Effective Classifier Pruning with Rule Information

Xiaolong Zhang[1,2], Mingjian Luo[1], and Daoying Pi[2]

[1] School of Computer Science and Technology,
Wuhan University of Science and Technology, Wuhan 430081 P.R. China
{xiaolong.zhang, mingjian.luo}@mail.wust.edu.cn
[2] Dept. of Control Science and Engineering,
Zhejiang University, Hangzhou 310027 P.R. China
dypi@iipc.zju.edu.cn

Abstract. This paper presents an algorithm to prune a tree classifier with a set of rules which are converted from a C4.5 classifier, where rule information is used as a pruning criterion. Rule information measures the goodness of a rule when discriminating labeled instances. Empirical results demonstrate that the proposed pruning algorithm has high predictive accuracy[1].

1 Introduction

Decision tree pruning is a kind of method to improve predict accuracy and avoid overfitting. There are some approaches previously proposed in this area. Minimum description length pruning [1] and minimal cost complexity pruning [2] generate a sequence of pruned trees and later select better one from them. It should be noted these methods prune a decision tree in a direct way, pruning nodes among the paths from a tree node to a tree leaf. It cannot delete a condition if the reorganization of the tree fails. This paper proposes a tree pruning method, pruning a tree with a rule set. The rule set is converted from a decision tree [3], where rule information proposed in [4] is used as pruning criteria. It prunes a tree in an indirect way. The tree pruning with a set of rules has some advantages over those done by a direct way. It can avoid the over pruning caused in the previous pruning methods especially when the training set is small.

Rule information [4] is used to calculate relationship between antecedent (conditions) and consequent (prediction) of a rule. Rule information is suited to describe the relationship among the conditions and the prediction. Moreover, a rule belief is employed to identify a rule if the rule is necessary to be pruned before a rule pruning starts. Empirical tests and comparisons show that our algorithm outperforms C4.5 in predictive accuracy.

[1] The work is supported by the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry, and the Project (No.2004D006) from Hubei Provincial Department of Education, P. R. China.

A. Hoffmann, H. Motoda, and T. Scheffer (Eds.): DS 2005, LNAI 3735, pp. 392–395, 2005.

2 Rule Information and Rule Belief

A decision tree can be converted to a set of rules. The rules with IF-THEN form are used to prune a tree. A rule is expressed with $\sum_{i=i}^{n}(A_1 = v_{ij}) \rightarrow (C = c_m)$, where $A_i = v_{ij}$ is a condition of the rule and $C = c_m$ the prediction of the rule, n is the number of the attributes; v_{ij} a value of the attribute A_i, c_m a class of the prediction C (assume m is among $1 \cdots k$, k denotes the number of the class values of C).

Rule information describes the mutual relationship for a rule between its conditions and its prediction. Rule information [4] is defined as

$$I(R) = \log_2 \frac{P(C = c_m | \prod_{i=1}^{n}(A_i = v_{ij}))}{P(C = c_m)}$$

where $P(C = c_m|(A_i = v_{ij}))$ is the proportion of attribute A_i with value v_{ij} under $C = c_m$, $P(C = c_m)$ is the proportion of class C with value c_m in the training set.

As the definition of $I(R)$, a rule with a larger value denotes the relationship between the condition part and the prediction part in a rule is tighter. If $C = c_m$ and all the examples in the training data are covered, $I(R)$ is the maximum $-\log_2 P(C = c_m)$. If the prediction of a rule cannot derive from inadequate conditions, its rule information is probably negative. Our algorithm only deals with the rules whose rule information ranges from 0 to $-\log_2 P(C = c_m)$. To identify a rule if it can be a candidate to be pruned, we define a concept "rule belief": $B(R) = \frac{I(R)}{-\log_2 P(C=c_m)}$. B(R) is a normalization of $I(R)$ with $-\log_2 P(C = c_m)$, which can be directly used to identify whether a rule should be deleted before pruning.

Both rule information and rule belief are used in our rule pruning algorithm. Rule information is used as a pruning criterion. Given a rule r, every step in rule pruning of r should not decrease the rule information $I(r)$. Rule belief is used to identify whether a rule should be directly deleted without entering rule pruning process.

3 Pruning Algorithm and Experiments

Our tree growing process is similar to that of C4.5. A built decision tree is then converted to a rule set. The algorithm mainly consists of: 1) Calculation of rule information and rule belief for each rule; 2) Deletion of the rules whose rule belief is less than a given threshold δ; 3) For each rule, any condition of the rule can be removed once it does not worsen the rule information of the rule. In every rule pruning, so that the increment of rule information should be greater than a value ϵ (given according to the noisy rate in the training data). For every condition of the rule, the algorithm tries to find the most effective pruning, where the increment of the rule information is maximum. The pruned rule with the most effective pruning is inserted into the pruned rule set R^p.

The experiments have extensively been evaluated with 4 data sets (Connect-4, Breast-cancer, Iris Plants, Credit-screening) selected from the well-known UCI data repository [5]. The predictive accuracy on their testing set is calculated. The results demonstrate our rule pruning algorithm outperforms C4.5 in predictive accuracy. For each data set, we have carried out several training and testing respectively. In case of Connect-4 (10-600-100), 10 training-testing tunes are performed; the first training set includes 600 examples; and each of the next training set is increased with 100 examples which are randomly selected from its data source. For Breast-cancer, Iris Plants, and Credit-screening, the above number is (10-50-50), (10-50-10), (9-300-30), respectively. Predictive accuracy is calculated with every training-testing tune by applying the pruned tree on the original data source except the training examples.

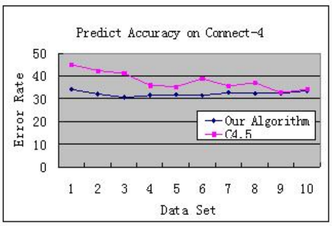

Fig.1-1. Predictive Accuracy on Connect-4

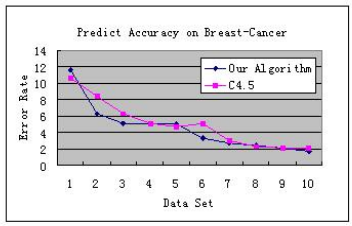

Fig.1-2. Predictive Accuracy on breast-cancer

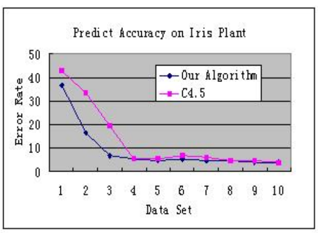

Fig.1-3. Predictive Accuracy on Iris Plant

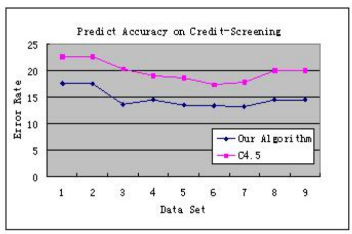

Fig.1-4. Predictive Accuracy on credit-screening

Fig. 1. Predict accuracy comparison

The results (see Fig. 1) show the predictive accuracy of these four conducted domains. Fig.1-1 shows both C4.5 and our algorithm have high error ratios (more than 30%) on Connect-4. The predictive accuracy obtained on Breast-Cancer (Fig. 1-2) and Iris Plant (Fig. 1-3) is reasonable, where the error ratios decrease when the training examples increase. However, for Credit-screening (Fig. 1-4), the increasing examples seem to be not improving the predictive accuracy. Both C4.5 and our algorithm have similar accuracy curves on these 4 domains. Our algorithm performs better than C4.5 on these 4 domains, and learns more accurate classifiers.

4 Conclusion

The proposed tree pruning method with rule information and rule belief performs better than C4.5 in the predictive accuracy. One of the future works is to explore the model optimization with the decision tree.

References

1. Mehta M., Rissanen J., Agrawal R. MDL-Based Decision Tree Pruning. *Proceedings of the First International Conference on KDD*, 216-221, 1995.
2. Breiman L., et al. Classification and Regression Trees. *Wadsworth & Brooks Press*, 1984.
3. Quinlan J. R. C4.5: Programs for Machine Learning. Morgan Kaufmann, 1993.
4. Hu D., Li H.X. Rule Mining and Rule Reducing Based on the Information of Rules. *Pattern Recognition and Artificial Intelligence*, 17(1), 2004.
5. Blake C., Merz C. UCI Repository of Machine Learning Databases. *Dept. of Information and Computer Science, University of California.*