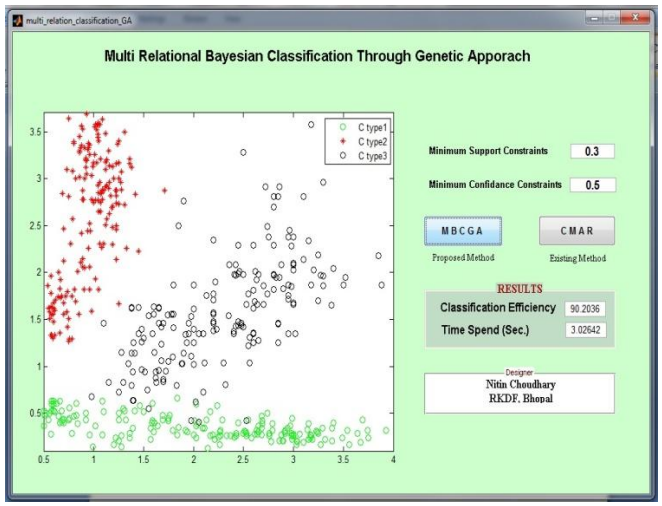

Figure 5: Result of MBCGA

V RESULT AND ANALYSIS

.

ISSN : 2278 – 1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 1, Issue 7, September 2012

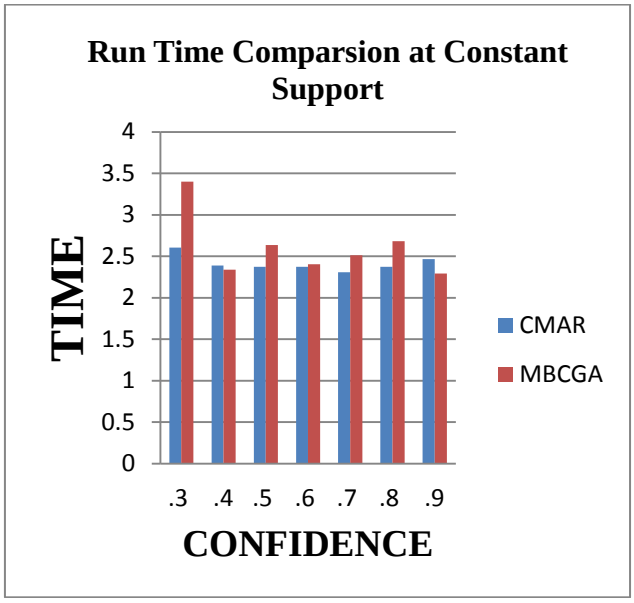

Figure 12 Comparison Graph for Time

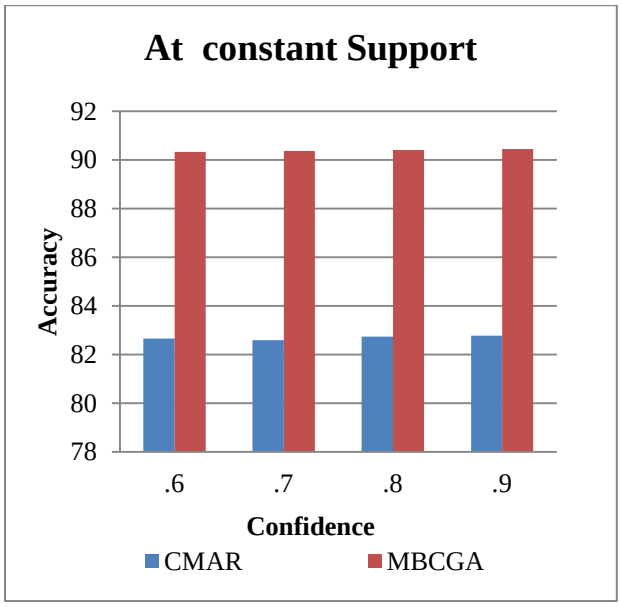

Figure 12 Comparison Graph for Accuracy

There are two graph shown the time and accuracy with respect to variable confidence and constant support. It seems to us that we may need more time as compare to existing technique but the accuracy is always high. so the time will considerable. The results shows that the proposed technique is better than the existing technique,

VI. CONCLUSION

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, CMAR, 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 CMAR gets higher accuracy comparing with the existing multi-relational algorithm. Furthermore, rules discovered by CMAR have a more comprehensive characterization of databases. There are several possible extensions to CMAR. Currently, CMAR 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.

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 dentrites cell algorithm.

VII. ACKNOWLEDGMENT

I would like to say thanks to my guide “ Mahesh malviya” who gives their knowledge and time in order to complete this paper. This paper will never complete without the support facility member CSE department RKDF IST Bhopal.

REFERENCES

[1]. 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 2009 IEEE.

[2]. 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 California, 2001, pp. 369-376.

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

[4]. Agrawal R., T. Imielinski, A. Swami, “Mining association rules between sets of items in large databases”, in Proceedings of the ACM SIGMOD Conference on Management of Data. pp. 207-216' 1993.

[5]. Zhen- Hui Song &Yi Li, “Associative classification over Data Streams”, IEEE, PP.2-10, 2010.

[6]. S.P.Syed Ibrahim1 K. R. Chandran2 M. S. Abinaya3 “Compact Weighted Associative Classification” IEEE pp.8-11, 2011.

[7]. 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.

[8]. Jun He , Bo Hu and Xiaoyong Du “A Multi-Relational Classification Algorithm based on Association Rule”, IEEE, pp. 4-09, 2009.

[9]. Kalyanmoy Deb, "Introduction to Genetic Algorithms", Kanpur Genetic Laboratory (Kangal), Depart of Mechanical Engineering, IIIT Kanpur 2005.
[10]. Rupali haldulakar, prof. Jitendra agrawal" Optimization of Association Rule Mining through Genetic Algorithm" (IJCSE) Vol. 3 No. 3 Mar 2011.
[11]. L.Dehaspe, and L.D. Raedt,"Mining Association Rules in Multiple Relations", Proceeding of the ILP, Springer-verlag, London U.K, 1997,pp.125-132
[12]. X.Yin, J.Han,J.Yang, and P.S. Yu,"CrossMine:Efficient Classification across Multiple Database Relations", Proceeding of the ICDE, IEEE,2004,PP.399-411.
[13]. B.Liu,Y.Ma and C.K Wong,"Improving an association rule based classifier", in proc.4th Eur. Conf.Principles Practice Knowledge Discovery Databases(PKDD-2000),2000
[14]. W. Li,J Han, and J.Pei, "CMAR:Accurate and Efficient Classification Based on Multiple Class-Association Rules,"Proc.IEEE Int'l Conf.Data Mining (ICMD 01),San Jose, California Nov.2001