Learning Predictive Clustering Rules

Bernard Ženko
Jožef Stefan Institute, Jamova cesta 39, SI-1000 Ljubljana, Slovenia
E-mail: bernard.zenko@ijs.si
http://www.fri.uni-lj.si/file/73407/zenko-phd-thesis.pdf

Thesis Summary

Keywords: machine learning, predictive clustering, rule learning

Received: December 14, 2007

The article presents the abstract of doctoral dissertation on learning predictive clustering rules.

Povzetek: Članek predstavlja povzetek doktorske disertacije o učenju pravil za napovedno razvrščanje.

1 Introduction

In the thesis [10] we developed and empirically evaluated a method for learning predictive clustering rules. The method [10, 9] combines ideas from supervised and unsupervised learning and extends the predictive clustering approach to methods for rule learning. In addition, it generalizes rule learning and clustering. The newly developed algorithm is empirically evaluated, in terms of performance, on several single and multiple target classification and regression problems. The new method compares favorably to existing methods. The comparison of single target and multiple target prediction models shows that multiple target models offer comparable performance and drastically lower complexity than the corresponding sets of single target models.

2 Thesis overview

The predictive clustering approach [1, 2] builds on ideas from two machine learning areas, predictive modeling and clustering [6]. Predictive modeling is concerned with the construction of models that can be used to predict some object's target property from the description of this object. Clustering, on the other hand, is concerned with grouping of objects into classes of similar objects, called clusters; there is no target property to be predicted, and usually no symbolic description of discovered clusters. Both areas are usually regarded as completely different tasks. However, predictive modeling methods that partition the example space, such as decision trees and rules are also very similar to clustering [7]. They partition the set of examples into subsets in which examples have similar values of the target variable, while clustering produces subsets in which examples have similar values of all descriptive variables. Predictive clustering builds on this similarity. As is common in 'ordinary' clustering, predictive clustering constructs clusters of examples that are similar to each other, but in general taking both the descriptive and the target variables into account. In addition, a predictive model is associated with each cluster which describes the cluster, and, based on the values of the descriptive variables, predicts the values of the target variables.

Methods for predictive clustering enable us to construct models for predicting multiple target variables which are normally simpler and more comprehensible than the corresponding collection of models, each predicting a single variable. So far, this approach has been limited to the tree learning methods. The aim of the thesis was to extend predictive clustering towards methods for learning rules, i.e., to develop a method for learning predictive clustering rules. Of the existing rule learning methods, majority are based on the sequential covering algorithm [8], originally designed for learning ordered rule lists for binary classification domains. We have developed a generalized version of this algorithm that enables learning of ordered or unordered rules, on single or multiple target classification or regression domains. The newly developed algorithm is empirically evaluated on several single and multiple target classification and regression problems.

3 Conclusion

The work presented in the thesis comprises several contributions to the area of machine learning. First, we have developed a new method for learning unordered single target classification rules. It is loosely based on the commonly used rule learning method CN2 [4, 3], but uses a generalized *weighted covering* algorithm [5].

Second, the developed method is generalized for learning ordered or unordered rules, on single or multiple target classification or regression domains. It uses a search

heuristic that takes into account several rule quality measures and is applicable to all the above mentioned types of domains.

The third contribution is the extension of the predictive clustering approach to models in the form of rules. The developed method combines rule learning and clustering. The search heuristic takes into account the values of both the target and the descriptive attributes. Different weighting of these two types of attributes enable us to traverse from predictive modeling to clustering.

The final contribution is an extensive empirical evaluation of the newly developed method on single target classification and regression problems, as well as multiple target classification and regression problems. Performance of the new method is compared to some existing methods. The results show that on single target classification problems, the performance of predictive clustering rules (PCRs) is comparable to that of CN2 rules and predictive clustering trees (PCTs), while in the case of unordered rules, PCRs are better than CN2 rules. Unordered PCRs are in general better than ordered PCRs. On multiple target classification problems, PCRs are comparable to PCTs, but PCRs tend to produce smaller rule sets than (transcribed) trees. Single target regression PCRs are comparable to existing regression rule methods, however, their performance is much worse than that of PCTs; on multiple target regression problems, PCRs are also much worse than PCTs. We believe the main reason that PCTs are better than PCRs on regression problems is the fact that PCTs use a state-of-the-art post-pruning method, while PCRs use no post-pruning. The comparison of the performance of single target and multiple target PCRs on multiple target problems shows, that multiple target prediction provides comparable accuracy as single target prediction, but multiple target prediction rule sets are much smaller than the corresponding single target rule sets.

References

[1] H. Blockeel (1998) *Top-down Induction of First Order Logical Decision Trees*. PhD thesis, Katholieke Universiteit Leuven, Belgium.

[2] H. Blockeel, L. De Raedt, J. Ramon (1998) Top-down induction of clustering trees. In *Proc of the ICML 98*, pp 55–63, San Francisco, Morgan Kaufmann.

[3] P. Clark, R. Boswell (1991) Rule induction with CN2: Some recent improvements. In *Proc of the* 5^{th} *EWSL*, pp 151–163, Berlin, Springer.

[4] P. Clark, T. Niblett (1989) The CN2 induction algorithm. *Machine Learning*, 3(4): 261–283.

[5] D. Gamberger, N. Lavrač (2002) Expert guided subgroup discovery: Methodology and application. *Journal of Artificial Intelligence Research*, 17:501–527.

[6] L. Kaufman, P. J. Rousseeuw (1990) *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley & Sons, New York.

[7] P. Langley (1996) *Elements of Machine Learning*. Morgan Kaufmann, San Francisco.

[8] R. S. Michalski (1969) On the quasi-minimal solution of the general covering problem. In *Proc of the FCIP 69*, pp 125–128, Bled.

[9] B. Ženko, S. Džeroski, J. Struyf (2006) Learning predictive clustering rules. In *Knowledge Discovery in Inductive Databases,* 4^{th} *Int Wshp, Revised Selected and Invited Papers*, pp 234–250. Springer.

[10] B. Ženko (2007) *Learning predictive clustering rules,* PhD thesis, University of Ljubljana, Slovenia.