International Journal of Scientific Research & Engineering Trends
Volume 3, Issue 2, March-2017, ISSN (Online): 2395-566X

Efficient and Scalable Multiple Class Classification using Bee Colony based Probabilistic Approach

Tarun Yadav
Email: yadavec@gmail.com

***Abstract* – Designed for multi-relational explore and learn about important device data classification, and can be widely used in many fields. New classification algorithm Union, naive Bayes, which is the main function of what is known in the literature for the application of multiple classification Union relational environment. The results showed that naive Bayes achieves greater accuracy compared to existing multi-relational algorithm. In addition, the rules of naive Bayes Over draft has a comprehensive database of more properties. There are many possible extensions Baye naive. Currently, naive patterns Baye and confidence to discover LCR repeated use and generation of classification rules. You can find the most important features of each category label using the procedures relating to the extension of the existing framework. Moreover, the current algorithm can improve in terms of improving the efficiency of techning. Relational multiple of the classification algorithm modified by optimization of bee colonies and Naïve Bayes classification rate and a better comparison, Baye Naive. In the process of Bee colony are the complexity increases calculation time complexity also increases. our overall proposal of test data was algorithm. In this dataset, the clearance rate was 92% .Also use another data set (data set abalone) and estimate some little difference in the clearance rate was 91%.**

***Keywords* – Data Mining, Classification, Multi Class Classification, Bee Colony, Naïve Bayes, Probabilistic Classification.**

I. INTRODUCTION

Exploring [18] operates at data from a large sets of data on the non-trivial novel extraction, and knowledge is a developing technology, which is a direct result of the increasing use of bases computer data to store and retrieve information efficiently .It also known as knowledge discovery in databases (KDD) and allows data extraction, data analysis and visualization of large data sets in a high level of abstraction, without a specific hypothesis in mind. The work of data mining is heard using a method called modeling with him to make predictions. Technical data mining are the result of a long process of research and product development, including neural networks, decision trees and genetic algorithms. This data recovery as needed using data mining technology. Data mining can be considered as a result of the natural evolution of information technology. This technology provides high availability of large amounts of data and the imminent convert this data into useful information and knowledge needs. Data mining is the extraction of patterns or knowledge of many interesting facts. It may be known by different names, such as knowledge discovery (mining) in databases (ECD), knowledge extraction, data analysis / design, archeology data leaks data, data collection, business intelligence and more. The term "data mining" [19] is simply the analysis of data in a database using tools taking into account the trends or anomalies without the knowledge of the meaning of data and is primarily used by statisticians, database data of the research and the business community. A data extraction software is not limited to modify the presentation, but before discovering the unknown data relationships. The information contained in the operation of the process of extracting data is contained in a historical database of past interactions. In principle, data mining are not specific to one type of media or data. Data mining should be applicable to any type of data warehouse.

II. RELATED WORK

Classification [13] is a data mining and machine learning important, which has been studied extensively and has a wide range of subject applications. The classification based on association rules, also called associative classification, is a technique that uses association rules to build the classifier. usually has two stages: the first is the set of class association rules (CAR), the right side is a class of labels, then select solid car to build a classifier rules. Thus, associative classification rules can generate more confidence and better readability compared to traditional approaches. Therefore, the associative classification has been widely studied in academia and industry, and more efficient algorithms [14, 15] proposed on. However, all above algorithms focus only on the processing of data organized in a single relational table. In practical application, the data is often stored in a dispersed manner on several tables in a relational database. Simply converting multi-relational data in a single flat table time can lead to high blood pressure and the cost of space, on the other hand, some semantic information essential for multi-relational data can be lost. Thus, existing associative classification algorithms can not be applied directly to relational data. We propose a new algorithm CMAR for associative classification can be applied to the multi-relational database environment. CMAR main idea is to extract the relevant characteristics of each class label in each table, respectively, and generate strong classification rules. By relevant characteristics, we refer to two types of sets of frequent closed items: a set of individual table objects in

© 2017 IJSRET

the destination table and sets of objects crosstabs non-target tables. The results of the experiment show that the two types of sets of the above objects contained enough relevant features of class labels. Then first width generate strict rules to classify these sets of elements with a pruning strategy used at this stage. After that, a classifier can be easily constructed to predict the class of objects invisible tags.

Multi-Relational Classification (MRC/RC)[16], which focuses on classification from relational databases comprising multiple tables, is one important task in multi relational data mining (MRDM/RDM)[17,18] 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 [17, 18] 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.

III. Proposed Methodology

Recent research has lower predictive accuracy leading to trends, combined [1] existing log-linear model with probabilistic techniques. While searching for information added features is computationally expensive, if successful, the new added features can increase the predictive accuracy. There are several possibilities for a combined hybrid approach. (I) Once good characteristics are added, they can be treated and other features used in a decision tree. (Ii) a simple decision Forest [2] is quick to learn and can establish a sound basis for assessing the information gain due to an added function candidate. (Iii) the regression weights can be used to carve information quickly join or small tables with weights, allowing the search features added to concentrate on the most relevant link roads. While in [9] hybrid mining algorithms to improve tree classification accuracy rate decision (DT) and Naïve Bayes classifier (NB) for the classification of multi-class problems, but they have no genetic algorithm, approaches rough and fuzzy set is used to address the multi-class classification tasks in real time in sets of dynamic characteristics.

IV. Approach Used

There are some limitations and problems sorting algorithm. Now we choose classification Association classification algorithm and used to optimize the bee colony algorithm to optimize the rate of classification Association. In our case, the results have improved due to the optimization of bee colonies is a heuristic function. The heuristic works best. Naive Bayes approaches applied to historical data and work better.

V. Methodology

Multi-relational classification is a data exploration and learning about important machine and can be widely used in many fields. New associative classification algorithm, Naive Bayes, which is the primary function of what is known in the literature to apply associative classification multi-relational environment. Experimental results show that Naive Bayes achieves greater accuracy compared to existing multi-relational algorithm. In addition, the rules of naive Bayes overdraft have a more complete characterization of databases. There are several possible extensions Naive Bayes. Currently, Naive Bayes uses a CSF-confidence to discover frequent patterns and generate classification rules. . You can discover the most relevant characteristics of each class label using measures related extend existing framework. So the current algorithm could be improved in terms of efficiency using the optimization technique proposed a method to construct a hybrid model through sorting and the algorithm of bee colonies partnership to increase the data rate classification amend the classification rules multi-relational association. classification rate more leads a better ranking. So our proposed algorithm provides a higher classification rate. Our proposed work is divided into two parts ,:

- For finding frequent item set and candidate key – we used Naïve bayes
- For Rule generation and optimization- we used Bee colony optimization along with Naïve bayes.

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

© 2017 IJSRET

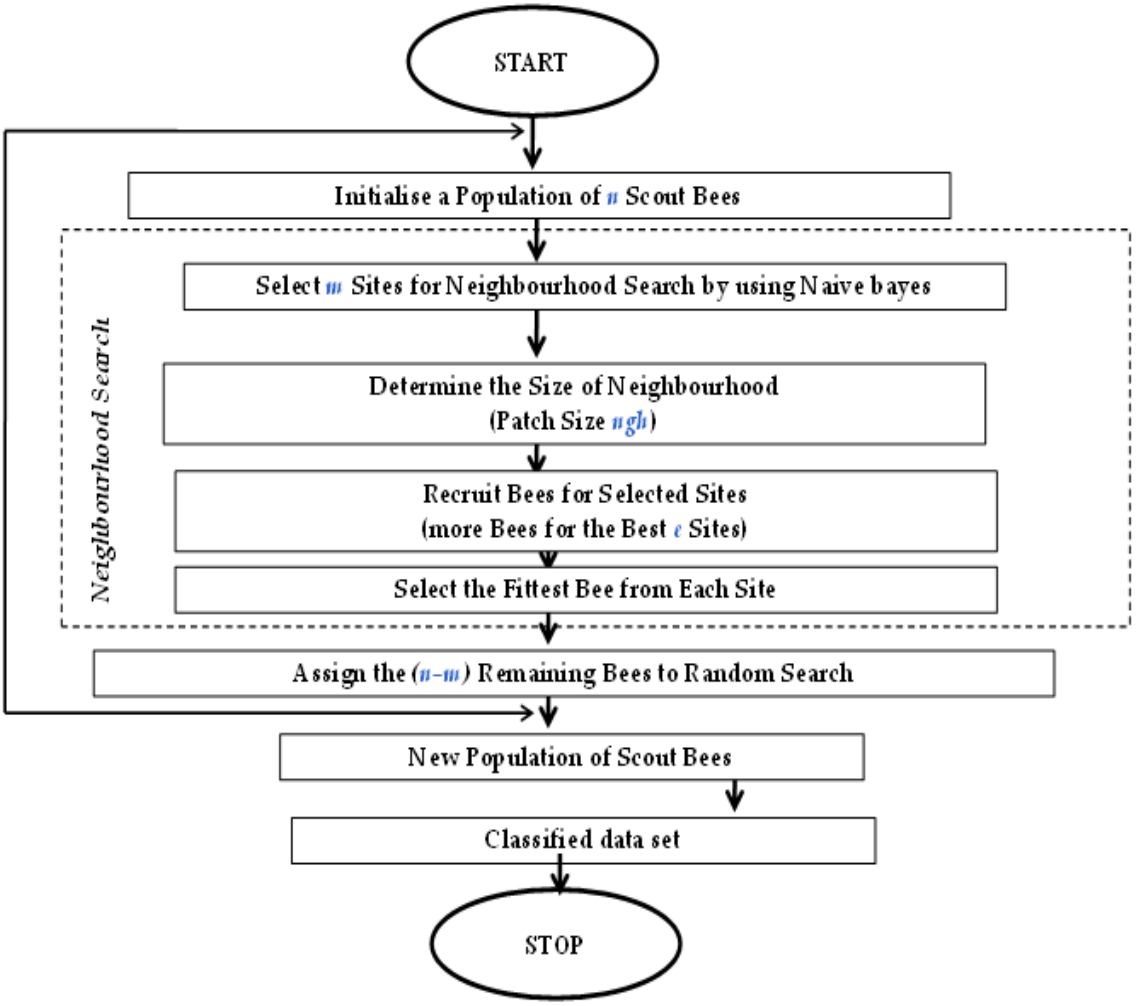

Fig.1. Hybrid Model for Multiple Relational Classification Algorithms Using Naïve Bayes Bee colony optimization

VI. Experimental Setup & Results Analysis

Experimental results show that bee colony-naive bayes tree gets higher accuracy comparing with the existing naive bayes tree. Rules discovered by bee colonynaive bayes tree. Have a more comprehensive characterization of databases. There is large possibility to extend Naive Bayes Tree rule set. Currently, Naive Bayes Tree uses a different initialization of data set in proposed framework to discover feature set 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 Bee colony so improved rate of classification in comparison of Naive Bayes Tree. 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 table given below shows the comparative analysis of efficiency and processing time. This table contains all the results regarding proposed and existing techniques.

© 2017 IJSRET

Table 1: Resultant table for Naive Bayes Tree and Bee-Naive Bayes Tree

			Efficiency		Time	
S. No	**Support**	**Confidence**	**Naive Bayes Tree**	**Bee Colony - Naive Bayes Tree**	**Naive Bayes Tree**	**Bee Colony - Naive Bayes Tree**
1	0.88	0.79	83.7024	90.9452	4.758	3.5412
2	0.834	0.814	83.2857	90.9365	3.6348	3.6816
3	0.931	0.916	83.7745	92.2051	4.0092	3.7392
4	0.496	0.485	83.8292	90.6719	5.616	3.1356
5	0.708	0.81	84.0481	91.6781	5.4756	5.2104
6	0.849	0.842	83.3032	91.7467	6.1824	5.6472
7	0.533	0.523	83.0443	91.8914	7.9016	6.1152
8	0.945	0.944	83.3839	91.032	3.9632	3.6816
9	0.836	0.823	83.6979	90.9408	3.7908	3.6192

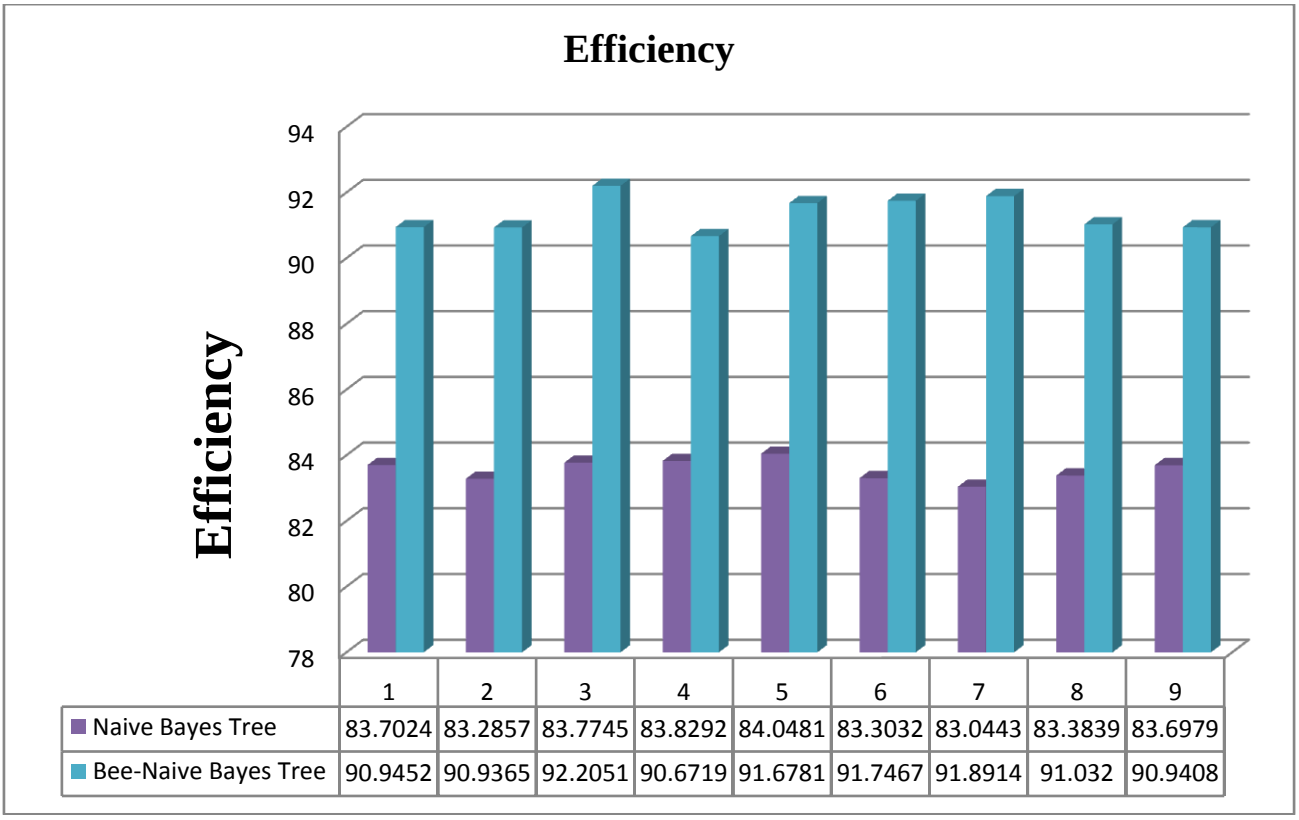

Fig.2. Comparative accuracy of Naive Bayes Tree and Bee colony Naive Bayes Tree

In Naive Bayes Tree, DCT algorithm is used which classified only one type of data means high order data not low and average by using this type of algorithm low order data is unclassified and high order data is classified .So this lead to negative rule generation and classification rate would not be above 90% as show in figure 2.

Whereas in proposed bee colony naive bayes tree low, average and high order data are classified easily because proposed methodology used bee colony which is computerized and optimization algorithm based on the mechanism of natural genetic and natural selection ,used genetic operation such as selection ,crossing and mutation and fitness function on the basis of data is optimized.

© 2017 IJSRET

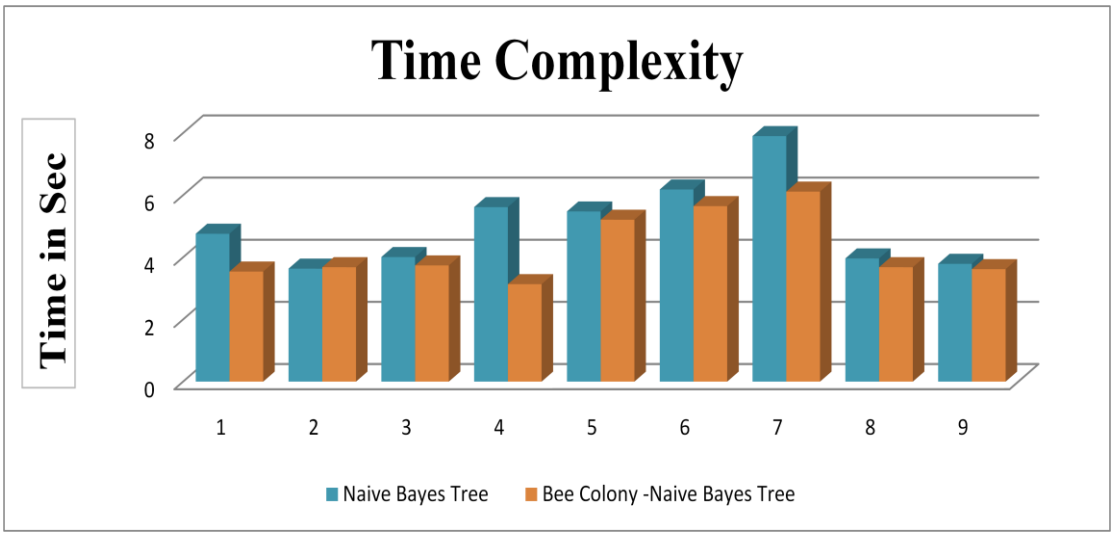

Fig.3. Comparative time complexity of Naive Bayes Tree and Bee colony Naive Bayes Tree

Classification rate/accuracy increased above 90% as show in figure 2. Total execution time of multiple relation classification algorithm on wine data is 3.27602 Sec and classification rate accuracy is 82.6137% whereas total execution time of multiple relational classification algorithm using Bee colony ie Naive Bayes Tree Using Bee colony on wine data set is 2.24641 Sec as shown in figure 6.8 and classification rate accuracy is 90.2829% as show in table 1, which is showing that classification rate accuracy increased above 90%.

VII. Conclusion

Relational multiple of the classification algorithm modified by optimization of bee colonies and Naïve Bayes classification rate and a better comparison, Baye Naive. In the process of Bee colony are the complexity increases calculation time complexity also increases. our overall proposal of test data was algorithm. In this dataset, the clearance rate was 92% .Also use another data set (data set abalone) and estimate some little difference in the clearance rate was 91%.

References

[1] B.N. Lakshmi. #1, G.H. Raghunandhan. #2 "A conceptual Overview of Data Mining"Proceedings of the National Conference on Innovations in Emerging Technology-2011 Kongu Engineering College, Perundurai, Erode, Tamilnadu, pp.27-32. India.17 & 18 February, 2011.

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

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

[4] Xun Zhu1, Hongtao Deng2, Zheng Chen3 "A Brief Review On Frequent Pattern Mining"PP-4-11 2011 IEEE.

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

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

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

[8] 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.

[9] You Wan#1, Chenghu Zhou*2" QuCOM: k nearest features neighborhood based qualitative spatial co-location patterns mining algorithm" IEEE pp.8-11, 2011.

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

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

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

[13] 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.

[14] W. Li, J. Han, and J. Pei, "CMAR: Accurate and efficientClassification Based on Multiple Class-Association Rules",Proceedings of the ICDM, IEEE Computer Society, SanJose California, 2001, pp. 369-376.

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

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

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

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

© 2017 IJSRET

[19] M.J. Zaki, and C.j. Hsiao, "CHARM: An EfficientAlgorithm for Closed Item set Mining", Proceedings ofSIAMOD International Conference on Data Mining, 2002,pp. 457-473.

[20] L. Xu, and K. Xie, "A Novel Algorithm for FrequentItem set Mining in Data Warehouses", Journal of ZhejiangUniversity, Journal of Zhejiang University, Zhejiang China, pp.216-224,2006.

[21] R. agrawal, t. imielinski and a. swami. "Mining association rules between sets of items in large databases". In proc. of the ACM sigmoid conference on management of data, Washington, D.C. May 1993.

[22] B. Liu, W. Hsu, and Y. Ma. "Integrating classification and association rule mining". In KDD 98, New York, NY, Aug.1998.

[23] B. Liu, Y. Ma, and C.-K. Wong, "Improving an association rule basedclassifier," in Proc.4th Eur. Conf. Principles Practice KnowledgeDiscovery Databases (PKDD-2000), 2000.

[24] Rajdev Tiwari, Manu Pratap Singh "Correlation-based Attribute Selection using Genetic Algorithm" International Journal of Computer Applications (0975 – 8887) Volume 4– No.8, August 2010.

[25] Kalyanmoy Deb, "Introduction to Genetic Algorithms", Kanpur Genetic Laboratory (Kangal), Depart of Mechanical Engineering, IIIT Kanpur 2005.

© 2017 IJSRET