Decision Support Systems 74 (2015) 57–66

Contents lists available at ScienceDirect

Decision Support Systems

journal homepage: www.elsevier.com/locate/dss

Cost-sensitive Global Model Trees applied to loan charge-off forecasting

Marcin Czajkowski [a,*], Monika Czerwonka [b], Marek Kretowski [a]

[a] Faculty of Computer Science, Bialystok University of Technology, Wiejska 45a, 15-351 Bialystok, Poland
[b] Collegium of Management and Finance, Warsaw School of Economics, Al. Niepodleglosci 162, 02-554 Warsaw, Poland

ARTICLE INFO

Article history:
Received 17 April 2014
Received in revised form 10 February 2015
Accepted 30 March 2015
Available online 9 April 2015

Keywords:
Cost-sensitive regression
Model trees
Evolutionary algorithms
Asymmetric costs
Loan charge-off forecasting

ABSTRACT

Regression learning methods in real world applications often require cost minimization instead of the reduction of various metrics of prediction errors. Currently in the literature, there is a lack of white box solutions that can deal with forecasting problems where under-prediction and over-prediction errors have different consequences. To fill this gap, we introduced the Cost-sensitive Global Model Tree (*CGMT*), which applies a fitness function that minimizes an average misprediction cost. Proposed specialized genetic operators improve searching for optimal tree structure and cost-sensitive linear regression models in the leaves. Experimental validation is performed on loan charge-off data. It is known to be a difficult forecasting problem for banks due to the asymmetric cost structure. Obtained results show that specialized evolutionary algorithm applied to model tree induction finds significantly more accurate predictions than tested competitors. Decisions generated by the *CGMT* are simple, easy to interpret, and can be applied directly.

1. Introduction

A lot of real world problems are cost-sensitive, which means that different types of prediction errors are not equally costly [5]. As a result, typical minimization of prediction errors is not the best scenario. A cost-sensitive term encompasses all types of learning where cost is considered [50,19], and different types of costs (e.g., costs of attributes, cost of instances, and costs of errors) can be distinguished. The current research focuses on a single cost for decision making, however, multiple costs are also investigated [35].

For example, in medical diagnoses, there are several types of costs that can be minimized, such as the cost of misclassification (e.g., overlooking an ill patient can be fatal in contrast to a false positive test) or the cost of treatment (e.g., financial or risk). When speculating on stock exchange, investors directly compare future gains and losses and usually give more weight to losses. Researchers show that potential gains need to be approximately twice as large to offset potential losses [51]. As a consequence, investors tend to realize their gains more often than their losses as they sell winning stocks more readily. There are many other examples for such asymmetry, such as in bankruptcy prediction [57], behavioral finances [45], expected stock returns [2], criminal justice settings [6], physician prognostic behavior [1], product recommendations [32], and so on.

Cost-sensitive regression is still not adequately addressed in the data mining literature, as most existing research in this area deals with classification problems. Conventional algorithms usually operate with symmetric loss functions and minimize absolute or squared errors that do not distinguish differences between under-prediction and over-prediction, as each is weighted equally (the cost of under-prediction and over-prediction is equal). There is a need for solutions with asymmetric loss functions that can successfully forecast cost-sensitive regression problems. Such models also minimize absolute or squared errors, however, the under-predicted and over-predicted instances have different weights that depend on the costs.

Our study makes several important contributions to the literature. First, we propose a new method called the Cost-sensitive Global Model Tree (*CGMT*), which extends the cost-neutral solution called *GMT* [17]. By applying evolutionary algorithms (*EA*) in the model tree induction, we managed to successfully search for optimal tree structure and cost-sensitive regression models in the leaves under different asymmetric loss functions. What is more, *CGMT* predictions on loan charge-off forecasting data as one of the cost-sensitive problems faced by banks are significantly more accurate than the results of their tested counterparts.

Next, the hierarchical tree structure, in which appropriate tests from consecutive nodes are sequentially applied, closely resembles a human way of decision making. Therefore, the *CGMT* prediction model is natural and easy to understand and interpret, which is extremely important in financial forecasting. Finally, we propose an improvement of existing algorithms in terms of their performance. In solutions proposed in [5] and [56], the linear tuning function is calculated by the heuristic approach (hill climbing algorithm). We managed to find a direct minimization that returns the exact value of an adjusted regression model. The proposed approach significantly extends upon previously performed research on cost-sensitive extensions for *GMT* [16]. In

* Corresponding author.
E-mail address: m.czajkowski@pb.edu.pl (M. Czajkowski).

http://dx.doi.org/10.1016/j.dss.2015.03.009
0167-9236/

particular, proposed solution can work with any convex function. New specialized variants of genetic operators were proposed, the fitness function was improved and mode detailed experimental analysis was performed.

The rest of the paper is organized as follows. The next section presents background information and Section 3 proposes the *CGMT* approach. The experimental evaluation is performed in Section 4 and the paper is concluded in the last section, in which future work is also discussed.

2. Background

In this section, we want to present some background information about decision trees for regression and asymmetric loss function as well as the problem of loan charge-off forecasting.

2.1. Regression and model trees

The most common predictive tasks in data mining are classification and regression, and the decision tree [42] is one of the most frequently applied prediction techniques. Tree-based approaches are easy to understand, visualize, and interpret. Their similarity to the human reasoning process makes them a powerful tool [29] among data analysts.

The problem of learning the optimal model tree is known to be NP-complete [39]. Consequently, practical decision-tree inducers are based on heuristics such as the greedy approach, where locally optimal decisions are made in each node. This process is known as recursive partitioning [41]. Two main variants of decision trees can be distinguished by the type of problem they are applied to. Tree predictors can be used to classify existing data (classification trees) or to approximate real-valued functions (regression trees). In each leaf, the classification tree assigns a class label (usually a 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). The model tree can be seen as an extension of the regression tree. The most important difference is that the constant value in each leaf of the regression tree is replaced in the model tree by the linear (or nonlinear) regression function. An example of classification, regression, and model tree induced by the top-down greedy approach is illustrated in Fig. 1. The color of each region in the classification tree represents a different class. In regression and model trees, the height of each region corresponds to the value of the prediction function.

One of the first and probably the most well-known top-down regression tree solutions is the *CART* system [7]. The algorithm searches for a locally optimal split that minimizes the sum of squared residuals and builds a piecewise constant model with each terminal node fitted with the training sample mean. Other solutions have managed to improve the prediction accuracy by replacing single values in the leaves with more advanced models. The *M*5 system [52] induces a tree that contains multiple linear models in the leaves, and thus the tree is similar to a piecewise linear function. All these solutions are fast and generally efficient in many practical problems, but they usually produce locally optimal solutions.

To limit the negative effects of greedy induction, multiple authors have proposed various techniques [36,54]. However, the true global approach for decision tree induction was possible with evolutionary computation. In the literature, there are attempts to apply the evolutionary approach for the induction of decision trees [4], but only a few solutions concern the regression problem. In *TARGET* [20], the authors propose to evolve a *CART*-like regression tree with simple operators and the Bayesian Information Criterion (*BIC*) [43] as a fitness function. Evolutionary induced regression trees with linear models in the leaves were proposed in a solution called *E-Motion* [3]. The authors applied a standard 1-point crossover and two different mutation strategies (shrinking and expanding). The algorithm optimizes the tree error

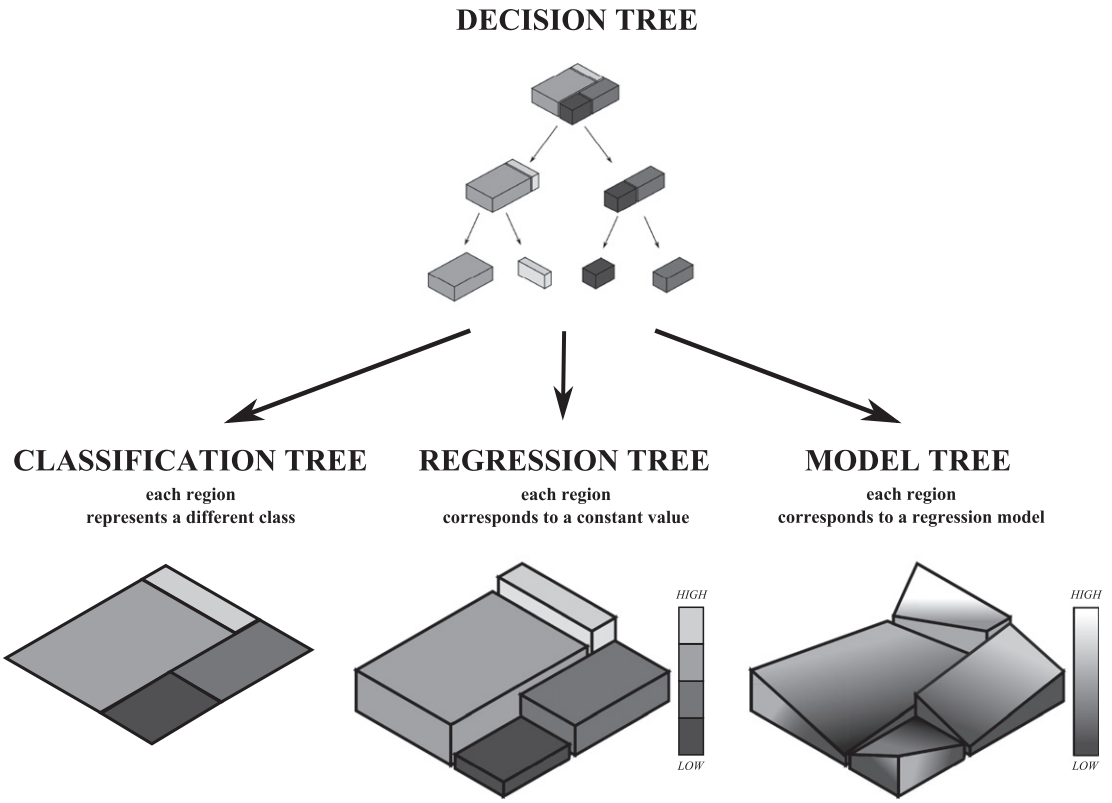

Fig. 1. An example of top-down induction of classification, regression, and model tree.