

OPTIMIZATION OF MULTIPLE RELATIONAL CLASSIFICATION ALGORITHMS USING GENETIC ALGORITHM

Pradeep Singh Raghav¹, Nikhil Singh², Satendra Singh Thakur³ and Priyank Jain⁴

^{1,2,3}Dept. of Computer Science, Patel Collage of Science and Technology, Bhopal-462023, India

Information Technology Dept, TIT Science Bhopal, Bhopal-462001, India, E-mail: priyankjain1984@gmail.com

ABSTRACT

Data classification is a challenging task in the field of machine learning. Now a days in every field such as online data processing, pattern recognition, pattern classification, medical diagnosis are required classification of data. In Current scenario various classification methods are available such as Decision tree, Naive bayes, KNN. Above classification rate is 80%(MAX) hence the 20% data are unclassified. This is a challenge in the field of data classification. In this dissertation, we used multiple relational classification algorithm based on genetic algorithm used for optimization of classification rate, generated by association rule. Association rules are generated on the basis of Support and Confidence. This value generate a classification rate.

1. INTRODUCTION

Classification rule mining aims to discover a small set of rules in the database that forms an accurate classifier. Association rule mining finds all the rules existing in the database that satisfy some minimum support and minimum confidence constraints. The integration is done by focusing on mining a special subset of association rules, called class association rules (CARs). In addition, this integration helps to solve a number of problems that exist in the current classification systems Optimization. The Search for the best solution among alternatives or the extreme value of a variable or a function. *Genetic Algorithm* GAs is a methods, which have been inspired by the process of biological evolution, Because of GAs' is a robust and it has uniform approach to solve the problems. It is used for optimization of solution, Creates an initial feasible solution and iteratively creates new "better" solutions, Based on human evolution and survival of the fittest., Must represent a solution as an individual. There are two terms Popoulation and individual, Individual it is like to child. Population there is combination of parents group like father, grand father, mother, grand mother etc A Genetic Algorithm (GA) is used to find the similarity between population and individual and computational model consisting of five parts: A starting set of individuals, P , Crossover is a process of taking more than one parent solutions and producing a child solution from them, Mutation the solution may change entirely from the previous solution, Fitness determine the best individuals, Algorithm which applies the crossover and mutation techniques to P iteratively using the fitness function to determine the best individuals in P to keep. GA Advantage is Genetic algorithm are easily

parallelizable and have been used for classification as well as other optimization problems. GA disadvantages are Difficult to understand and explain to end users, Determining fitness function is difficult, Determining how to perform crossover and mutation is difficult. Gas are stochastic search methods, which have been inspired by the process of biological evolution. because of Gas robustness and their uniform approach to large number of different classes of problems . they have been used in many applications. data mining is also one of the important application fields of GAs. In data mining, a GA can be used either to optimize parameres for other kinds of data miming algorithms or to discover knowledge by itself. In this latter task the rules that a GAs finds are usually more general because of its global search nature. In contrast, most of the other data mining methods are based on the rule induction paradigm. where the algorithm usually performs a kind of local search. The advantage of GAs become more obvious when the search space of a task is large.

2. LITERATURE SURVEY

Zhen- Hui Song & Yi Li, "Associative classification over Data Streams", IEEE, PP. 2-10,2010 [1]: Describe in the field of data classification as AC-DS, a new associative classification algorithm for data streams which is based on the estimation mechanism of the Lossy Counting and landmark window model. The rate of classification is only 78%. This algorithm will not output an accurate result". Jun He^{1,2}, Bo Hu^{1,2} and Xiaoyong Du^{1,2} "A Multi-Relational Classification Algorithm based on Association Rule", IEEE, PP. 4-09,2009[2]: Describe in the field of data classification as MCAR mines relevant features in each table to predict the class label. close itemsets technique and tuple ID

Propagation method are used to improve the performance of the algorithm. MCAR has higher accuracy and better understandability 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.

3. PROPOSED APPROACH

Combination of classification association(ca) and genetic algorithm (GA): In this dissertation we used association classification and genetic algorithm build a hybrid model for data classification. In this process, we used apriori algorithm and genetic algorithm for the formation of hybrid algorithm. An association rule is an expression of the form $X \rightarrow Y$. Where X, Y are the sets of items. Association rule mining is a method for finding the frequent set of items from the data set. Frequent set is the set whose value is maximum use. Support and Confidence: The support for an association rule $X \rightarrow Y$ is the percentage of transactions in the database that contain $X \cup Y$. The confidence or strength for an association rule $X \rightarrow Y$ is the ratio of the number of transactions that contain $X \cup Y$ to the number of transactions that contain X . $\text{conf}(X \rightarrow Y) = \frac{\text{supp}(X \cup Y)}{\text{supp}(X)}$. The condition for generation of rule: item \geq min support

BEGIN

Generate initial population;

Compute fitness of each individual;

REPEAT /* New generation */

FOR population size / 2 DO

Select two parents from old generation;

/* biased to the fitter ones */

Recombine parents for two offspring;

Compute fitness of offspring;

Insert offspring in new generation

END FOR

UNTIL population has converged

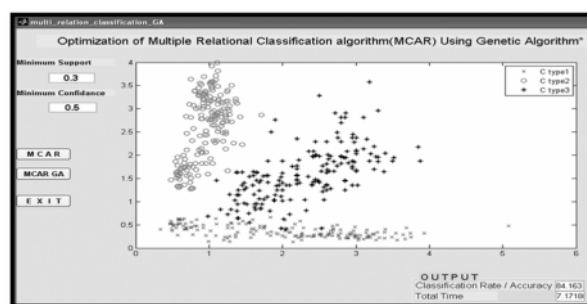
4. IMPLEMENTATION DETAILS

Our hybrid model implements in the Matlab 7.8.0 and used data set for the validation of classification.

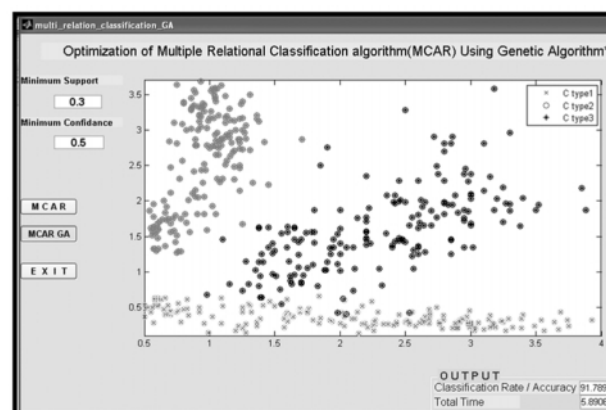
a. MATLAB

(MATrix LABoratory) A programming language for technical computing from The MathWorks, Used for a wide variety of scientific and engineering calculations, especially for automatic control and signal processing, MATLAB runs on Windows, Mac and a variety of Unix-based systems.

b. Result of MCAR



c. Result of MCAR using GA



d. Performances on Both Databases:

Approach.	Accuracy	Run Time
MCAR	84.163%	7.1718 Sec
MCARS USING GA	91.789%	5.8906 Sec

5. CONCLUSION

We find a problem and solved that problem with the help of genetic algorithm also implementing this logic in a matlab and come some good result. Classification rate/accuracy increased above 90%.

6. FUTURE ENHANCEMENT

To make genetic algorithm more effective and efficient it can be incorporated with other technique so it can provide a best result.

REFERENCES

- [1] Zhen-Hui Song and Yi Li, "Associative Classification Over Data Streams", 978-1-4244-7941-2/10/©2010 IEEE.
- [2] Jun He^{1,2}, Bo Hu^{1,2} and Xiaoyong Du^{1,2} "A Multi-Relational Classification Algorithm based on Association Rule", *IEEE*, pp. 4-09, 2009[2].
- [3] Pei-yi hao and yu-de chen, "A Novel Associative Classification Algorithm: A Combination of LAC and CMAR with New Measure of Weighted Effect of Each

- Rule Group", *Proceedings of the 2011 International Conference on Machine Learning and Cybernetics*, Guilin, 10-13 July, 2011.
- [4] You Wan#1, Chenghu Zhou*2, "QuCOM: k Nearest Features Neighborhood Based Qualitative Spatial Co-location Patterns Mining Algorithm", 978-1-4244-8351-8/11/©2011 IEEE.
- [5] Achilleas Tziatzios and Jianhua Shao and Grigorios Loukides, "A Heuristic Method for Deriving Range-Based Classification Rules", *2011 Eighth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*.