HYBRID PROBABILITY-BASED ENSEMBLES FOR BANKRUPTCY PREDICTION

Chihli Hung[1], Jing-Hong Chen[2], Stefan Wermter[3],
[1,2]Department of Management Information Systems, Chung Yuan Christian University, Taiwan ROC
[3]School of Computing and Technology, University of Sunderland, UK
chihli@cycu.edu.tw

ABSTRACT

Bankruptcy prediction has attracted a lot of research interests as it is one of the major business topics. Both statistical approaches such as discriminant analysis, logit and probit models and computational intelligence techniques such as expert systems, artificial neural networks and support vector machines have been explored for this topic and most research compares the prediction performance via different techniques for a specific data set. However, there is no consistent result that one technique is consistently better than another. Different techniques have different advantages on different data sets and different feature selection approaches. Therefore, we divide the prediction performance into using different techniques for two parts: bankruptcy prediction and non-bankruptcy prediction. Based on analyzing the expected probability of both bankruptcy and non-bankruptcy predictions for a training set, we have built an ensemble of three well known classification techniques, i.e. the decision tree, the back propagation neural network and the support vector machine. This ensemble provides an approach which inherits advantages and avoids disadvantages of different classification techniques. In this paper we describe results which demonstrate that our expected probability-based ensemble outperforms other stacking ensembles based on a weighting or voting strategy.

Keyword: Bankruptcy Prediction, Ensemble Learning, Decision Tree, Neural Network, Support Vector Machine

INTRODUCTION

Bankruptcy prediction is an important and serious topic for business. An effective prediction in time is valued priceless for business in order to evaluate risks or prevent bankruptcy. A fair amount of research has therefore focused on bankruptcy prediction. Generally speaking, there are two main groups of techniques for handling this topic. The first group is statistical techniques, such as regression analysis, correlation analysis, discriminant analysis, logit model, probit model etc. The second group belongs to computational intelligence techniques such as decision trees, artificial neural networks (ANN), support vector machines (SVM) etc. Most researchers use one of techniques to compare the prediction performance with other techniques for a specific data set [e.g. Odom and Sharda, 1990; Altman et al., 1994; Jo et al., 1997; Koh and Tan, 1999; Shin et al., 2005; Lensberg et al., 2006]. However, there is no single conclusion that one technique is consistently better than another for general bankruptcy prediction.

In terms of prediction accuracy, there are two prediction directions: bankruptcy and non-bankruptcy. Some techniques have a bias for bankruptcy prediction and other techniques have a bias for non-bankruptcy prediction. In other words, techniques having a bias for bankruptcy prediction have a higher probability to predict a business as bankruptcy. Thus, it is more credible when they predict a business as non-bankruptcy. This probability is an expected probability, which provides a hint to make an ensemble of classifiers. However, no work in the literature refines the prediction performance from the viewpoint of analyses of these two directions.

In this research, we compare several models such as decision trees, ANNs and SVMs by analyzing expected probabilities from both bankruptcy and non-bankruptcy predictions. We find that each classifier has its own benefits for a specific prediction direction. Our aim is to provide an approach which inherits advantages and avoids disadvantages of different classification techniques for bankruptcy prediction, and want to examine how the prediction accuracy of the whole model can be enhanced. Thus we propose an expected probability-based ensemble model for bankruptcy prediction. We will demonstrate, according to our experimental results, that our expected probability-based strategy outperforms the traditional majority voting and weighted voting strategies.

EXPECTED PROBABILITY

For a bankruptcy prediction task, each classifier has two prediction directions: one for bankruptcy and the other for non-bankruptcy. We usually evaluate such a classifier based on the classification accuracy (CA), which is generally presented by a confusion matrix [Kohavi and Foster, 1998] as shown in Table 1. “A” indicates the number of non-bankrupt companies which are also predicted as non-bankrupt companies. “B” indicates the number of non-bankrupt companies which are predicted as bankrupt companies. “C” indicates the number of bankrupt companies which are predicted as non-bankrupt companies. “D” indicates the number of bankrupt companies which are also predicted as bankrupt companies. Therefore, the classification accuracy of a classifier, CA_{model}, is defined as (1). The classification accuracy for bankrupt companies, $CA_{bankruptcy}$, is defined as (2). The classification accuracy for non-bankrupt companies, $CA_{non\text{-}bankruptcy}$, is defined as (3). We propose expected probabilities for bankruptcy and non-bankruptcy predictions as (4) and (5) respectively.

Table 1. Confusion Matrix

		Predictive value	
		Positive (non-bankruptcy)	Negative (bankruptcy)
Actual value	Positive (non-bankruptcy)	A	B
	Negative (bankruptcy)	C	D

$$CA_{\text{model}} = \frac{A+D}{A+B+C+D} \quad (1)$$

$$CA_{\text{bankruptcy}} = \frac{D}{C+D} \quad (2)$$

$$CA_{\text{non-bankruptcy}} = \frac{A}{A+B} \tag{3}$$

$$P_{\text{bankruptcy}} = \frac{D}{B+D} \tag{4}$$

$$P_{\text{non-bankruptcy}} = \frac{A}{A+C} \tag{5}$$

THE PROPOSED EXPECTED PROBABILITY-BASED ENSEMBLE

Generally speaking, the stacking strategy of ensemble learning is able to inherit advantages from the different classifiers, but also suffers from disadvantages of those classifiers simultaneously [Witten and Frank, 2005]. In our research, based on the concept of expected probability, we propose a novel ensemble approach, which overcomes the traditional dilemma for integrating different classifiers. By analyzing the decision bias of each classifier, we arrange the sequence of classifiers in an ensemble decision tree. For example, in Fig. 1, classifier 1 has the maximum expected probability for non-bankruptcy and this classifier is the top node of the ensemble tree. Thus, when a sample in the test set is classified by classifier 1 as a non-bankrupt company, we believe that this decision has a high expected probability or otherwise this sample is introduced to classifier 2 for further classification and etc.

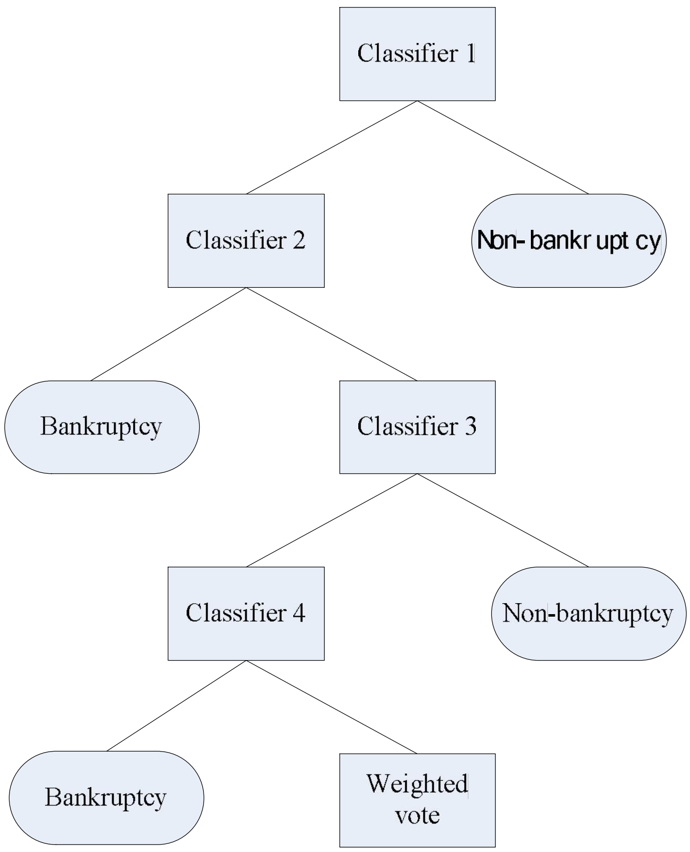

Fig. 1. An example of expected probability based ensemble tree. A rectangle is a classifier and an ellipse is decision leaf node.

EXPERIMENTAL DESIGN

We use the Wieslaw data set (Wieslaw, 2004) which includes 30 financial ratios from 56 bankrupt companies and 64 non-bankrupt companies between 1997 and 2001. Each company contains two continuous yearly data so there are 240 samples in total. The technique of feature selection is often an essential method for filtering the noise of the data set. We remove any variable which has less significant discrimination ability by using the significance test based on a significance level of 5% of F-test and finally only fourteen variables are left. This feature selection technique is very common and well established [e.g. Serrano-Cinca, 1996].

Then, the data set is randomly divided into two parts: a training set with 200 samples and a test set with 40 samples. The proportion of bankrupt and non-bankrupt companies of each set equals to that of the entire data set. In order to avoid the problem of overfitting during training for each classifier, we use the 10-fold cross validation technique. There are many decision tree algorithms. We use J4.8, which is a modification of C4.5 revision 8 (Witten and Frank, 2005). For the technique of SVM, we use sequential minimal optimization (SMO) algorithm because of its quick convergence (Platt, 1999). For ANN, we use back propagation neural networks (BPN) with 0.2 of a momentum and the number of units in its input, hidden and output layers is 14, 10 and 1 respectively. The initial learning rate is 0.75 and decays over the training time.

EXPERIMENTAL RESULTS

We first calculate all expected probabilities for J4.8, SVM and BPN as in Table 2. The maximum expected probability is in the direction of non-bankruptcy prediction for J4.8 and thus it is the first branch node of the ensemble tree. If samples in the test set are filtered by J4.8 as non-bankrupt companies, they are final results due to the maximum expected probability. Otherwise, the rest samples should be filtered by classifiers in other child branches of the ensemble tree. At the next cycle, we find the maximum expected probability only from bankrupt companies. Thus, BPN goes for the next and SVM is the last. We compare our proposed ensemble with majority and weighted voting strategies and find that our model outperforms those models using only majority and weighted voting strategies by getting a higher classification accuracy as shown in Table 3.

Table 2. Expected Probabilities of Three Classifiers

	J4.8	BPN	SVM
$P_{\text{bankruptcy}}$	64.29%	70.21%	70.11%
$P_{\text{non-bankruptcy}}$	76.14%	74.53%	71.68%

Table 3. A Comparison of the Expected Probability-Based Ensemble and the Ensemble Using Voting Strategy for tset-set

	Voting ensemble	Weighting ensemble	Expected probability ensemble
CA_{model}	70.00%	70.00%	72.50%
$CA_{bankruptcy}$	73.68%	73.68%	73.68%
$CA_{non\text{-}bankruptcy}$	66.67%	66.67%	71.43%

CONCLUSIONS

Bankruptcy prediction is an important and serious topic for both researchers in data mining and decision-makers in business. Most researchers use a single technique for bankruptcy prediction and compare this with another one. However, there is no conclusion for the best bankruptcy prediction technique. In this research, we propose an expected probability-based ensemble for bankruptcy prediction in order to inherit advantages and avoid disadvantages of different classifiers. Based on our experimental results, the proposed ensemble outperforms other stacking ensemble using the weighting or voting strategy.

REFERENCES

Altman, E.I. 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy, *Journal of Finance*, 23 (4), pp. 589-609.

Altman, E.I., Marco, G.V. & Varetto, F. 1994. Corporate distress diagnosis: comparisons using linear discriminant analysis and neural networks, *Journal of Banking and Finance*, 18, pp. 505-529.

Balcaen, S. & Ooghe, H. 2004. 35 years of studies on business failure: an overview of the classical statistical methodologies and their related problems, *Working paper*, No. 248, Department of Accountancy and Corporate Finance, Ghent University, Belgium.

Bauer, E. & Kohavi, R. 1999. An empirical comparison of voting classification algorithms: bagging, boosting, and variants, *Machine Learning*, 36 (1-2), pp. 105-139.

Beaver, W. 1966. Financial ratios as predictors of failure, *Journal of Accounting Research*, 4, pp. 71-111.

Eisenbeis, R.A. 1977. Pitfalls in the application of discriminant analysis in business, finance and economics, *Journal of Finance*, pp. 875-900.

Jo, H., Han, I. & Lee, H. 1997. Bankruptcy prediction using case-based reasoning, neural networks, and discriminant analysis, *Expert Systems with Applications*, 13 (2), pp. 97-108.

Koh, H.C. & Tan S.S. 1999. A neural network approach to the prediction of going concern status, *Accounting and Business Research*, 29 (3), pp. 211-216.

Kohavi, R. & Foster, P. 1998. Glossary of terms, *Machine Learning*, 30 (23), pp. 271-274.

Lensberg, T., Eilifsen, A. & McKee, T.E. 2006. Bankruptcy theory development and classification via genetic programming, *European Journal of Operational Research*, 169, pp. 677-697.

Min, J.H. & Lee, Y.-C. 2005. Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters, *Expert Systems with Applications*, 28 (4), pp. 603-614.

Myer, P.A. & Pifer, H.W. 1970. Prediction of bank failure, *Journal of Finance*, pp. 853-868.

Odom, M. & Sharda, R. 1990. A neural network model for bankruptcy prediction, *Proceedings of International Joint Conference on Neural Networks*, pp. 163-168.

Ohlson, J. 1980. Financial ratios and the probabilistic prediction of bankruptcy, *Journal of Accounting Research*, 18 (1), pp. 109-131.

Platt, J. 1999. Fast training of support vector machines using sequential minimal optimization, Schoelkopf, B., Burges, C.J.C., Smola, A.J. (eds.), *Advances in Kernel Methods - Support Vector Learning*. Cambridge, MA: MIT Press, pp. 185-208.

Press, S.J. & Willson, S. 1978. Choosing between logistic regression and discriminant analysis, *Journal of the American Statistical Association*, pp. 699-705.

Serrano-Cinca, C. 1996. Self organizing neural networks for financial diagnosis, *Decision Support Systems*, 17 (3), pp. 227-238.

Shin, K.-S., Lee, T.S. & Kim, H.-J. 2005. An application of support vector machines in bankruptcy prediction model, *Expert Systems with Applications*, 28 (1), pp. 127-135.

Tam, K. & Kiang, M. 1992. Managerial applications of neural networks: the case of bank, *Management Science*, 38 (7), pp. 926-947.

West, D., Dellana, S & Qian, J. 2005. Neural network ensemble strategies for financial decision applications, *Computers & Operations Research*, 32, pp. 2543-2559.

Wieslaw, P. 2004. Application of Discrete Predicting Structures in an Early Warning Expert System for Financial Distress, *PhD Thesis*, Szczecin Technical University, Szczecin.

Witten, I.H. & Frank, E. 2005. *Data Mining*, 2nd ed, Elsevier, Morgan Kaufmann Publishers.

Zmijewski, M. 1984. Methodological issues related to the estimation of financial distress prediction models, *Journal of Accounting Research*, 22 (1), pp. 59-82.