Golden Research Thoughts
Volume 2, Issue. 12, June. 2013
ISSN:-2231-5063 Available online at www.aygrt.isrj.net

ORIGINAL ARTICLE

ENSEMBLE APPROACH FOR RULE EXTRACTION IN DATA MINING.

HITESH NINAMA

Assistant Professor
School of Computer Science & Information Technology,
Devi Ahilya Vishwavidyalaya, Takshashila Campus, Khandwa Road,
Indore (M.P.) India

Abstract:

A major drawback with neural networks is that the models produced are opaque; i.e. they do not permit human inspection or understanding. A decision tree model, on the other hand, is regarded as comprehensible since it is transparent, making it possible for a human to follow and understand the logic behind a prediction. Although accuracy is the prioritized criterion for predictive modeling, the comprehensibility of the model is often very important. A comprehensible model makes it possible for the user to understand not only the model itself but also why individual predictions are made. Traditionally, most research papers focus on high accuracy, although the comprehensibility criterion is often emphasized by business representatives. Clearly, comprehensibility is very important for data mining technique. Since techniques producing opaque models normally will obtain highest accuracy, it seems inevitable that the choice of technique is a direct trade-off between accuracy and comprehensibility. With this trade-off in mind, several researchers have tried to bridge the gap by introducing techniques for transforming opaque models into transparent models, keeping an acceptable accuracy. Most significant are the many attempts to extract rules from trained neural networks. And this technique of transforming opaque model into transparent model is called as Rule Extraction (RE). Within the machine learning research community it is, however, also well known that it is possible to obtain even higher accuracy, by combining several individual models into ensembles. The overall goal when creating an ensemble is to combine models that are highly accurate, but differ in their predictions. The common ensemble techniques are probably bagging, boosting and stacking, all of which can be applied to different types of models and perform both regression and classification. Most importantly; bagging, boosting and stacking will, almost always, increase predictive performance over a single model. The proposed approach converts the opaque model into transparent model and at the same time maintains acceptable level of accuracy. Our idea is to first produce an opaque model by applying a data mining algorithm neural network to produce accurate model that optimizes the accuracy and then extract the rules from this opaque model by applying rule extraction algorithm to convert the opaque model into transparent model. Further we apply the same algorithm multiple times i.e. to devise ensemble method for rule extraction that leverages the power of multiple methods to achieve greater comprehensibility than any individual approach. Experiments are conducted to show that the proposed work reduces the accuracy vs comprehensibility tradeoff.

KEYWORDS:

Rule Extraction, Data Mining, Decision Tree, Ensemble approach, Bagging, Boosting, Neural Network, Accuracy, Comprehensibility.

Title :ENSEMBLE APPROACH FOR RULE EXTRACTION IN DATA MINING.
Source:Golden Research Thoughts [2231-5063] HITESH NINAMA yr:2013 vol:2 iss:12

INTRODUCTION

The problem of predictive modeling is that the produced model is either accurate or comprehensible. If we want to achieve accuracy, comprehensibility gets sacrificed and vice versa. Predictive models are mainly used for decision making and accuracy is the main motive of these models but comprehensibility is often very important. Because a comprehensible model makes it possible for the data miner to understand not only the model itself but also why individual predictions are made. So the first problem that is to be solved is to obtain an opaque model to achieve accuracy and the second is to transform opaque model to transparent model. So the main aim of proposed research work is to implement a novel technique that reduces the accuracy vs. comprehensibility trade off. That is to use rule extraction to transform an opaque model with high accuracy to comprehensible model while retaining approximately same level of accuracy.

OBJECTIVES

Following are the objectives of the proposed work:

A comparative study of existing Rule Extraction.
To implement ensemble approach for Rule Extraction to convert opaque model into transparent model.
Result analysis of the proposed algorithm in terms of accuracy.
To compare the performance of the proposed algorithm with the existing algorithms in terms of accuracy and comprehensibility.

LITERATURE REVIEW

We focused over literature survey on the following research work: H. Johan et al. given a formal definition of rule extraction and comment on the inherent tradeoff between accuracy and comprehensibility [2]. S. Huber et al. suggested that Artificial Neural Networks (ANNs) is a powerful tool for pattern recognition, decision problems or predication applications [3]. T. Löfström et al. had proposed that whether it is possible to predict that rule extraction will produce an accurate model by just analyzing the characteristics of the data set [4]. S. M. Kamruzzaman et al. had proposed that back propagation Artificial Neural Networks provides more accuracy for classification as compared to decision trees. They argued that the ANNs are generally regarded as black boxes because the predictions generated by them are not understandable as those of decision trees [5]. K. Saito and R. Nakano (1997) proposed a connectionist method called as RF5 (Rule extraction from Facts version 5) for numeric law discovery. RF5 uses a new second order learning algorithm called as BPQ based on quasi- Newton method for training the neural network. It also uses MDL (Minimum Description Length) criterion for finding an adequate number of hidden units in the network [8]. Schmitz et al. (1999) proposed a novel artificial neural-network decision tree algorithm (ANN-DT). This algorithm is used to extracts binary decision trees from a trained neural network [9]. Olcay Boz (2000) introduced a new method DecText for extracting the decision trees from trained neural network. They have developed four new splitting methods (SetZero Split, SSE Split, ClassDiff Split and Fidelity Split) for extracting the decision tree from the network [10]. K. Saito and R. Nakano (2002) mainly overcome the problem of RF5 algorithm. So they proposed a new framework and method called as RN2 for extracting regression rules from trained neural network for the data set that contain both nominal and numeric variables [11]. Trelak (2003) et al. proposed a method called REX for extracting the fuzzy rules from trained neural network. Their proposed provide a description of fuzzy sets in addition to the fuzzy rules [12]. F. Chen (2004) proposed a novel algorithm BUR (Bust unordered rule) for extracting the rules from SVM. The approach consists of two phases learning and pruning [13]. Huysmans (2006) et al. presented a new algorithm ITER for pedagogical regression rule extraction. ITER can extract regression rules from a trained black box model [14]. D. M. Escalante (2009) et al. proposed a novel method called as ENREDD that extract the knowledge from inherently distributed data without moving it from its original location, completely or partially, to other locations for legal or competition issues [1]. Huynh T.Q. (2011) et al. suggested a different approach for extracting the rules in which they modify the error backpropagation training process so that it learns a different hidden layer representation of input patterns than would normally occur [15]. M. Craven et al. devised a novel RE algorithm TREPAN that address the generality and scalability issues. Their algorithm can be applied to a wide variety of hard to understand models including ensembles [6]. L. Niklasson et al. had surveyed that techniques producing transparent models, either directly from the dataset, or from an opaque model, could benefit from using an oracle guide

[7].

PROPOSED RESEARCH METHODOLOGY

The proposed approach Ensemble method for rule extraction in data mining applies the rule extraction algorithm in ensemble fashion that leverages the power of multiple models to achieve greater comprehensibility than any individual approach.

TOOLS & TECHNOLOGIES USED

Following is the overview of the tools and technologies used for the proposed work:

Tool : Weka (3.6.2)
Weka is a machine learning toolkit.
Technology: VB 6.0

Visual Basic is the third-generation event-driven programming language and integrated development environment (IDE) from Microsoft for its COM programming model.
Operating System: Windows
Hardware specification: P4 processor

RESULT ANALYSIS

We had run our work on Wine and Iris data set downloaded from UCI repository site. The details for the data sets are as follows:

Wine data set

Data set: wine_data
Instances: 178
Attributes: 14

Alcohol, Malic acid, Ash, Alcalinity of ash, Magnesium, Total phenols, Flavanoids, Nonflavanoid phenols, Proanthocyanins, Color intensity, Hue, OD280/OD315 of diluted wines, Proline

Class
Class 1, Class 2, Class 3
Iris data set
Data set: iris_data
Instances: 150
Attributes: 5
sepal length, sepal width, petal length, petal width
class
Class Iris-setosa, Class Iris-versicolor, Class Iris-virginica

RESULTS FOR NEURAL NETWORK ALGORITHM

We have executed the neural network algorithm on Wine and iris data set and obtained the results as shown in table 1. The table also compares the implemented neural network algorithm with the existing neural network (weka implementation) algorithm.

SN	Data Set	No of Instances	Accuracy (%)	
			Existing NN	Implemented NN
1	Wine	178	97.19	98.31
2	Iris	150	97.33	98.00

Table 1: Accuracy measure for neural network algorithm

From the generated results it is clear that the accuracy obtained for neural network is approximately 98 %. The accuracy obtained for the implemented neural network algorithm is better as compared to existing algorithm for both the data sets Wine and Iris as shown in graph of figure 5.10.

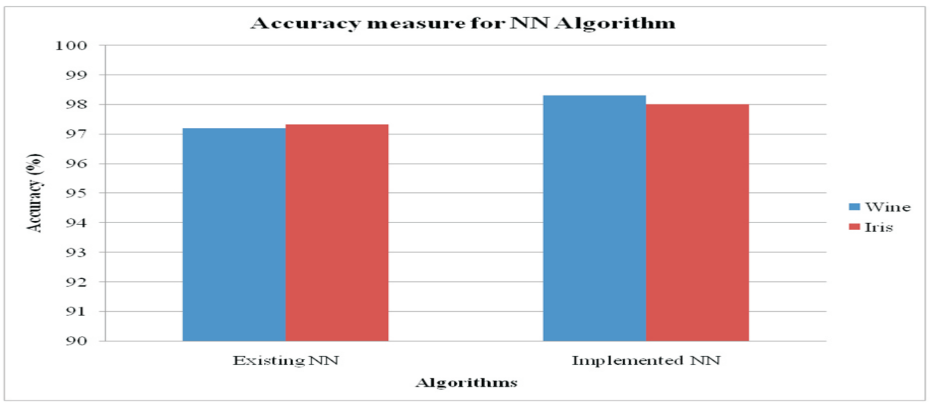

Figure 5.10: Accuracy Comparison for NN algorithm

Results for Rule extraction algorithm

Table 2 contains the results obtained by applying the Rule Extraction (DecText) algorithm on opaque model (output of neural network).

SN	Data Set	No of Instances	% of Correctly Classified Instances	% of Incorrectly Classified Instances
1	Wine	178	93.83	6.17
2	Iris	150	97.33	2.66

Table 2: Accuracy measure for Rule Extraction algorithm

From the generated result it is clear that the accuracy obtained for neural network is approximately 98% but after applying the Rule Extraction algorithm on neural network output the accuracy obtained is approximately 94% for Wine data set and 97% for Iris data set. There is a decrease in the accuracy but the output is comprehensible that is understandable because the output is in the form of rules. Hence we achieve the aim of converting the opaque model into transparent model. Now we create ensembles in order to improve accuracy. Results for ensemble creation are discussed in the next section.

Results for Ensemble creation

Table 3 contains the results obtained after creating ensembles using boosting approach.

SN	Data Set	No of Instances	% of Correctly Classified Instances	% of Incorrectly Classified Instances
1	Wine	178	96.32	3.68
2	Iris	150	97.83	2.17

Table 3: Accuracy measure for Ensemble Creation

From the generated result it is clear that the accuracy obtained for Rule Extraction is approximately 94% for Wine data set and 97% for Iris data set. There is a decrease in the accuracy but the output is comprehensible that is understandable. For achieving comprehensibility, accuracy get sacrifice. In order to reduce accuracy vs comprehensibility trade - off our approach create ensembles. Accuracy obtained for ensemble creation is approximately 96% for Wine data set and 97% for Iris data as shown in table 3. Hence we achieve the aim of converting opaque model into transparent model and at the same time maintaining acceptable level of accuracy that is comparable with the accuracy of neural network algorithm. Table 4 compares the accuracy obtained on each step of proposed approach.

SN	Data Set	No of Instances	Accuracy (%)		
			NN	RE	EA
1	Wine	178	98.31	93.83	96.32
2	Iris	150	98.00	97.33	97.83

Table 4: Accuracy comparison for each step of proposed approach

The accuracy obtained for the neural network algorithm is approximately 98% for Wine and Iris data set. As we apply Rule extraction algorithm on this opaque model (neural network) to convert it to transparent model accuracy gets decreased to 94% for Wine data set and 97% for Iris data set. Hence we achieve comprehensibility but there is loss in accuracy. Finally accuracy gets improved by creating ensembles. The accuracy of the proposed approach is comparable with the accuracy of opaque model as shown in graph of figure 5.11.

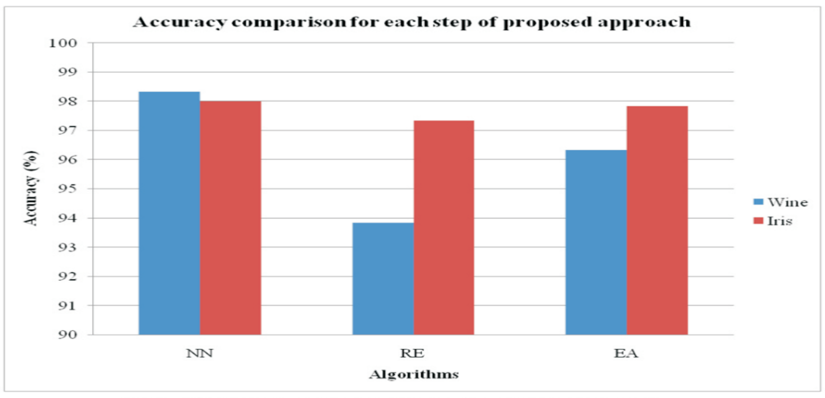

Figure 5.11: Accuracy comparison for each step of proposed approach

Comparison of proposed approach with existing techniques

This section compares the accuracy obtained for proposed approach (Ensemble Approach for Rule Extraction) with the existing predictive modeling algorithms for Wine and Iris data set. Results for C4.5, neural network (NN), CART and proposed approach Ensemble Approach for Rule Extraction (EA) are summarized in table 5.

SN	Data Set	No of Instances	Accuracy (%)			
			NN	C4.5	CART	EA
1	Wine	178	97.19	93.82	89.32	96.32
2	Iris	150	97.33	96.00	95.33	97.83

Table 5: Accuracy comparison of proposed approach (EA) with existing techniques

From the obtained results it is clear that the proposed approach reduces the accuracy vs comprehensibility trade - off. Approach converts the opaque model into transparent model while retaining approximately same level of accuracy as shown in graph of figure 5.12.

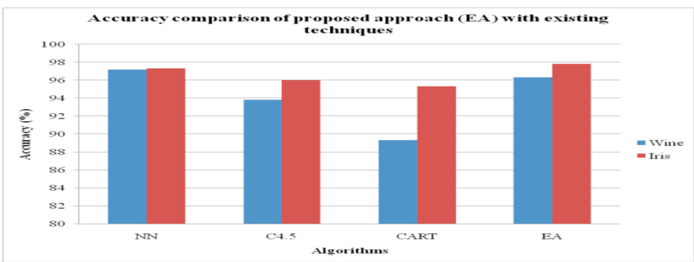

Figure 5.12: Accuracy comparison of proposed approach (EA) with existing techniques

CONCLUSION

The limitation of neural network and support vector machine leads to the accuracy vs. comprehensibility tradeoff. Rule extraction can achieve this tradeoff because it possess the ability to extract human understandable description from these models. This observation is mentioned by many researchers but they had not provided the output of these algorithms to show that really the output is comprehensible or not. We had shown the output for both opaque model and transparent model by running neural network and decision tree induction algorithm on Iris data set by using Weka tool. The generated output makes it clear for one that why the output of the neural network is not comprehensible but the output of decision tree induction algorithm is human understandable. Also the accuracy measures for both the algorithms are also given that is the accuracy for neural network is 97.33% whereas for decision tree induction algorithm it is 96.0%. This also shows that the neural networks are more accurate than decision tree algorithms. The main contribution of the proposed research work is the implementation of Ensemble approach for rule extraction in data mining. The proposed work obtains better comprehensibility.

FUTURE WORK

In this paper we have evaluated the proposed approach only for accuracy and comprehensibility. There are several other criterions for evaluating rule extraction algorithm such as fidelity, generality and consistency. So in future we can concentrate to cover these factors also.

REFERENCES

D. M. Escalante, M. A. Rodriguez, A. Peregrin, “An Evolutionary Ensemble-based Method for Rule Extraction with Distributed Data⋆”. In Proceedings of the 4th International Conference on Hybrid Artificial Intelligence Systems HAIS '09, pp. 638-645, 2009.

H. Johan, B. Bart and V. Jan, “Using Rule Extraction to Improve the Comprehensibility of Predictive Models”. In Open Access publication from Katholieke Universiteit Leuven, pp.1-56, 2006.

S. Huber, M. Rohde, M. Tamme, “Rule Extraction from Artificial Neural Networks”, PS Natural Computation, SS 2006.

T. Löfström, U. Johansson, “Predicting the Benefit of Rule Extraction -A Novel Component in Data Mining”. In Human IT 7.3, pp. 78-108,2005.

S. M. Kamruzzaman, Md. M. Islam, “Extraction of Symbolic Rules from Artificial Neural Networks”. In Journal of WASET Transactions on Science, Engineering and Technology, Vol. 10, pp. 271-277, Dec. 2005.

M. Craven, J. Shavlik, “Rule Extraction: Where Do We Go from Here?”, University of Wisconsin Machine Learning Research Group working Paper, pp. 99-1, 1999.

U. Johansson and L. Niklasson, “Evolving Decision Trees Using Oracle Guides”, In IEEE Symposium on Computational Intelligence and Data Mining, CIDM '09, pp. 238-244, 2009.

K. Saito and R. Nakano, “Law discovery using neural networks.” In Proceedings of the 15th International Joint Conference on Artificial Intelligence, pp. 1078-1083, 1997.

G.P.J. Schmitz, C. Aldrich and F.S. Gouws, “ANN-DT: An algorithm for extraction of decision trees from artificial neural networks.” IEEE Transactions on Neural Networks, 10(6):1392-1401, 1999.

O. Boz., “Converting A Trained Neural Network To A Decision Tree DecText - Decision Tree Extractor.” PhD thesis, Lehigh University, Department of Computer Science and Engineering, 2000.

K. Saito and R. Nakano, “Extracting regression rules from neural networks.” Neural Networks, 15(10). pp. 1279-1288, 2002.

U. Markowska-Kaczmar and W. Trelak, “Extraction of fuzzy rules from trained neural network using evolutionary algorithm.” In European Symposium on Artificial Neural Networks (ESANN), pp. 149-154, 2003.

F. Chen, “Learning accurate and understandable rules from SVM classifiers.” Master's thesis, Simon Fraser University, 2004.

J. Huysmans, B. Baesens and J. Vanthienen, “ITER: an algorithm for predictive regression rule extraction.” In 8th International Conference on Data Warehousing and Knowledge Discovery (DaWaK 2006). Springer Verlag, lncs, 2006.

Huynh T.Q., Reggia J.A., “Guiding Hidden Layer Representations for Improved Rule Extraction From Neural Networks.” In IEEE Transactions on Neural Networks, pp. 264-275, 2011.