A Survey on Rule Extraction for Achieving a Trade off Between Accuracy and Comprehensibility

Vikash Gupta
Global institute of Technology,sitapura
vikash.gupta@gitjaipur.com
Mamta Sakpal Surve
Global institute of Technology,sitapura
Mamta.simran15@gmail.com

----------ABSTRACT----------

Over the last decade, various data mining techniques such as Artificial Neural Network and support vector machine have demonstrated superior performance to many other classification techniques in a variety of application areas. However, they provide highly accurate model but they have an inability to provide an explanation, or comprehensible justification, for the solutions they provide as the output produce is in the form of weights and nodes [2]. Some techniques like decision tree provide comprehensible solution which is easy to interpret as these are in the form of rules or tree like structure but are not as accurate as neural network [4]. But for making any decision both accuracy and comprehensibility is required. Therefore an approach or technique is required through which we can generate a model which can provide a tradeoffs between both accuracy and comprehensibility. Rule extraction is the technique which we can use to achieve this trade-off.

In this paper, we conduct a formal review of the area of rule extraction from ANN and decision tree. We presented the motivation and meaning of rule extraction. Finally a survey of several well known rule extraction techniques is presented It is concluded that there is no one algorithm that can be favoured in general. However we find that ANN produces the most accurate rule set and have the highest fidelity but we required both comprehensibility and accuracy so we are introducing genetic rule extraction which somehow achieve a trade-off between accuracy and comprehensibility.

Keywords: Data mining; Rule extraction; ANN, genetic algorithm.

Introduction

Recent advances in data collection and storage technology has made it possible for organizations to accumulate huge amounts of data at moderate cost. Exploiting this stored data, in order to extract useful and actionable information, is the overall goal of the generic activity termed data mining.

Data mining techniques are the result of a long process of research and product development. This evolution began when business data was first stored on computers, continued with improvements in data access, and more recently, generated technologies that allow users to navigate through their data in real time. Data mining takes this evolutionary process beyond retrospective data access and navigation to prospective and proactive information delivery. Data mining is ready for application in the business community because it is supported by three technologies that are now sufficiently mature.

According to Professor Arun K. Pujari:
"*Data mining is the non trivial extraction of implicit, previously unknown and potentially useful information from the data.*"

Classification, regression, clustering, association rules are some of the well known data mining approaches. The models generated by data mining techniques can be either descriptive or predictive.

A. Descriptive data mining

Descriptive data mining is also called as unsupervised learning because it analyzes the data objects without having knowledge about the class label. It works on the data set in which the class labels are not present. Data mining techniques such as clustering, association and feature extraction are descriptive data mining techniques. Clustering decompose the data objects into groups or clusters such that the objects of the same cluster have high similar characteristics in comparison to one another, but are very dissimilar to objects in other clusters.

B. Predictive data mining

On the other hand predictive data mining is also called as supervised learning because it analyzes the data objects for which the class label of each tuple is available. Predictive model is useful for making the predictions for an unknown value of an attribute. It can be considered as a two step

process learning and prediction. Learning means analyzing the training data by any data mining algorithm. Once the model is trained using training data it can be further used in the second step called as prediction. The model can be used to make the predictions for future data tuples, as well as provide deeper insight into the database. For example a forecast manager can obtain a predictive model for the weather data and further utilize it to predict about play, whether it will be "yes" or "no" The attribute for which we are making the predictions is often called as *target attribute*. In this example attribute play is the target attribute. If the target attribute is a discrete variable then the data mining technique is called as *Classification*. On the other hand if it a real number, the technique is *Regression*. [2]

Rule extraction

Rule extraction is the process of converting an opaque model into transparent model while retaining approximately same level of accuracy. It is a technique that represent the opaque model in a form that is human understandable as well as the accuracy level is comparable with the opaque model. So the aim of the rule extraction is not just to achieve comprehensibility but it also takes care of the accuracy of the model. Hence the technique is used for achieving accuracy vs. comprehensibility trade off. A number of definitions for Rule Extraction have been proposed by various researchers. One of them given by H. Johan, B. Bart and V. Jan:

"*Given an opaque predictive model and the data on which it was trained, produce a description of the predictive model's hypothesis that is understandable yet closely approximates the predictive model's behavior. The task of converting Opaque model into more easily understood representation is commonly referred as Rule Extraction.*" [4]

A large number of algorithms are available for rule extraction such as TREPAN, REX and SVM+ etc. These algorithms presents a reasonable description of the underlying model may be in the form of rules. Although the process is called as rule extraction but it is not necessary that the output of rule extraction is always in the form of rules. Any method that produces human understandable description of the model can be considered as a rule extraction technique. The output can be in different formats such as finite state machine, decision tree, graphical representation or rules.

A. Motivation for rule extraction

Following figure-- shows the output produced by Weka 3.6 tool for neural network algorithm applied on Iris data set. Iris data set is a publicly available data set from UCI Repository. There are 150 instances, 3 classes and 4 real valued attributes in the Iris data set.

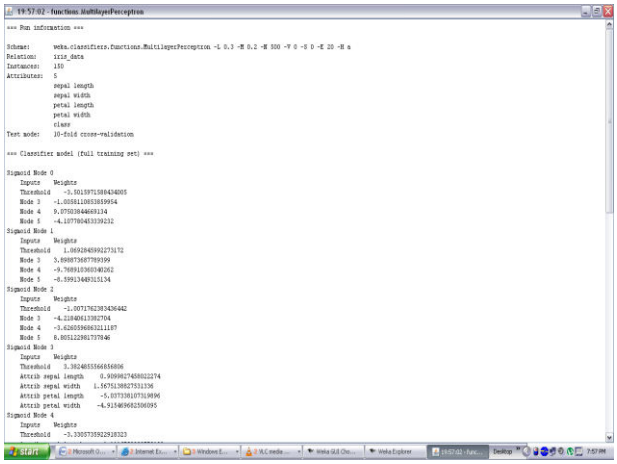


Figure : Output of neural network for Iris Data Set

Produced output is in the form of sigmoid nodes and weights. Although the accuracy obtained is 97.333%, but the output is not comprehensible. One cannot understand the logic behind the output. It is difficult to understand that why a particular prediction has been made.

On the other hand figure shows the output produced by Weka 3.6 tool for C4.5 algorithm (Decision tree induction algorithm) applied on the same Iris data set.

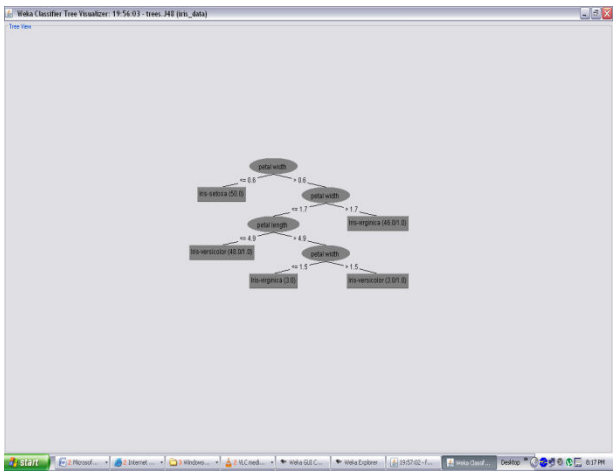


Figure : Output of C4.5 algorithm for Iris Data Set

Produced output is in the form of Decision tree. The tree can be easily converted to IF - then rule. Although the accuracy obtained is 96.0% (Less than neural network output for same data set), but the output is comprehensible. One can easily understand the logic behind the output. It is easy to understand that why a particular prediction has been made with the help of tree.

If comprehensibility is the main issue for predictive model then there is no need to introduce a middle step for creating an opaque model and we can directly obtain a transparent model for the training data set. The main basis for introducing a technique or model of creating an opaque model for various applications so that the accuracy achieved by opaque models should be better than transparent models.

There are a number of motivating factors for rule extraction such as [3]:

- **Justification***:* It is necessary for a user to know how the decision support or learning system arrives at a particular decision. Justification is necessary in many applications such as security and medical diagnosis systems in which the user must have faith on the system. For having faith on system justification for decision is important. Hence if the user understands the logic behind the decision then he or she is confident about the performance of the system.
- ***Elucidation*:** It is not always compulsory to generate full explanation about the opaque model; it will be sufficient to explain only the prediction or decision about individual instances. If the knowledge embedded in the opaque model is comprehensible then it can be easily used to explain the decision made for a particular case.

C. Properties of rule extraction algorithms

According to R. Andrews et al., rule extraction algorithms are subdivided by different properties which are as follows:

- ***Expressive power***: It targets the representation language of the extracted rules by a RE algorithm. It gives an idea about the different types of rules produced by various RE algorithms. There are algorithms which produce "if-then" rules based on Boolean algebra or propositional logic. Another possibility would be fuzzy rules like "if temperature is high then set control to low" based on fuzzy logic.
- ***Translucency*:** One more way of classifying rule extraction algorithms is based on how they view the trained neural net. One possibility to view ANN as a black box. Input is given to the ANN, the output is received and rules are extracted based on the relation between input and output. This technique is called as black-box or pedagogical. The other approach looks inside the trained neural net and utilize knowledge about connections and weights. This class of rule extraction algorithms is called decomposition.
- **Dependency on ANN architecture**: The third way to classify rule extraction algorithms is based on ANN architectures. ANN architecture can be Hopfield or binary feed forward networks.
- ***Quality of extracted rules*:** It describes how well the required explanation is performed by extracted rules.
- **Algorithmic complexity**: It classifies the rule extraction algorithms based on how efficient the underlying rule extraction algorithm is. [5]

D. Evaluation criterion for rule extraction algorithms

There are several criteria used for evaluating rule extraction algorithms. In [10] M. Craven and J. Shavlik listed five criteria:

- ***Comprehensibility*:** The extent to which extracted representations are humanly comprehensible.
- ***Fidelity*:** The extent to which extracted representations accurately model the networks from which they were extracted.
- ***Accuracy*:** The ability of extracted representations to make accurate predictions on previously unseen cases.
- ***Scalability*:** The ability of the method to scale to networks with large input spaces and large numbers of weighted connections.
- ***Generality*:** The extent to which the method requires special training regimes or places restrictions on network architectures.

II Background

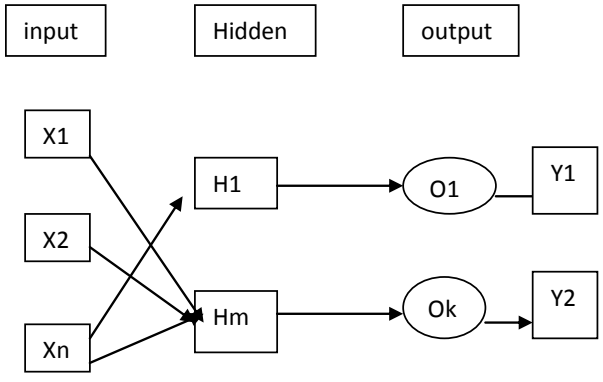
We will use two data mining technique in this paper and compare output of both for accuracy and comprehensibility on different dataset.

[1] Artificial Neural Network

Artificial neural network have become a powerful tool in tasks like pattern recognition, decision problem or predication applications. It is one of the newest signal processing technology. We can say ANN is an adaptive, most often non linear system that learn to perform a function from data and that adaptive phase is normally training phase where system parameter are change during operations. After the training is complete the parameter are fixed. If there are lots of data and problem is poorly understandable then using ANN model is accurate, the non linear characteristics of ANN provide it lots of flexibility to achieve input output map. Artificial Neural Networks, provide user the capabilities to select the network topology, performance parameter, learning rule and stopping criteria. ANN based solutions are extremely efficient in terms of development time and resources, and in many difficult problems artificial neural networks provide performance that is difficult to match with other technologies.

[1.1] Multi layer preceptron

An ANN is often represent as a graph with neuron as nodes and connection as weight arcs. We show a sample layer ANN with N inputs layer with m unit hidden layer and k unit output layer.



[1.2] Back propagation

The MLP is the function f that produces outputs values from the equation y=f(x;q). The purpose of the training is consequently to find the best q values, which for MLPs correspond to the architecture and the weights associated with the arcs. Since the architecture is often chosen before training starts, the term neural network training normally refers to the process of changing the weights from observed differences between desired and actual output. The overall purpose of supervised MLP training is to minimize an error function over the training set. Although there are many different algorithms for MLP learning, the most common is still some variation of back propagation.

[2] Genetic rule extraction

We have suggested a method for rule Extraction based on genetic programming (GP). The method named G-REX (Genetic Rule Extraction), is very general as it representation language depend upon the terminal and functional set and fitness function crossover rate and mutation determine the accuracy and comprehensibility. The easiest way to Understand these process is to show black-box rule extraction as an attribute of predictive modeling where the original input patterns are used with the corresponding training set and target variable.

G-REX uses GP when searching for the best solution is required. It work as first the initial population is created randomly following the ramped half and half strategy. Rules are chosen for reproduction. The genetic operators (crossover and mutation) are applied to the selected rules in a standard fashion, and it is at all times made sure that the resulting programs are still syntactically correct. The exact parameters like crossover and mutation rates, number of individuals in the population and number of generations are normally found from initial experimentation. After several generations, the fittest program is chosen as the extracted rule

CONCLUSION

References

[1] A. C. Oza, Ph.D. “Ensemble Data Mining Methods”. In Encyclopedia of Data Warehousing and Mining, pp. 448–453, Idea Group Reference, 2006.

[2] 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. [2]R. Konig, U. Johansson and L. Nikalasson, “Genetic Programming – A Tool for Flexible Rule Extraction”. In Proceedings of IEEE Congress on Evolutionary Computation, pp. 1304-1310, 2007.

[3] 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.

[4] 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.

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

[6] 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.

[7] 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.

[8] 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.

[9] U. Johansson, R. KÄonig and L. Niklasson. “Automatically balancing accuracy and comprehensibility in predictive modeling”. In Proceedings of the 8th International Conference on Information Fusion, 2005.

[10] U. Johansson, T. Lofstrom, R. Konig, C. Sonstrod, L. Niklasson, “Rule Extraction from Opaque Models--A Slightly Different Perspective”. In Proceedings of the 5th International Conference on Machine Learning and Applications (ICMLA'06), pp. 22-27, Dec 2006.