Constructing Hierarchical Rule Systems

Thomas R. Gabriel and Michael R. Berthold

Data Analysis Research Lab
Tripos, Inc., 601 Gateway Blvd., Suite 720
South San Francisco, CA 94080, USA.
{tgabriel,berthold}@tripos.com

Abstract. Rule systems have failed to attract much interest in large data analysis problems because they tend to be too simplistic to be useful or consist of too many rules for human interpretation. We present a method that constructs a hierarchical rule system, with only a small number of rules at each stage of the hierarchy. Lower levels in this hierarchy focus on outliers or areas of the feature space where only weak evidence for a rule was found in the data. Rules further up, at higher levels of the hierarchy, describe increasingly general and strongly supported aspects of the data. We demonstrate the proposed method's usefulness on several classification benchmark data sets using a fuzzy rule induction process as the underlying learning algorithm. The results demonstrate how the rule hierarchy allows to build much smaller rule systems and how the model—especially at higher levels of the hierarchy—remains interpretable. The presented method can be applied to a variety of local learning systems in a similar fashion.

1 Introduction

Data sets obtained from real world systems often contain missing, noisy, or simply wrong records. If one attempts to build a rule model for such data sources, the result is either an overfitted and inherently complex model, which is impossible to interpret, or the model ends up being too simplistic, ignoring most of the interesting aspects of the underlying system as well as the outliers. But even this effect of outlier ignorance is often also not desirable since the very example that was excluded from the model may be caused by a rare but still extremely interesting phenomena. The real challenge is therefore to build models that describe all interesting properties of the data while still being interpretable by the human expert.

Extracting rule models from data is not a new area of research. In [9] and [13] algorithms were described that construct hyperrectangles in feature space. The resulting set of rules encapsulates regions in feature space that contain patterns of the same class. Other approaches, which construct fuzzy rules instead of crisp rules, were presented in [1,6,11] and [12]. All of these approaches have in common that they tend to build very complex rule systems for large data sets originating from a complicated underlying system. In addition, high-dimensional feature

M.R. Berthold et al. (Eds.): IDA 2003, LNCS 2810, pp. 76–87, 2003.

spaces result in complex rules relying on many attributes and increase the number of required rules to cover the solution space even further. An approach that aims to reduce the number of constraints on each rule individually was recently presented in [4]. The generated fuzzy rules only constrain few of the available attributes and hence remain readable even in case of high-dimensional spaces. However, this algorithm also tends to produce many rules for large, complicated data sets.

In this paper we describe a method that attempts to tackle this inherent problem of interpretability in large rule models. We achieve this by constructing a hierarchy of rules with varying degrees of complexity. Early attempts to build model hierarchies have been published previously. In [5] an approach to build unsupervised hierarchical cluster models was described, however, the resulting system of clusters does not offer great interpretability. A method to build hierarchies of rule systems was described in [8]. Here, an ensemble of rule sets with different granulation is built at the beginning. Starting with a coarse granulation (usually having only two membership functions for each attribute), the remaining rule sets exhibit increasingly finer granulation. In the resulting multi-rule table the grade of certainty for each rule is then used for pruning, that is, rules with a low grade of certainty will be removed from the rule set. However, this method still exhibits the complexity problem described above—in case of high-dimensional feature spaces, the resulting number of rules will increase exponentially with the dimensionality of the space. As a result, each rule relies on all attributes, making it very hard to interpret.

The approach presented here builds a rule hierarchy for a given data set. The rules are arranged in a hierarchy of different levels of precision and each rule only depends on few, relevant attributes thus making this approach also feasible for high-dimensional feature spaces. Lower levels of the hierarchy describe regions in input space with low evidence in the given data, whereas rules at higher levels describe more strongly supported concepts of the underlying data. The method is based on the fuzzy rule learning algorithm mentioned above [4], which builds a single layer of rules autonomously. We recursively use the resulting rule system to determine rules of low relevance, which are then used as a filter for the next training phase. The result is a hierarchy of rule systems with the desired properties of simplicity and interpretability on each level of the resulting rule hierarchy. We evaluate the classification performance of the resulting classifier-hierarchy using benchmark data sets from the European StatLog–Project [7]. Experimental results demonstrate that fuzzy models at higher hierarchical levels indeed show a dramatic decrease in number of rules while still achieving better or similar generalization performance than the fuzzy rule system generated by the original, non-hierarchical algorithm.

2 The Underlying Fuzzy Rule Algorithm

The underlying, non-hierarchical fuzzy rule learning algorithm is described in [4]. The algorithm constructs a set of fuzzy rules from given training data. The