C

Composite Classifiers for Bankruptcy Prediction

Efstathios Kirkos
ATEI of Thessaloniki, Greece

INTRODUCTION

Business failures can cause financial damages to investors, creditors, or even society. For this reason bankruptcy prediction is one of the most challenging tasks in the field of financial decision-making. Business failure prediction has been an active research area since the 60s. The work of Beaver (1966) who performed univariate analysis of financial ratios and the work of Altman (1968) who employed Multiple Discriminant Analysis (MDA) mark the starting point of the relevant research. In the following years many researchers have proposed statistical and artificial intelligence techniques to predict bankruptcies. In most cases the artificial intelligence techniques outperformed the statistical techniques.

In the current research one can notice that there is a strong trend concerning the proposed classification methods. This trend is to design and apply composite classifiers, i.e. hybrid and ensemble classifiers. According to the results, composite classifiers perform better than the single classification techniques. The purpose of the present chapter is to cover the topic of the design and application of composite classifiers for bankruptcy prediction. Key issues related to composite classifiers are analyzed to familiarize the reader with these techniques. Subsequently nine selected papers that were recently published in high reputation journals of an impact factor score of more than 0.5, are presented. This presentation concentrates on problems, methodological issues, proposed designs and findings that refer to the employment of composite classifiers for bankruptcy prediction. Other important issues relevant to bankruptcy prediction, such as improved single classifiers, alternative input data, feature selection etc are beyond the scope of the present article.

BACKGROUND

Recent developments in the field of intelligent bankruptcy prediction have pointed out that the combination of a number of techniques may yield models able to outperform individual classifiers. The combination of individual techniques is not a trivial task. Several methodological issues give rise to alternative approaches and may lead to different categorizations. A common categorization of composite classifiers divides them into ensemble and hybrid ones. In what follows, a brief introduction to these types of classifiers is presented.

According to the ensemble approach a number of different classifiers, each of which solves the same original problem, are trained. The individual decisions are aggregated and a final classification decision is reached.

Since it is pointless to multiply the same original classifier, the individual base classifiers must substantially differ. This diversification can be achieved in a number of ways.

- Employment of different methods and development of corresponding models. Neural Networks, Decision Trees and Bayesian Networks are examples of individual methods. Normally, all models are trained by using the same samples.
- Employment of different training sets. The idea is to create alternative data sets from

DOI: 10.4018/978-1-4666-5202-6.ch043

an original data set and to train corresponding models. Bagging (Breiman, 1996) is a common example of this case.

- Employment of different subsets of features. Different models are trained by using different input variables. This can be achieved by using different feature selection techniques.
- Different initial settings of the base method. The same technique and the same data are used. The difference arises from the tuning of the base method. For a neural network for example this may mean different topologies, learning rates, training epochs etc.

Another important aspect concerns the way each classifier affects the other classifiers. The classifiers can be combined in a sequential or concurrent manner. The sequential approach is an iterative process. Knowledge obtained in one iteration influences the learning task of the next iteration. The obtained knowledge can be expressed as manipulation of the training data set. Another approach is to use the classifier developed in one iteration to build the classifier of the following iteration. A well known example of sequential combination with data set manipulation is boosting. In AdaBoost (Freund & Schapire, 1996), weights are assigned to each training sample. After the training of a classifier, the weights of the misclassified samples are increased. In the next training iteration the new classifier pays more attention to the previously misclassified samples.

According to the concurrent ensemble method, several alternative training sets are created from the original data set and for each training set a classifier is developed. These classifiers are aggregated to the final classifier. Bagging is a famous concurrent ensemble algorithm. In Bagging the training subsets are created with random sampling with replacement. In order to classify an observation, the individual classifiers vote and the observation is assigned to the class that concentrates the majority of votes.

A key issue in building ensembles is to define how the individual classification decisions are combined. There are two available approaches, simple combining methods and meta-combining methods. Simple combining methods aggregate, according to a formula, the individual decisions. Uniform voting is the simplest scheme in this case. Many other alternative schemes such as the Distribution Summation, the Borda count and the Dempster-Shafer Theory (DST) have also been proposed. The Distribution Summation represents the sum of membership probabilities obtained from each classifier. According to the Borda count, the individual classifiers rank the candidate class values in order of preference. A number of points correspond to each position in the rank and the class value that accumulates the most points is the winner. In the DST the winning class maximizes the value of a belief function that uses the basic probability assignment. Meta-combiners, that constitute the second possible approach, use the base classifiers and their predictions for further learning. In Stacked Generalization (Wolpert, 1992) the outputs of the base classifiers are passed to classifiers of the successive layer and are used by them as input.

In hybrid systems, as in ensembles, a number of techniques are used. However, there are significant differences with the ensemble methods. The first is that heterogeneous techniques, which solve different problems, are involved. Another difference is that only one classifier performs the final classification, whereas in ensemble methods the decisions of several classifiers are aggregated.

Lin, Hu and Tsai (2012) defined three types of hybrid classifiers, cascading different classifiers, combining clustering and classification and finally, intergrading two techniques in a way that one technique complements the other. In the cascaded hybrid classifiers, the output of the first level method is passed to the second level classifier. An example is the case in which calculated values obtained from the first level are passed as additional input to the final classifier. A broad interpretation of the notion of hybrid techniques