Chapter 7

Bankruptcy Prediction by Supervised Machine Learning Techniques:
A Comparative Study

Chih-Fong Tsai
National Central University, Taiwan

Yu-Hsin Lu
National Chung Cheng University, Taiwan

Yu-Feng Hsu
National Sun Yat-Sen University, Taiwan

ABSTRACT

It is very important for financial institutions which are capable of accurately predicting business failure. In literature, numbers of bankruptcy prediction models have been developed based on statistical and machine learning techniques. In particular, many machine learning techniques, such as neural networks, decision trees, etc. have shown better prediction performances than statistical ones. However, advanced machine learning techniques, such as classifier ensembles and stacked generalization have not been fully examined and compared in terms of their bankruptcy prediction performances. The aim of this chapter is to compare two different machine learning techniques, one statistical approach, two types of classifier ensembles, and three stacked generalization classifiers over three related datasets. The experimental results show that classifier ensembles by weighted voting perform the best in term of predication accuracy. On the other hand, for Type II errors on average stacked generalization and single classifiers perform better than classifier ensembles.

INTRODUCTION

Bankruptcy prediction has been a major research topic in accounting and finance for at least a century since corporate bankruptcy can affect the economy of every country seriously. Therefore, timely and correctly predicting bankruptcy is a great importance to various stakeholders (e.g. management, investors, employees, shareholders

and other interested parties) as it provides them some early warnings (Shin, et al, 2005; Lensberg, et al, 2006; Van Gestel et al., 2006; Hua et al. 2007).

Financial failure always occurs when the firm has chronic and serious losses, owns negative net worth that the market value of assets is less than total liabilities, and/or in a situation which is firm's inability to pay debts as they come due. The common assumption underlying bankruptcy prediction is that a firm's financial statements appropriately reflect all these characteristics. Therefore, almost all prior researches (such as Deakin, 1972; Ohlson, 1980; Richardson, et al., 1998; Van Gestel et al. 2006; Hua et al., 2007; Alfaro et al., 2008) have predicted financial distress through several classification techniques by using financial ratios (e.g. leverage, size of firm, and current liquidity) and data originating from these statements.

However, traditional statistics such as univariate approaches (Beaver, 1966), multivariate approaches, linear multiple discriminant approaches (MDA) (Altman, 1968; Altman, et al., 1977), and multiple regression (Meyer & Pifer, 1970) typically rely on the linearity assumption, as well as normality assumptions which is difficult to apply to the real world problem. To develop a more accurate and generally applicable prediction model, machine learning and artificial intelligence techniques including neural networks, decision trees, genetic algorithm (GA), support vector machine (SVM), etc., have been successfully applied in corporate financial bankruptcy forecasting recently (Wu et al., 2007; Hua et al., 2007; Huang et al., 2008; Alfaro et al., 2008). Especially, the neural network models trained by the back-propagation learning algorithm and decision trees are the popular techniques used for financial and accounting literatures.

The consideration of prior studies has been to identify the single best model for predicting financial distress. However, many researches have realized that there exists limitation on using a single classification technique. This observation has motivated the relatively recent studies utiliz-

ing classifiers combinations (i.e. Multi-classifier system or ensembles) for better accuracy (Zhou & Zhang, 2002; Kim, et al., 2002; West, et al., 2005; Tsai & Wu, 2008; Nanni & Lumini, 2009). Besides classifier ensembles, stacked generalization is another advanced learning approach which estimates the errors of using one single technique and then corrects those errors to maximize the accuracy (Wolpert, 1992; Tsai, 2003).

Although the two approaches may provide more accurate prediction results in various domains, there are very few researches comparing with different models based on these machine learning techniques to examine their prediction performances. Therefore, this paper develops a classifier ensemble and stacked generalization model, respectively and employs a multilayer perception neural network, decision trees, and logistic regression as the baseline classifiers to assess the accuracy and Type I/II errors of these models.

The remainder of this paper is organized as follows. Section 2 reviews the predictors of bankruptcy and describes the concept of pattern classification and different classification techniques. Section 3 describes the development of different models and the evaluation methods of them. In Section 4, the experiments are based on comparing the prediction performance of models in terms of average prediction accuracy and the type I and type II errors over three well-known and public downloadable datasets. Finally, the conclusion is made in Section 5.

LITERATURE REVIEW

Bankruptcy Prediction

Factors of Bankruptcy

According to the definition of Beaver (1966), financial failure is defined as the inability of a firm to pay its financial obligations as they mature. In

14 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/bankruptcy-prediction-supervised-machine-learning/46808

Related Content

Is It Privacy or Is It Access Control?
Sylvia L. Osborn (2014). *Information Security in Diverse Computing Environments (pp. 51-58).*
www.irma-international.org/chapter/is-it-privacy-or-is-it-access-control/114369

How Private Is Your Financial Data?: Survey of Authentication Methods in Web and Mobile Banking
Vidya Mulukutla, Manish Guptaand H. R. Rao (2017). *Identity Theft: Breakthroughs in Research and Practice (pp. 327-357).*
www.irma-international.org/chapter/how-private-is-your-financial-data/167234

Data Hiding Method Based on Inter-Block Difference in Eight Queens Solutions and LSB Substitution
Vinay Kumar, Abhishek Bansaland Sunil Kumar Muttoo (2014). *International Journal of Information Security and Privacy (pp. 55-68).*
www.irma-international.org/article/data-hiding-method-based-on-inter-block-difference-in-eight-queens-solutions-and-lsb-substitution/130655

Cybercafés of Nepal: Passage to Cyber Crime?
Deepak Rauniar (2008). *Security and Software for Cybercafes (pp. 253-269).*
www.irma-international.org/chapter/cybercafés-nepal-passage-cyber-crime/28541

Robust Image Data Hiding Technique for Copyright Protection
Siddharth Singhand Tanveer J. Siddiqui (2013). *International Journal of Information Security and Privacy (pp. 44-56).*
www.irma-international.org/article/robust-image-data-hiding-technique-for-copyright-protection/87414