Poly-Bagging Predictors for Classification Trees

Francisco Louzada
Universidade de São Paulo - ICMC-USP

Binary classification modelling compromises a common task on data mining, directing to discovering prediction rules which aid in planning and decision making. In many classification scenarios, such as financial, medical or industrial ones, many techniques are largely considered, such as classification trees , discriminant analysis, probit analysis, logistic regression and neural networks. We consider here classification trees. A critical factor is whether a classification modelling is accurate enough in order to provide correct classification of the client as a good or bad payer. In this context the concept of bootstraping aggregating (bagging) arises. The basic idea is to generate multiple classifiers by obtaining the predicted values from the fitted models to several replicated datasets and then combining them into a single predictive classification in order to improve the classification accuracy. In this paper we purpose a new bagging-type variant procedure, which we call poly-bagging, consisting of combining predictors over a succession of resamplings. The study is driven by credit scoring modelling. The proposed poly-bagging procedure was applied to some different artificial datasets and to a real granting of credit dataset up to three succession of resamplings. We observed better classification accuracy for the two-bagged and the three-bagged models for all considered setups. These results lead to an strong indication that the poly-bagging approach may promote improvement on the modelling performance measures, while keeping a flexible and straightforward bagging-type structure easy to implement.