

A Review of Rule Based Classification Technique using Soft Computing Technique

Neha Saxena¹, Pankaj Kawadkar²

¹M.Tech Scholar, PCST, Bhopal, INDIA

²Assistant Professor, PIES, Bhopal, INDIA

ABSTRACT

Classification technique is important tool in data mining for the classification and grouping of data. The techniques of data mining offer various types of classification technique such as binary classifier, statistical classifier; neural network based classifier and also gives rule based classifier. The applicability of classifier depends on the processing of data domain. Now a day's rule based classifier are used in various domain such as medical diagnoses data, computer vision and pattern recognition. For the improvement of the performance of rule based classifier used soft computing technique. Soft computing technique provide the process of optimization for the selection of rule for the building of for the improvement of rule based classifier.

Keywords-- Data Mining, Rule based Classification, Soft Computing and Rough Set.

I. INTRODUCTION

Classification is most important tool of data mining. The process of classification define in various ways such as binary classification, multi-class classification and neural network based classification. The process of classification suffered from the process of data diversity. Now in several field of engineering and medical science acquired the process of classification. The major challenge for the classification is how to improve the classification ratio. In the process of classification improvement various authors used different technique such as genetic algorithm ant colony optimization and graph set theory used for the improvement of classification task. Now a day's various authors used rough set based classification technique for the improvement of classification task. In the process of classification rule based classification technique play an important for the analysis of medical data. For the rule based classification technique used CBA classification algorithm. For classification purpose, basically CBA classification algorithm based on Apriori rule mining algorithm. Associative classification play big role in medical data diagnose. The diagnose of pattern faced the series of training process. The training process of classification technique generates the accuracy performance of classifier and method of pattern diagnose. In phase of dataset training imbalance of data arise a problem of minority and majority of class labeling. Classification for

other approach to imbalance data, the replace rules approach is focused on handling cardinality aspects of imbalance. Strengthening some sub-regions and leaving uncovered examples. Some difficult examples may be uncovered depending on the procedure for tuning parameters which is time consuming and sophisticated. Some authors are used graph based optimization technique for improvement of classification ratio. Section II discusses associative classification and soft computing, Section III discusses related work. Section IV discusses Problem formulation. And Section V discusses comparative result analysis. Finally, concluded in section VI.

II. ASSOCIATIVE CLASSIFICATION AND SOFT COMPUTING

Let D is the dataset. Let I be the set of all items in D and C be the set of class labels. We say that a data case $di \in D$ contains $X \subseteq I$, a subset of items, if $X \subseteq di$. A class association rule (CAR) is an implication of the form $X \rightarrow c$, where $X \subseteq I$, and $c \in C$. Bing Liu et al. [10] first proposed the AC approach, name classification based on association algorithm (CBA), for building a classifier based on the set of discovered class association rules. The difference between rule discoveries in AC and conventional frequent item set mining is that the former task may carry out multiple frequent item set mining processed for mining rules of different classes simultaneously. Data mining in associative classification (AC) framework usually consists of two steps [5]:

1. Generating all the class association rules (CARs) which has the form of $I \text{ set} \Rightarrow c$, where $I \text{ set}$ is an item set and c is a class.
2. Building a classifier based on the generated CARs. Generally, a subset of the association rules was selected to form a classifier and AC approaches are based on the confidence measure to select rules [23].

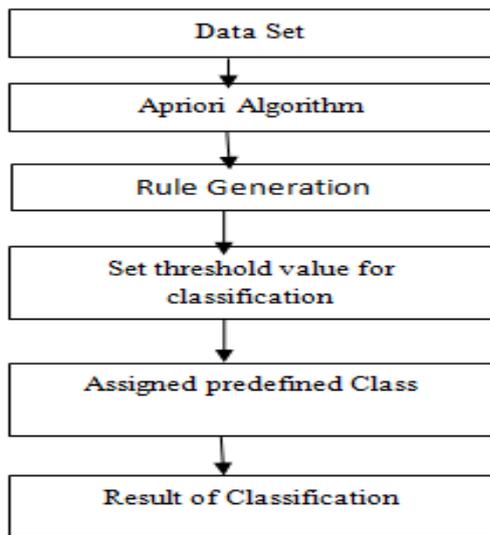


Figure 1: Association Classification

Soft computing

Soft computing is a consortium of methodologies, (like fuzzy logic, neural networks, genetic algorithms, rough sets), that works synergistically and provides, in one form or another, flexible information processing capabilities for handling real life problems. Its aim is to exploit the tolerance for imprecision, uncertainty, approximate reasoning and partial truth in order to achieve tractability, robustness, low solution cost, and close resemblance with human like decision-making. The process of knowledge discovery from data bases (KDD), on the other hand, is a real life problem solving paradigm and is defined as the non-trivial process of identifying valid, novel, potentially useful and understandable patterns from large data bases, where the data is frequently ambiguous, incomplete, noisy, and redundant and changes with time. Data mining is one of the fundamental steps in the KDD process and is concerned with the algorithmic means by which patterns or structures are enumerated from the data under acceptable computational efficiency. Soft computing tools, individually or in integrated manner, are turning out to be strong candidates for performing data mining tasks efficiently. At present, the results on these investigations, integrating soft computing and data mining, both theory and applications, are being available in different journals and conference proceedings mainly in the fields of computer science, information technology, engineering and mathematics.

III. RELATED WORK

In this section discuss the related work to rule based classification technique used in graph based theory and some other optimization technique. The process of classification deals the method of CBA and other rule based classification technique.

[1] In this paper author proposes a new classification rule inducing algorithm. Standard Rough Set has the attribute reduction step before rule generation and does not handle the inconsistent examples. In comparison with standard rough sets theory it calculates value core without attribute reduction in advance and does not remove examples

covered by the newly generated rule. In this paper author also introduce a new rough set based rule generation method to construct a higher accuracy classifier compared to C4.5 and RIPPERK. The new algorithm generates the core value for every consistent example first and generates a rule on basis of it directly.

[2] In this paper, gravitational search algorithm (GSA) which is one of the newest swarm based heuristic search technique, is employed to generate prototypes for nearest-neighbor (NN) classification. The proposed method is compared with several state-of-the-art techniques and results are presented. The comparison shows that our proposed method can achieve higher classification accuracy than the competing methods and has a good performance in the field of prototype generation. Gravitational search algorithm is one of the latest swarm intelligence-based optimization techniques, which has been inspired by Newtonian laws of gravity and motion.

[3] In this paper author compare four different algorithms for computing reduct or a reduct approximations based on Rough Set Theory. Reduct are used to create rule sets and then to classification of the test set. The data of patients suffering from heart disease is used to make algorithm's accuracy measure and create Receiver Operating Characteristic (ROC) curves. It has been concluded that out of these four methods of reduct calculation, classification through table rule set using covering algorithm yields best accuracy. Moreover statistical methods for algorithms comparison are presented. The ROSSETA software is used for reduce generation, rule set calculation and rough set classification.

[4] In this paper author propose a new associative classification method, CMAR, i.e., Classification based on Multiple Association Rules. The method extends an efficient frequent pattern mining method, FP-growth, constructs a class distribution-associated FP-tree, and mines large database efficiently. Moreover, it applies a CR-tree structure to store and retrieve mined association rules efficiently, and prunes rules effectively based on confidence, correlation and database coverage. The classification is performed based on a weighted analysis using multiple strong association rules.

[5] In this paper, the basic concepts of rough set theory and other aspects of data mining are introduced. The rough set theory offers a viable approach for extraction of decision rules from data sets. The extracted rules can be used for making predictions in the semiconductor industry and other applications. This contrasts other approaches such as regression analysis and neural networks where a single model is built. One of the goals of data mining is to extract meaningful knowledge. The power, generality, accuracy, and longevity of decision rules can be increased by the application of concepts from systems engineering and evolutionary computation introduced in this paper. A new rule-structuring algorithm is proposed.

[6] In this paper, a new associative classification technique, Ranked Multi label Rule (RMR) algorithm is introduced, which generates rules with multiple labels. Rules derived by current associative classification algorithms overlap in their training objects, resulting in many redundant and useless rules. However, the proposed algorithm resolves the overlapping between rules in the classifier by generating rules that does not share training objects during the training

phase, resulting in a more accurate classifier. Results obtained from experimenting on 20 binary, multi-class and multi-label data sets show that the proposed technique is able to produce classifiers that contain rules associated with multiple classes.

[7] In this paper, we use Rough Set Theory (RST) to address the important problem of generating decision rules for data mining. In particular, we propose a rough set-based approach to mine rules from inconsistent data. It computes the lower and upper approximations for each concept, and then builds concise classification rules for each concept satisfying required classification accuracy. Estimating lower and upper approximations substantially reduces the computational complexity of the algorithm. We use UCI ML Repository data sets to test and validate the approach.

[9] In this paper we consider the inclusion properties for upper and lower approximation of union and intersection of sets for both pessimistic and optimistic multi-granulations. We find that two inclusions for pessimistic cases are actually equalities. For other six cases we provide examples to show that actually the proper inclusions hold true. Also, we consider types of elements in classifications with respect to both types of multi-granulations and establish a general theorem on them.

[10] In this paper author extended the study of rough set on two universal sets further by defining approximation of classifications in rough set on two universal sets. We considered the types of union and intersection of rough sets on two universal sets. Also, we generalized the four theorems established by Busses' on approximation of classifications and obtain two theorems of necessary and sufficient type to the settings of rough sets on two universal sets. From these theorems several other results besides Busses' theorems could be derived as corollaries in the settings of rough set on two universal sets.

[11] In this work, author extend this framework to numerical feature spaces by replacing partition of universe with neighborhood covering and derive a neighborhood covering reduction based approach to extracting rules from numerical data. We first analyze the definition of covering reduction and point out its advantages and disadvantages. Then we introduce the definition of relative covering reduction and develop an algorithm to compute it. Given a feature space, we compute the neighborhood of each sample and form a neighborhood covering of the universe, and then employ the algorithm of relative covering reduction to the neighborhood covering, thus derive a minimal covering rule set.

[12] This paper presents a new weighted rough set framework for early intervention and prevention of neurological dysfunction and kernicterus that are catastrophic sequels of neonatal jaundice. The obtained results illustrate that the weighted rough set can provide significantly more accurate and reliable predictive accuracy than well known algorithms such as weighted SVM and decision tree considering the fact that physicians do not have any estimation about probability of jaundice appearance.

IV. PROBLEM FORMULATION

In the process of review we found that some performance affected problem related to the heart disease

data classification. These problem are affected the performance and accuracy of associative classification technique. The unclassified data increase, decrease the accuracy and performance of classifier. Some problems are mentioned here [4, 6, 9, 10].

Some number of rough rule are generated.

- Feature selection of data
- Proper selection of rule set.
- New class generation.
- imbalanced data problem
- algorithm selection

V. EXPERIMENTAL ANALYSIS

In this section discuss the comparative result analysis of pervious algorithm used for the classification process used soft computing approach and optimization technique. The all pervious algorithm implement in MATLAB 7.8.0 software and used some standard dataset form UCI machine learning. Form analysis of data used five dataset such as cancer, e coli, Liver Thyroid Cleve Land.

| Name of Data set | RGI | CBA | Decision Table | J Rip |
|------------------|-------|-------|----------------|-------|
| Breast Cancer | 71.69 | 64.27 | 72.07 | 73.79 |
| Liver | 65.82 | 71.03 | 57.98 | 59.39 |
| E coli | 83.32 | 80.94 | 64.88 | 82.15 |
| Thyroid | 95.78 | 93.46 | 79.98 | 93.92 |
| Cleve Land | 82.14 | 74.84 | 71.55 | 74.52 |

Table 1: Performance analysis of different methods with different data set.

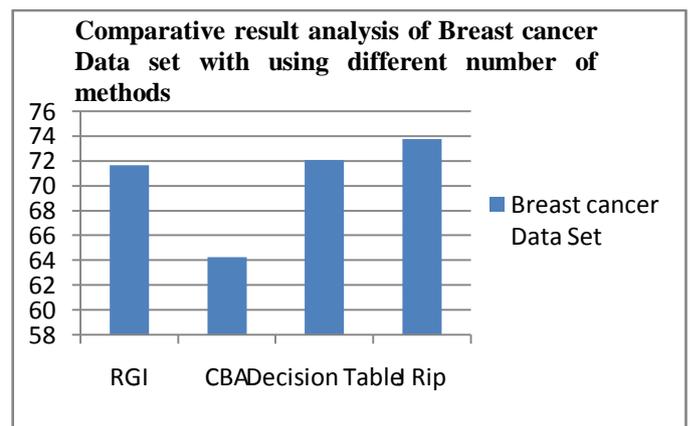


Figure 2: Shows that the Comparative result analysis of Breast cancer data set for different methods.

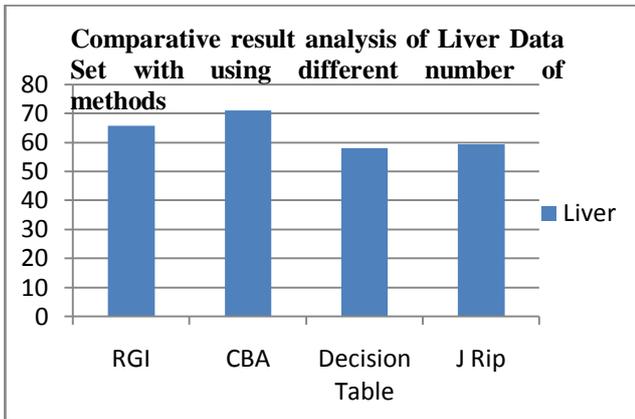


Figure 3: Shows that the Comparative result analysis of Liver data set for different methods.

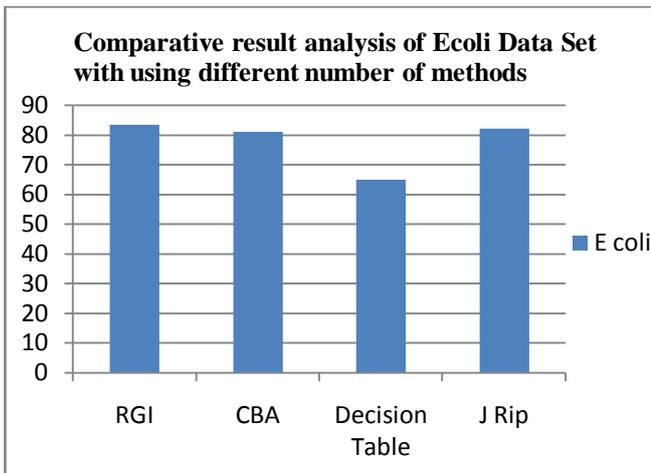


Figure 4: Shows that the Comparative result analysis of E coli data set for different methods.

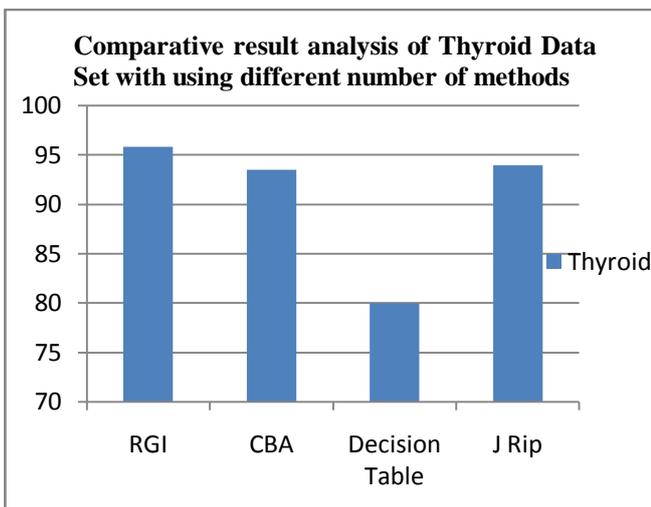


Figure 5: Shows that the Comparative result analysis of Thyroid data set for different methods.

VI. CONCLUSION AND FUTURE WORK

In this paper study of rule based classification technique using soft computing approach and optimization

technique for the improvement of rule based classification technique. Process of experimental task discuss four rule based classification algorithm such as RGI, decision tree, CBA and JRIP. In all these algorithm RGI have better performance in compression of all. In future improve the classification ratio using gravitational search algorithm.

REFERENCES

- [1] Honghai Feng, Yanyan Chen, Qing Ni, Junhui Huang "A new rough set based classification rule generation algorithm" International Conference on Computational Science and Computational Intelligence, IEEE 2014. Pp 380-385.
- [2] Mohadeseh Rezaei, Hossein Nezamabadi-pour "A prototype optimization method for nearest neighbor classification by gravitational search algorithm" IEEE, 2014. Pp 1-4.
- [3] Kanik Tomasz "Comparison of Algorithms for Computing Data Set Reducts" International Workshop on Biomedical Technologies, IEEE, 2013. Pp 99-105.
- [4] Wenmin Li, Jiawei Han, Jian Pei "CMAR: Accurate and Efficient Classification Based on Multiple Class-Association Rules" Natural Sciences and Engineering Research Council of Canada, 2010. Pp 346-353.
- [5] Andrew Kusiak "Rough Set Theory: A Data Mining Tool for Semiconductor Manufacturing" IEEE TRANSACTIONS ON ELECTRONICS PACKAGING MANUFACTURING, VOL- 24, 2001. Pp 44-50.
- [6] F.A. Thabtah , P.I. Cowling "A greedy classification algorithm based on association rule" Applied Soft Computing, Elsevier ltd. 2007. Pp 1102-1111.
- [7] Prasanta Gogoi , Dhruba K. Bhattacharyya "A rough set-based effective rule generation method for classification with an application in intrusion detection" Int. J. Security and Networks, Vol-8, 2013. Pp 61-71.
- [8] Decui Liang, Dun Liua, Witold Pedrycz, PeiHu "Triangular fuzzy decision-theoretic rough sets" International Journal of Approximate Reasoning, 2013. Pp 1087-1106.
- [9] R. Raghavan, B.K.Tripathy "On Some Comparison Properties of Rough Sets Based on Multi granulations and Types of Multigranular Approximations of Classifications" I.J. Intelligent Systems and Applications, 2013, Pp 70-77.
- [10] B. K. Tripathy , D. P. Acharjya "Approximation of Classification and Measures of Uncertainty in Rough Set on Two Universal Sets" International Journal of Advanced Science and Technology Vol-40, 2012. Pp 77-90.
- [11] Yong Du, Qinghua Hu , Pengfei Zhu, Peijun Ma "Rule learning for classification based on neighborhood covering reduction" Information Sciences, 2011. Elsevier ltd. Pp 1-11.
- [12] Hala S. Owna, Ajith Abraham "A new weighted rough set framework based classification for Egyptian NeoNatal Jaundice" Applied Soft Computing, Elsevier ltd. 2012. Pp 999-1005.