Chapter 50
Evolutionary Algorithms for Global Decision Tree Induction

Marek Kretowski
Bialystok University of Technology, Poland

Marcin Czajkowski
Bialystok University of Technology, Poland

ABSTRACT

Decision trees represent one of the main predictive techniques in knowledge discovery. This chapter describes evolutionary-induced trees, which are emerging alternatives to the greedy top-down solutions. Most typical tree-based systems search only for locally optimal decisions at each node and do not guarantee the optimal solution. Application of evolutionary algorithms to the problem of decision tree induction allows searching for the structure of the tree, tests in internal nodes, and regression functions in the leaves (for model trees) at the same time. As a result, such globally induced decision trees are able to avoid local optima and usually lead to better prediction than the greedy counterparts.

INTRODUCTION

Data mining (Fayyad, Piatetsky-Shapiro, Smyth & Uthurusamy, 1996) can reveal important and insightful information hidden in data. However, appropriate tools and algorithms are required to effectively identify correlations and patterns within the data. Decision trees (Kotsiantis, 2013) represent one of the main techniques for discriminant analysis prediction in knowledge discovery. The success of tree-based approaches can be explained by their ease of application, fast operation, and effectiveness. Furthermore, the hierarchical tree structure, in which appropriate tests from consecutive nodes are sequentially applied, closely resembles a human way of decision making. All this makes decision trees easy to understand, even for inexperienced analysts.

Despite 50 years of research on decision trees, many problems still remain (Loh, 2014) such as searching only for a locally optimal split in the internal nodes; appropriate pruning criterion, efficient analysis of cost-sensitive data or performing multi-objective optimization. To help resolve some of these problems, evolutionary algorithms (EAs) (Michalewicz, 1996) has been applied to decision tree induction

DOI: 10.4018/978-1-5225-7362-3.ch050

(Barros et al., 2012). The strength of this approach lies in the global search for splits and predictions. It results in higher accuracy and smaller output trees compared to popular greedy decision tree inducers.

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

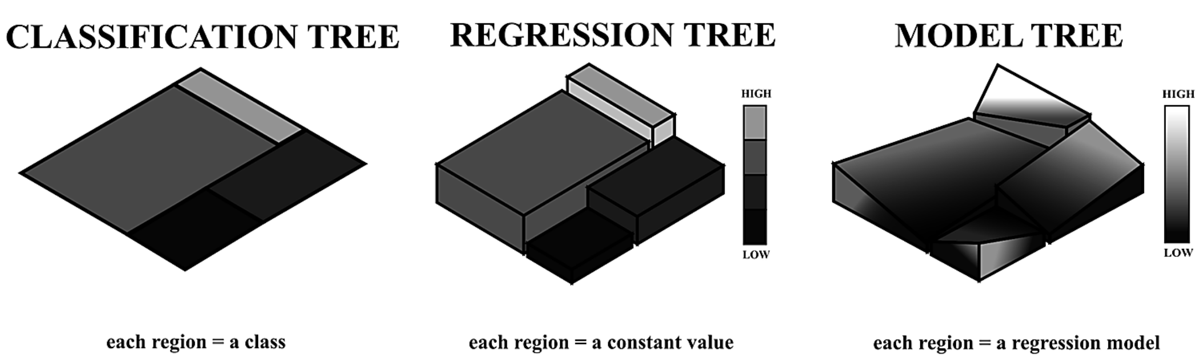
We may find different variants of decision trees in the literature (Loh, 2014). They can be grouped according to the type of problem they are applied to, the way they are induced, or the type of structure. Tree predictors can be used to classify existing data (classification trees) or to approximate 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). A model tree can be seen as an extension of the typical regression tree (Quinlan, 1992). The constant value in each leaf of the regression tree is replaced in the model tree by a linear (or nonlinear) regression function. To predict the target value, the new tested instance is followed down the tree from a root node to a leaf using its attribute values to make routing decisions at each internal node. Next, the predicted value for the new instance is evaluated based on a regression model in the leaf.

Examples of predicted values of classification, regression, and model trees are given in Figure 1.

The gray level color of each region represents a different class label (for a classification tree), and the height corresponds to the value of the prediction function (regression and model trees).

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

Figure 1. An illustration of predicted values of the classification, regression, and model trees



Related Content