Global Induction of Classification and Regression Trees

Marek Kretowski
Bialystok University of Technology, Poland

Marcin Czajkowski
Bialystok University of Technology, Poland

INTRODUCTION

Decision trees are one of the most widely used prediction techniques in data mining (Fayyad, Piatetsky-Shapiro, Smyth, & Uthurusamy, 1996). Their similarity to a human reasoning process makes them popular not only among data analysts but also business users. Tree-based approaches are easy to understand, visualize, interpret and their output model can be explained in typical business terms. The ease of application, fast operation and what may be the most important, i.e. the effectiveness of the decision trees, makes them powerful and popular tool (Kotsiantis, 2013).

Two main types of decision trees' approaches can be distinguished by the type of the problem they are applied to. Tree predictors can be used to classify existing data (classification trees) or to approximate real-valued functions (regression trees) (see Figure 1). In each leaf, classification tree assigns a class label (usually the majority class of all instances that reach that particular leaf), while the regression tree holds a constant value (usually an average value for the target attribute). In addition, model tree is an extension of regression tree (see Figure 1). Main difference between regression tree and model tree is that, for the latter, constant value in the terminal node is replaced by the regression function. Each leaf of the model tree holds a linear (or nonlinear) model whose output is the final prediction value.

Most typical tree-based system applies greedy search in decision tree induction. One major drawback of greedy algorithms is that, they search only for locally best splits at each node which does not guarantee the globally best solution. Hence, applications of Evolutionary Algorithms (EAs) to the problem of decision tree induction become increasingly popular alternative. Instead of local search, EAs perform a global search in the space of candidate solutions.

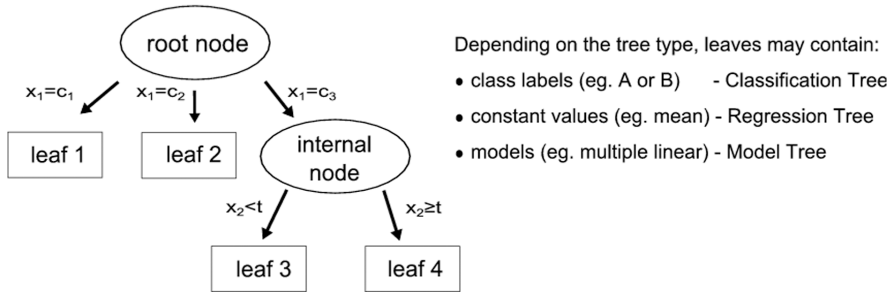
The purpose of this chapter is to illustrate the application of EAs to the problem of decision tree induction. The objectives are to show that evolutionary optimization, compared to the greedy search, may result in finding globally optimal solutions, whose complexity is significantly smaller and the prediction is highly competitive. We will cover the global induction of classification, regression, and model trees.

BACKGROUND

Inducing optimal decision tree is known to be NP-complete (Naumov, 1991). Consequently, practical decision-tree learning algorithms are based on heuristics such as greedy algorithms where locally optimal splits are made in each node. The most popular tree-induction is based on the top-down approach (Rokach & Maimon, 2005). Top-down induction starts from the root node, where locally best split (test) is searched, according to the given optimality measure (e.g. Gini, towing or entropy rule for classification tree and least squared or least absolute deviation error criterion for regression tree). Next, the training instances are redirected to the newly cre-

DOI: 10.4018/978-1-4666-5202-6.ch098

Figure 1. An example of univariate decision tree with tests on nominal and continuous-valued features. Depending on the tree type, leaves could contain class (classification tree), constant value (regression tree) or some kind of model (model tree).



ated nodes and this process is repeated for each node until some stopping-rule is satisfied. The recursive partitioning of the dataset may lead to the data over-fitting, therefore, the decision tree pruning (Esposito, Malerba, & Semeraro, 1997) is applied to improve the generalization power of the predictor.

Most of tree inducing algorithms partition the feature space with axis-parallel hyper-planes. These types of trees are called univariate decision trees. Split at each non-terminal node usually involves single feature. For a continuous-valued feature usually an inequality test with binary outcomes is applied and for a nominal attribute mutually exclusive groups of attribute values are associated with outcomes. One of the first and most well-known solution that can be applied to classification and regression problem is CART (Breiman, Friedman, Olshen, & Stone, 1984) system. Good representatives of univariate inducers are also systems developed by Quinlan: C4.5 (1993) for classification and M5 (1992) for regression.

When more than one feature is taken into account to build a test in non-terminal node, then we deal with multivariate decision trees. The most common form of such a test is an oblique split, which is based on a linear combination of features (hyper-plane). The decision tree which applies only oblique tests is often called oblique or linear, whereas heterogeneous trees with univariate, linear and other multivariate (e.g., instance-based) tests can be called mixed decision trees (Llora & Wilson, 2004). It should be emphasized that computational complexity of the multivariate induction is generally significantly higher than the univariate induction. OC1 (Murthy, Kasif, & Salzberg, 1994) is a good examples of multivariate decision tree system.

Inducing the decision tree with greedy strategy usually leads to suboptimal solutions. One of the alternatives is the ensemble of trees (Seni & Elder, 2010), which is created by the induction of different trees from the training sample. Ensemble classifiers like Random Forests (Breiman, 2001) induce many decision trees whose predictions are combined to make the overall prediction for the forest (a collection of trees). Multiple models improve predictive performance, however, single-tree comprehensibility is lost. Different approaches propose look-ahead algorithms like APDT (Shah & Sastry, 1999) for classification and LLRT (Vogel, Asparouhov, & Scheffer, 2007) for regression.

Evolutionary computing techniques are proven to be effective at escaping local optima and are able to successfully solve a general class of difficult computational problems. They improve handling of attribute interactions when compared to greedy methods. Evolutionary Algorithms (EA) (Michalewicz, 1996) are usually applied in decision tree induction (Barros et al., 2012; Podgorelec et al., 2013), however, in the literature there are some attempts of e.g. ant colony optimization algorithms (Boryczka & Kozak, 2010; Otero et al., 2012).

Related Content