Abstract

Classification is a data mining technique used to predict group membership for data instances. Cost-sensitive classifier is a relatively new field of research in the data mining and machine learning communities. They are basically used for classification tasks under the cost-based model, unlike the error-based model. Error based classifier AdaBoost is a simple algorithm that reweights the training instances to build multiple classifiers in training phase, without considering the cost of misclassification. Out of all generated classifiers in training, in classification, it collects the weighted votes from each and classifies the new sample (example) according to maximum votes collected. Intuitively, combining multiple models shall give more robust predictions than a single model under the situation where misclassification costs are considered. Boosting has been shown to be an effective method of combining multiple models in order to enhance the predictive accuracy of a single model. Thus, it is natural to think that boosting might also reduce the misclassification costs. All the cost-sensitive boosters are studied and five new extensions are proposed and their results are compared in this paper. A few future extensions are notified.
- Data-set downloaded from: http://tunedit.org/repo/UCI/credit-a.arff
- Tutorial on Oracle® Data Mining Concepts 11g Release 1, http://download.oracle.com/docs/cd/B28359_01/datamine.111/b28129/classify.htm
- OAIDTB: Boosting extensions for WEKA, web-link: http://pisuerga.inf.ubu.es/lsi/Software/oaiddtb/
- Books: Data mining - Concept and Techniques by Han & Kamber. Data mining: concepts and techniques. The Morgan Kaufmann series in data management systems, ISBN-
Index Terms

Computer Science

Keywords

Adaboost  Cost-sensitive Classifiers  Csextension1  Csextension2  Csextension3  Csextension4  Csextension5  Misclassification Cost

Number Of High Cost Errors