Top-down Induction of First Order Logical Decision Trees

Hendrik Blockeel
K.U.Leuven, Dept. of Computer Science
Celestijnenlaan 200A, B-3001 Heverlee, Belgium
Hendrik.Blockeel@cs.kuleuven.ac.be

Although top-down induction of decision trees is a very popular induction method, up till now it has mainly been used for propositional learning; relational decision tree learners are scarce. This dissertation discusses the application domain of decision tree learning and extends it towards the first order logic context of Inductive Logic Programming.

Keywords: Machine learning, inductive logic programming, decision trees
Thesis advisors: Luc De Raedt and Maurice Bruynooghe

1. Situation

Top-down induction of decision trees is a very popular induction method. Until now, however, this method has been used mainly within the framework of attribute-value learning (AVL). This framework imposes a number of constraints on the representation of data and hypotheses. A more powerful formalism, inductive logic programming (ILP) does not impose such constraints and is therefore more generally applicable. Unfortunately, ILP is less mature than AVL and does not contain as many sophisticated and specialized techniques as AVL.

The aim of this dissertation is to upgrade induction of decision trees and some of the related sophisticated techniques from AVL to ILP.

2. Predictive Clustering using Decision Trees

In a first part of the dissertation a relatively general framework for induction called "predictive clustering" is described. Predictive clustering can be seen as a special case of clustering; it encompasses those types of clustering where the evaluation criterion for the clustering is related to a certain predictive accuracy criterion.

In this text the intra-cluster variance is used as the evaluation criterion. The computation of this variance presupposes a distance metric. It is demonstrated in the dissertation that a number of important learning tasks (classification, regression, unsupervised learning, pattern completion, and some other kinds of clustering) are special cases of predictive clustering; one only needs to select a suitable distance metric. For instance, when a classifier represented by a set of rules is induced, the body of a rule can be seen as an intensional description of a cluster; the head contains the prediction for that cluster. In a classification context the predictive accuracy criterion usually states that the probability of making a correct prediction (i.e., the predicted class equalling the actual class) should be maximised. It can easily be shown that maximising this probability amounts to minimising the variance along the prediction space in a cluster, using the equality distance over the classes ($d_{=}(c_1, c_2) = 1$ if $c_1 \neq c_2$, 0 otherwise) as a distance metric.

In a following chapter, the dissertation goes into more detail by discussing how induction of decision trees can be described at the general level of predictive clustering. An generic algorithm for induction of decision trees is proposed that generalises over many existing algorithms for induction of classification or regression trees (it is discussed how the parameter procedures of the generic algorithm can be instantiated for these specific tasks) and is also usable for unsupervised learning, pattern completion and other types of (conceptual) clustering.

The predictive clustering framework does not only generalise the classical approaches to induction; a number of modified induction methods that combine explicit induction (of decision trees or rule sets) with distance-based prediction can also be situated within the framework.

AI Communications 0 () 0
ISSN 0921-7126 / $8.00 © , IOS Press

3. First Order Logical Decision Trees

In a second part decision trees are defined in the context of first order logic, which is the representation formalism used in ILP, and study their properties. Roughly, a first order logical decision tree is defined as a binary decision tree where the tests in the nodes are existentially quantified conjunctions of literals (just like Prolog queries) and different nodes may share variables.

It is shown that first order logical decision trees have an advantage with respect to expressiveness over the kind of logic programs that are normally induced by most ILP systems. We call the latter *flat* logic programs, because the target predicate is defined immediately in terms of the background knowledge, without any intermediate predicates being defined (in the latter case we would call them *layered*).

Given a set of predicates, the set of all first order logical decision trees is a strict superset of the set of flat logic programs. In order to enable an ILP system to produce any theory that is representable by a first order decision tree, the ILP system should either have the possibility to invent auxiliary predicates (so that layered logic programs can be induced), use a format for its theories that allows the use of both universal and existential quantification, or induce Prolog programs with cuts (so-called first order decision lists).

The results concerning expressivity of these different formats are novel and are quite different from the ones obtained in attribute value learning, where the difference in expressivity between those formats is such that *given certain complexity bounds*, theories representable in one format may not be representable in another format, but disregarding complexity bounds the formats are equivalent.

Once these first order logical decision trees are defined and understood, it becomes possible to apply the proposed technique for predictive clustering by induction of decision trees within ILP. This gives rise to a generic ILP algorithm that, given suitable instantiations for its parameter procedures, can be used for a wide variety of tasks (classification, regression, unsupervised learning, pattern completion, conceptual clustering, ...)

4. Implementation and evaluation

A prototype implementation of the algorithm, called TILDE, is presented and evaluated empirically. The performance of the algorithm over a wide range of domains (classification, regression, clustering; both relational and propositional domains) is measured with respect to several evaluation criteria: the predictive accuracy of the induced theories, their complexity, and the efficiency of the induction process.

The main conclusions from this experimental evaluation are that TILDE is competitive with current state-of-the-art ILP systems in that the theories obtained with it usually are of high predictive quality and relatively simple. Specific advantages of the approach are its efficiency (within attribute value learning induction of decision trees is known to be one of the most efficient techniques, and this property is in part inherited by TILDE), and its very broad applicability.

In a second part of the experimental evaluation, special attention is given to the scalability of the technique. Using an alternative implementation of TILDE specially adapted towards the processing of large data sets (based on an existing propositional level-wise algorithm for decision tree induction), experiments are reported that involve data sets of over 100MB, which demonstrates the feasibility of mining databases of reasonable size with this technique. The technique is theoretically and empirically shown to scale linearly in the number of examples.

5. Conclusions

The main contributions of this dissertation are the formulation of a general framework for predictive induction that puts several specific techniques in perspective, the reformulation of decision tree induction at this general level, and the lifting of this framework to ILP. The result is an efficient, flexible induction method that yields accurate and simple predictive theories.

6. Availability

The full dissertation is available at the following URL:
`http://www.cs.kuleuven.ac.be/~ml/PS/blockeel98:phd.ps.gz`

A prototype of the TILDE system is available for academic purposes upon request. Send mail to `Hendrik.Blockeel@cs.kuleuven.ac.be`.