International Journal of Scientific Research & Engineering Trends
Volume 2, Issue 1, Jan.-2016, ISSN (Online): 2395-566X

Efficient and Scalable Multiple Class Classification using Bee Colony Optimization

Esha Shrivastava*, Yogdhar Pandey
Department of Computer Science & Engineering
Sagar Institute of Research & Technology
*Email: shrivastavaesha@gmail.com

***Abstract* – Multiple relational classification algorithm modified by Bee Colony optimization and Navies bayes so improved rate of classification in comparison of Navies bayes. In the process of NB Bee Colony optimization 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%.**

***Keywords* – Data Mining, Classification, Clustering, Association Rule, Bee Colony Approach.**

I. INTRODUCTION

Data mining [1] a nontrivial extraction of the novel, implicitly and actionable knowledge from large data sets is an evolving technology, which is a direct result of the increasing use of computer databases for the purpose of storing and retrieve effective way of information may also known as knowledge discovery in databases (EDC) and enables data mining, data analysis and visualization of data from large databases to a level 'high abstraction without specific hypotheses in mind. The operation of the extraction is to use 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 that the technology helps mining. Data mining can be considered the 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 (KDD), knowledge extraction, data analysis / design, archeology, data dredging data collection information, the business intelligence and other. The term "data mining" [2] is nothing but analysis of data in a database using tools which look for trends or anomalies without the knowledge of meaning of the data and is primarily used by statisticians, database researchers and business communities. A data mining software does not just change the presentation, but discovers previously unknown relationships among the data. The information on which the data mining process operates is contained in a historical database of previous interactions. In principle, data mining is not specific to one type of media or data.

Exploring [3,16,17] 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" [4-9] 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. Some types of information collected are:

II. RELATED WORK

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 [2,13,14,15] 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 [5-12] 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. In this paper, a probabilistic approach of bee colonies for optimizing association rules proposed.

III. PROPOSED METHODOLOGY

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. 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 well.

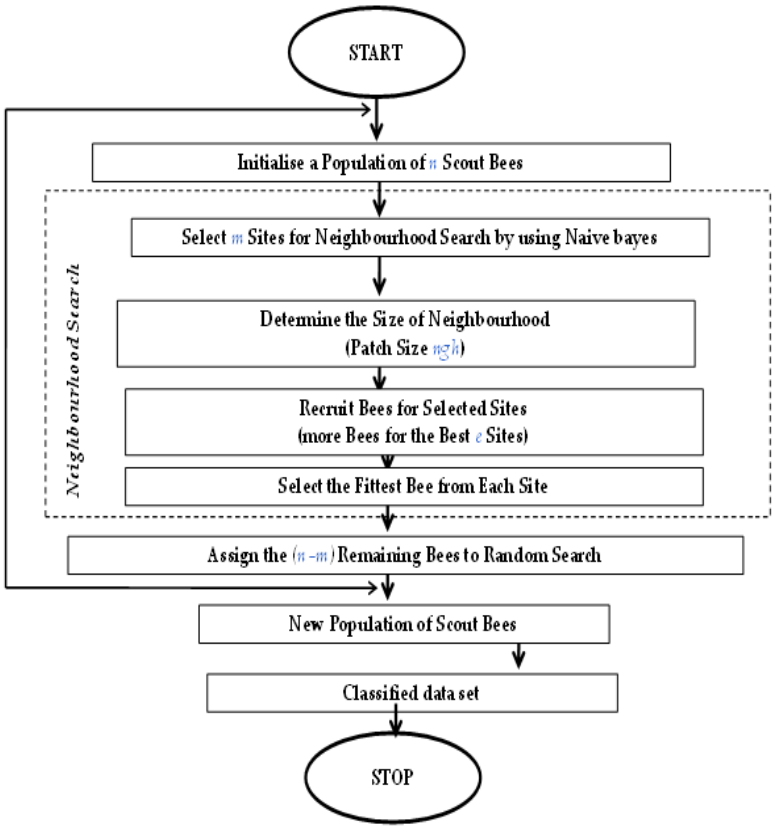

Fig.1. Hybrid Model for Multiple Relational Classification Algorithms Using Naïve Bayes Bee colony optimization

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.

IV. Simulation Results

The experimental results of all Naive Bayes .wine output data contain three types of data type 1 C, C and C type type 2 3. The data fall into several low, medium and high categories. We use Bayesian that is effective to find all frequent item sets and implements the search using the property level as frequent element algorithm. Bayesian class generate a single data type algorithm means that the data for no more high and middle .for the use of these data is less algorithm lower order unclassified data and high-order this cable is classified .so rule negative and the clearance rate would not be greater than 90%.

Performance of multiple relational classification algorithm shows that when we used Navies Bayes on Abalone Dataset then value of runtime is 7.4375 Sec and classification rate accuracy is 84.570%.

Table 1: Naive Bayes Tree

Experimental Data base	Approach	Data Initialization	Cross Validation	Processing Time	Accuracy
Abalone Dataset	**Naive Bayes Tree**	**0.9**	**7**	**6.5 Sec**	**84.48%**

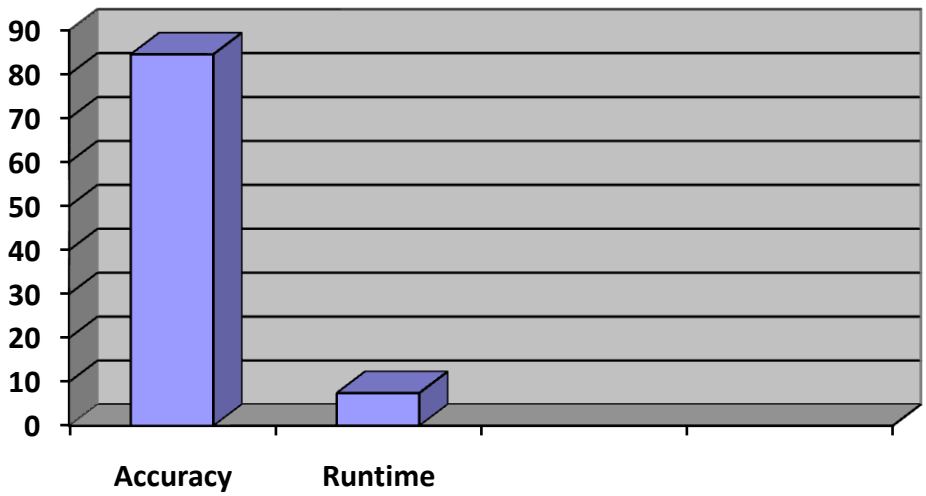

Fig.2. Simulation results with Naive Bayes Tree

The experimental results of the proposed NB bee colonies. set of wine data contains three types of data type 1 C, C and C type2 type3. NB association rule algorithm to generate rules and algorithm of bee colonies for optimization was used and build a hybrid model for classifying data by combining both the receiving part of the algorithm. NB data bee colonies with low, medium and high to easily classified because we use a bee colony, which is computerized and optimization algorithm based on the natural genetic mechanism of natural selection, they use operation such as genetic selection, crossover and mutation data base and the fitness function is optimized and applied in conjunction with the naive Bayes approach lead to association rules to optimize the use of probabilistic classification. We have implemented this logic in a MATLAB 7.8.0. And will get a good result. classification / accuracy rate increased above 90%

Performance of multiple relational classification algorithms using Bee Colony Optimization shows that when we used NB Bee Colony optimization approach on Abalone Dataset then value of runtime is 5.7812 Sec and classification rate accuracy is 91.789%.

Table 2: NB Bee Colony Optimization Approach

Experimental Data base	Approach	Data Initialization	Cross Validation	Processing Time	Accuracy
Abalone Dataset	**NB Bee Colony Optimization Approach**	**0.9**	**7**	**2.53 Sec**	**90.52%**

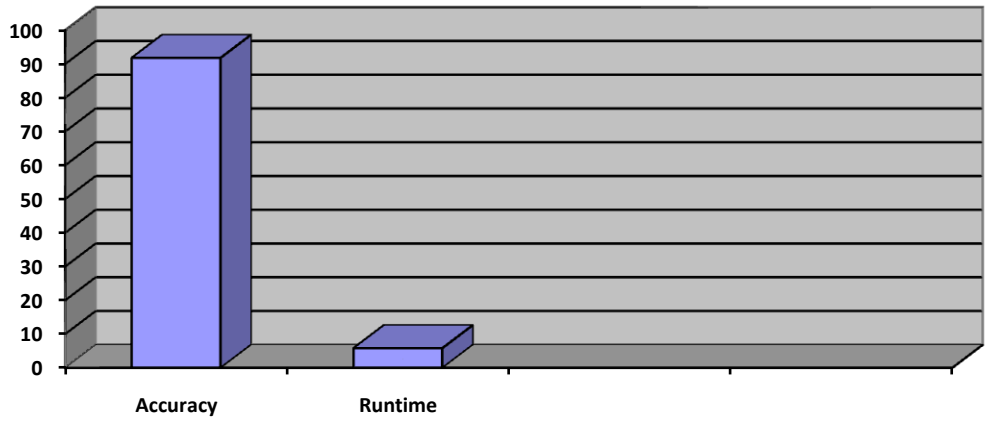

Fig.3. Simulation results with NB Bee Colony Optimization Approach

Performance of multiple relational classification algorithm shows that when we used Naive Bayes Tree on Abalone Dataset then value of runtime is 7.4375 Sec and classification rate accuracy is 84.570%.Performance of multiple relational classification algorithm using Genetic algorithm shows that when we used Naive Bayes Using Bee Colony Optimization Approach on Abalone Dataset 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 3: Naive Bayes Tree and NB Bee Colony Optimization Approach

Experimental Data base	Approach	Data Initialization	Cross Validation	Processing Time	Accuracy
Abalone Dataset	NB Bee Colony Optimization Approach	0.9	7	2.53 Sec	90.52%
Abalone Dataset	Naive Bayes Tree	0.9	7	6.5 Sec	84.48%

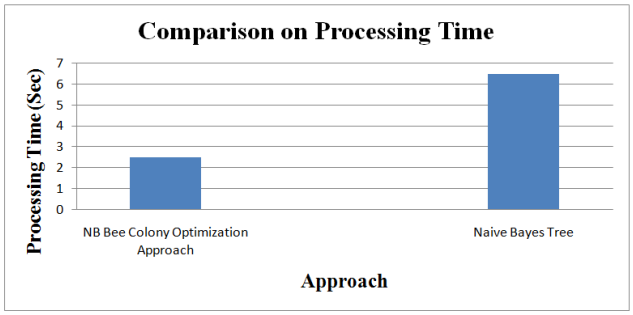

Fig.4. Processing Time results with Naive Bayes Tree and NB Bee Colony Optimization Approach

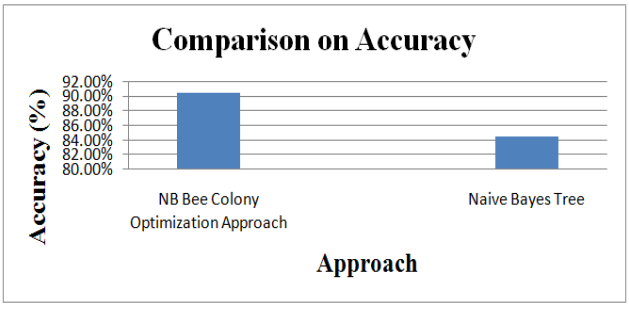

Fig.5. Accuracy results with Naive Bayes Tree and NB Bee Colony Optimization Approach

V. CONCLUSION

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 Naïve 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 Baye. Currently, Baye Naïve uses a CSF-confidence to discover frequent patterns and generate classification rules.In the process of NB 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%. 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.

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

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

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

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

[7] 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.

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

[9] 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.

[10] 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.

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

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

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

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

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

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

[17] 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.