IDEA GROUP PUBLISHING
701 E Chocolate Avenue, Hershey PA17033, USA
Tel: 717/533-8845; Fax: 717/533-8661; URL: www.idea-group.com
ITB7252

Chapter III

Cost-Sensitive Classification Using Decision Trees, Boosting and MetaCost

Kai Ming Ting
Monash University, Australia

This chapter reports results obtained from a series of studies on cost-sensitive classification using decision trees, boosting algorithms, and MetaCost which is a recently proposed procedure that converts an error-based algorithm into a cost-sensitive algorithm. The studies give rise to new variants of algorithms designed for cost-sensitive classification, and provide insights into the strength and weaknesses of the algorithms. First, we describe a simple and effective heuristic of converting an error-based decision tree algorithm into a cost-sensitive one via instance weighting. The cost-sensitive version performs better than the error-based version that employs a minimum expected cost criterion during classification. Second, we report results from a study on four variants of cost-sensitive boosting algorithms. We find that boosting can be simplified for cost-sensitive classification. A new variant which excludes a factor used in ordinary boosting has an advantage of producing smaller trees and different trees for different scenarios; while it performs comparably to ordinary boosting in terms of cost. We find that the minimum expected cost criterion is the major contributor to the improvement of all cost-sensitive adaptations of ordinary boosting. Third, we reveal a limitation of MetaCost. We find that MetaCost retains only part of the performance of the internal classifier on which it relies. This occurs for both boosting and bagging as its internal classifier.

INTRODUCTION

Cost-sensitive classification allows one to assign different costs to different types of misclassifications. For example, in the field of natural disaster prediction, the cost of having a disaster undetected is much higher than the cost of having a false alarm. Thus, cost-sensitive classification seeks to minimize the total misclassification cost. In contrast, conventional classification seeks to minimize the total errors regardless of cost.

Cost-sensitive tree induction employing the greedy divide-and-conquer algorithm has attracted much interest recently. Breiman, Friedman, Olshen and Stone (1984) describe two different methods of incorporating variable misclassification costs into the process of tree induction. These methods adapt the test selection criterion in the tree growing process. Pazzani, Merz, Murphy, Ali, Hume and Brunk (1994) reported negative empirical results when using one of Breiman et al.'s formulations to induce cost-sensitive trees. They found that the cost-sensitive trees do not always have lower misclassification costs, when presented with unseen test data, than those trees induced without cost consideration. Using a post-processing approach, Webb (1996) shows that applying a cost-sensitive specialization technique to a minimum error tree can reduce its misclassification costs by a small margin. In contrast to Pazzani et al.'s study, Ting (in press) shows convincingly that, by applying a simple heuristic, a truly cost-sensitive tree can be effectively learned directly from the training data, employing the greedy divide-and-conquer algorithm. The paper extends this line of research into improving the performance by combining multiple trees.

Boosting has been shown to be an effective method of combining multiple models in order to enhance the predictive accuracy of a single model (Quinlan, 1996; Freund & Schapire, 1996; Bauer & Kohavi, 1999). Boosting is amenable to cost-sensitive adaptation and recent research has reported some success (Ting & Zheng, 1998; Fan, Stolfo, Zhang & Chan, 1999). However, the relative performance between the proposed methods has yet to be investigated, and other forms of adaptations are also possible.

In this paper, we study two new variants of cost-sensitive boosting, and two recently proposed variants by Fan et al (1999) and Ting and Zheng (1998). All these variants must relearn their models when misclassification cost changes. An alternative method that converts an error-based model to a cost-sensitive model simply applies a minimum expected cost criterion (Michie, Spiegelhalter & Taylor, 1994) to the error-based model, and the same model can be reused when cost changes. Therefore, it is important to investigate whether the cost-sensitive variants have any advantage over this simple alternative. This study aims at improving our understanding of the behavior of the four cost-sensitive boosting algorithms and how variations in the boosting procedure affect misclassification cost.

MetaCost (Domingos, 1999) is a recently proposed method for making an arbitrary classifier cost-sensitive. The method has an interesting design that uses a "meta-learning" procedure to relabel the classes of the training examples and then

Related Content

From Silos to Sharing: An Institutional Research View of the Conversion to an ERP
Dana L. Dalton (2012). *Cases on Institutional Research Systems (pp. 39-51).*
www.irma-international.org/chapter/silos-sharing-institutional-research-view/60839

Applications of Domain-Specific Predictive Analytics Applied to Big Data
Ravi Kumar Poluru, S. Bharath Bhushan, Basha Syed Muzamil, Praveen Kumar Rayaniand Praveen Kumar Reddy (2019). *Sentiment Analysis and Knowledge Discovery in Contemporary Business (pp. 289-306).*
www.irma-international.org/chapter/applications-of-domain-specific-predictive-analytics-applied-to-big-data/210976

Intellectual Capital Measurement and Reporting Models
Jamal A. Nazari (2015). *Knowledge Management for Competitive Advantage During Economic Crisis (pp. 117-139).*
www.irma-international.org/chapter/intellectual-capital-measurement-and-reporting-models/117846

A Comparison of Revision Schemes for Cleaning Labeling Noise
Chuck P. Lamand David G. Stork (2008). *Mathematical Methods for Knowledge Discovery and Data Mining (pp. 220-232).*
www.irma-international.org/chapter/comparison-revision-schemes-cleaning-labeling/26142

Internet Forums: What Knowledge can be Mined from Online Discussions
Mikolaj Morzy (2011). *Knowledge Discovery Practices and Emerging Applications of Data Mining: Trends and New Domains (pp. 315-336).*
www.irma-international.org/chapter/internet-forums-knowledge-can-mined/46902