International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-8 Issue-5, June 2019

Enhancing Classification Accuracy by Attribute Reduction Technique

A. Rama, C. Nalini

Abstract*: Data mining methods helps in analyzing the data set efficiently by reducing the size of the search space, so as to choose significant attribute for recognition of the type of appropriate data. This study deals with the classification of categories of glass, which helps in criminology investigation. The glass material got as evidence in the crime scene is to be correctly identified. The fundamental purpose of this work is to deal with a large glass data set with high accuracy of identifying king of the glass. The models are constructed with supervised learning algorithm in weka tool. It is important to minimize the dimensions of data by constructing the models by selecting the attributes that is implemented as search methods, which are applied to predict the evaluating for the possible test cases.

Index Terms: Data mining, Bayes algorithm, Glass data set, Search Methods, Tree, Classification, Meta boost.

I. INTRODUCTION

Obtaining accurate model by aggregating the classification models is the trend established in bioinformatics field. Work of this nature is found more, similar to that of author's [4,5]. Designing the experiments for discovering the patterns covers variety of configurations for iterative steps in the literature2. In this paper we improve the accuracy of the model for a glass dataset by applying the iterative steps as shown in Fig 1. The idea of iterations using various types of learning models including meta classifiers is novel in this case. We describe the dataset attributes in section 2 and in section 3 the descriptions of learning methods. Finally the experimental results and conclusion are given in section 4.

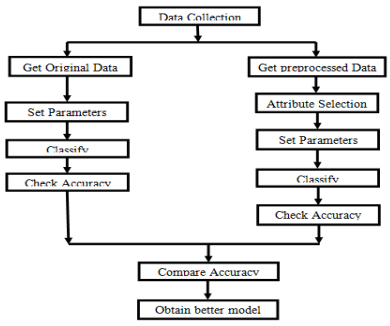

Fig.1 Iterative steps in Data Mining

II. EXPERIMENTAL ANALYSIS

Various classification algorithms are implemented and evaluated for better accuracy.

Manuscript published on 30 June 2019.
***** Correspondence Author (s)
A.Rama, Assistant Professor, Department of Information Technology, Bharath Institute of Higher Education and Research
Dr. C. Nalini, Professor, Department of Computer Science and Engineering, Bharath Institute of Higher Education and Research

A. Dataset

The glass data set [8] is extracted from the UCI repository which was used for this experiment,

a. Dataset Description

The dataset describes eight attributes which are listed below, The resultant values are categorized as either negative or positive. The resultant class represents the potential class of the glass a material. Number of Instances in this database is 214.

b. List of Description of Attributes.

For each attribute (all numeric-valued), the description and the units are shown:

B. Statistical Analysis of the Glass Dataset:

Attribute	Mean	Standard Deviation
RI	14.408	0.817
Na	13.408	0.817
Mg	2.685	1.442
AI	1.445	0.499
Si	72.651	0.775
K	0.497	0.652
Ca	8.957	1.423
Ba	0.175	0.497

Table (a): statistical analysis of glass dataset

C. Related work in Glass Dataset

For the long time the research in glass prediction have been conducted. The key goal of this experiment is to predict the variables that lead to the cause of infection at high risk for glass and to offer a precautionary action for every person who is at increased risk of the disease. Though acquiring the accuracy for recommendation by the physician assistance is a principal issue. Prevention measures are essential by providing increased awareness and treatment of glass. Most of the focuses on the impact and significance of precautionary actions for disease occurrences and particularly with cost effective way are resulted by these measures. [4] proposes a risky model which includes the attributes like Age, BMI, waist circumference, history of antihypertensive, drug treatment, high blood pressure etc.. The physical activity and daily consumption of fruits like berries, or vegetables as categorical variables. [2]proposes the NN model to identify and specify the risk factors of the final model and the cholesterol level with symptoms like back pain, blood pressure, fatty food, weight index or alcohol index [5].

proposes an application using data mining technique considering the data of diabetic patients. They consider the medical variables such as blood pressure, BMI, glycaemia, or cardio-vascular, cholesterol that act as a risk in the model

III. METHODS DESCRIPTION

For this experiment three categories of classifiers are considered for implementation such as decision tree based, Meta level based classifier and Bayes approach. The classification techniques and the respective outcomes of each implementation are tabulated. Then the final outputs are demonstrated as results.

A. Rules

a. Conjunctive

The numeric and nominal values of the class labels are predicted by the class implementation using single conjunctive rule learner. A rule contains the antecedents "AND" and the consequent value together as class by applying classification/regression. Here, the consequent may be distribution of the accessible classes such as mean for a numeric values in the dataset. If the test instances are not with the rule constrains, then the default class distributions predicts the value of the data that are not included in the rule are the training data. This learner chooses the antecedent by computing the increased information by each antecedent and it reduces the generated rule using Reduced Error Pruning (REP) or by simple prepruning method based on the use of antecedent number.

b. Decision Table

Classes are constructed using simple decision table using majority classifier. The options include crossVal, which sets the amount of folds employed for cross validation the second parameter is debug that is set as true for the classifier that may provide output with additional information to the console. The display Rules is helps to set when corresponding rule values are to be printed. Evaluation measure helps to determine the performance of the corresponding attribute permutations that can be applied for the decision table. The search technique discovers the better attribute combinations that can be used for decision table. Use IBk classifier for the Sets while IBk can be used as an alternative of the majority class.

c. Zero R

Zero – R Classifier helps in building and predicting the mean or the mode. It has the option namely debug that is set to true only if the classifier may obtain additional output infomation to the console.

B. Bayes Classifiers

It has two kinds of classifier

a. Naive Bayes

Naive Bayes classifiers are constructed using estimator classes with numeric estimator with precision values that are chosen for the analysis of training data. Therefore the classifiers that are modifiable classifier need the functionality to use the Naive Bayes modifiable classifier. The Naive Bayes classifier uses default precision value of 0.1 for all numeric attributes while constructing the classifier and so it is called as zero training instances.

b. Bayes Net

Bayes Network helps in classifying by learning the patterns using different search algorithms and by quality measures. The base class used for Bayes Network classifier supplies the data structures that provide regular facilities for all learning algorithms in Bayes Network such as K2 and B algorithm.

C. Meta Classifiers

a. Adaboost

Adaboost techniques are applied for improve the nominal class by applying the M1 technique, therefore to tackle the nominal class problems. It helps in enhancing the performance but sometimes leads to an issue of over fitting.

b. Bagging

The classifier bagging is used for reducing the variance therefore the value of classification and regression depends on the base learner. It generates samples of B bootstrap that used for training data by applying random sampling technique with substitute. Training the classifier or the regression function by applying such bootstrap samples are majority results in classification. The regression function computes the average predicted values that reduces the variation and enhances the performance as it is the unstable classifiers that differ considerably by little modification in the data set, e.g., CART.

c. Attribuite selected classifier

Dimensionality reduction methods are applied for both the training data and validation set by attribute selection technique, later the data conceded on to the classifier for training**.** The base classifier is the Classifier function itself, while the parameter such as debug is set to true, then the classifier may obtain additional information as output for the console. The parameter evaluator is set, therefore it can be used throughout the attribute selection phase before invoking the classifier and the search parameters are set with subsequent search method.

IV. TRAINING AND TESTING WITH SELECTED CLASSIFIERS

The above twelve classifiers are considered for learning the glass dataset given in 2.1. The ‘a’ tables show the accuracies without selection of attributes, whereas ‘b’ tables show the accuracies with selection of attributes. We use best first search, greedy search for selecting the subset evaluation of attributes using Weka tool and the output shows the reduced attribute set [2] [6] –[8].

A. Results obtained for the corresponding classifier

1 (a) Results with all attributes

Rules Classifiers	Accuracy
Conjunctive	44.394
Decision Table	52.122
ZeroR	35.514

Table (a) Rules Classifier Iterations to get Maximum Accuracy

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

1 (b). Results with selected attributes

Rules Classifiers	Accuracy
Conjunctive	46.266
Decision Table	53.121
ZeroR	35.524

Table (b). Rules Classifier Iterations to get Maximum Accuracy

2 (a) Results with all attributes

Meta Classifiers	Accuracy
Adaboost	44.856
Bagging	60.45
Attribuite selected classifier	59.34

2 (b) Results with selected attributes

Meta Classifiers	Accuracy
Adaboost	74.349
Bagging	75
Attribuite selected classifier	61.0469

Table2 (a,b) . Meta Classifier Iterations to get Maximum Accuracy

3 (a) Result with all attributes

Bayes	Accuracy
Naïve Bayes	48.591
Bayes Net	65.543

3(b) Results with selected attribute

Naïve Bayes	Accuracy
Naïve Bayes	59.561
Bayes Net	68.432

Table 3 (a,b) .Bayes Classifier Iterations to get Maximum Accuracy

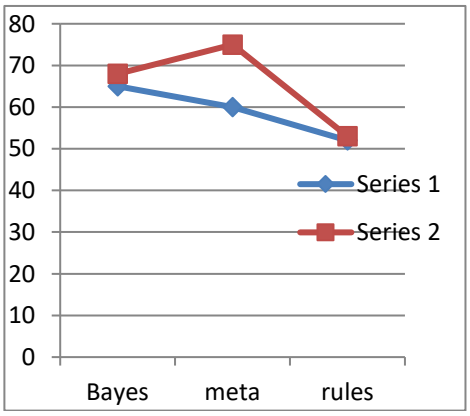

Fig 2: Comparison of original reduced dataset Vs for accuracy

The above table clearly shows the effect of attribute reduction in the process of classification of given glass dataset. The graph in fig shows the difference in accuracy results

V. CONCLUSION

The above results shows that the reduced attribute dataset obtains more accurate value than the original dataset that are previously used and this experiment helps in formulating the better system for identifying the glass material. However while considering the dimension of the dataset, in future research it is planned to put up with large dataset or with aggregating small data set into better size.

REFERENCES

1. https://www.waset.org/journals/ waset/v68/v68-21.pdf world academy of science, engineering and technology, 2012.
2. H.Dunham, Data Mining, Introductory and Advanced Topics, Prentice Hall, 2002
3. Source about wekahttp: //www.cs.waikato. ac.nz/ ml/weka/
4. A.Kumaravel, Pradeepa.R, Efficient molecule reduction for drug design by intelligent search methods Int J Pharm Bio Sci 2013 Apr; 4(2): (B) 1023 - 1029
5. A.Kumaravel, Udhayakumarapandian.D, Consruction Of Meta Classifiers For Apple Scab Infections , Int J Pharm Bio Sci 2013 Oct; 4(4): (B) 1207 - 1213
6. A.Gelman, Y. S. Su, M.Yajima, J. Hill, M. Pittau, J. Kerman, and T. Zheng, "arm: Data Analysis Using Regression and Multilevel/Hierarchical Models," R package version 1.5-02.://CRAN.Rproject. org/package=arm,2012.
7. L. Breiman, " RandomForests," inMachine Learning, vol. 45, pp. 5-32, 2001.
8. http://ipm.ncsu.edu/apple/chptr5.html
9. Dietterich, T. G., Jain, A., Lathrop, R., Lozano-Perez, T. (1994). A comparison of dynamic reposing and tangent distance for drug activity prediction.Advances in Neural Information Processing Systems, 6. San Mateo, CA: Morgan Kaufmann. 216--223.
[10] A.Stensvand, T. Amundsen, L. Semb, D.M. Gadoury, and R.C. Seem. 1997. Ascospore release and infection of apple leaves by conidia and ascospores of Venturia inaequalis at low temperatures. Phytopathology 87:1046-1053.

AUTHORS PROFILE

A.Rama is working as an Assistant Professor in the department of information technology, Bharath Institute of higher Education and Research, Chennai. Her research area includes visual computing and machine learning. Email: rama_j1@yahoo.com

Dr. Nalini Chidambaram is working as a professor in the department of computer science and Engineering, Bharath Institute of Higher Education and Research, Chennai. Her research area includes Data mining, social networking and Image Processing., She has Published more than 70 papers in scopus indexed journal and approximately 100 papers in high indexed journal. Email: nalini.cse@bharathuniv.ac.in