Expert Systems with Applications 36 (2009) 4830–4837

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Failure prediction of dotcom companies using hybrid intelligent techniques

D. Karthik Chandra [a], V. Ravi [a,*], I. Bose [b]

[a] Institute for Development and Research in Banking Technology, Castle Hills Road #1, Masab Tank, Hyderabad 500 057, AP, India
[b] School of Business, The University of Hong Kong, Room 730 Meng Wah Complex Pokfulam Road, Hong Kong SAR, PR China

ARTICLE INFO

Keywords:
Dotcom companies
Failure prediction
Feature selection
Majority voting
t-statistic
Ensemble
Boosting

ABSTRACT

This paper presents a novel hybrid intelligent system in the framework of soft computing to predict the failure of dotcom companies. The hybrid intelligent system comprises the techniques such as a Multilayer Perceptrons (MLP), Random Forest (RF), Logistic Regression (LR), Support Vector Machine (SVM), Classification and Regression Trees (CART). The dataset collected from Wharton Research Data Services (WRDS) consists of 240 dotcom companies (also known as click-and-mortar companies), of which 120 are failed and 120 are healthy. Ten-fold cross validation is performed on the data set for all the techniques considered in their stand-alone mode. Further, two hybrid techniques viz., ensembling and boosting are employed to improve the accuracies. Moreover, *t*-statistic is performed on the dataset for feature selection purpose and the reduced feature subset with 10 features is extracted. The reduced feature subset is tested with all the techniques and then ensembling and boosting is also done for the reduced feature subset. Results supported by Receiver Operating Characteristic (ROC) curve indicate that the important features extracted by the *t*-statistic based feature subset selection yielded very high accuracies for all the techniques. Furthermore, the results are superior to those reported in previous studies on the same data set.

1. Introduction

The late 1990s have witnessed a phenomenal growth of the Internet based companies. During the last decade, the Internet has taken the form of an alternate "shopping mall". Companies have started selling on the Web wide range of merchandise such as pet supplies, garden tools, cosmetics, books etc. Some of these companies like Barnes and Nobles, Wal-Mart, etc. have a physical presence in retailing. These are the brick-and-mortar corporations. Many of these companies also have online subsidiaries, which produce their own financial reports separately from the parent company. There are other companies, which have only an online existence in the Internet. These are known as the dotcoms or the click-and-mortar corporations, examples of which are Amazon.com, expedia.com, etc. As the number of Internet users across the world increases at a steady rate, retailing over the Internet continues to gain popularity. It is stated in a report (Jupiter Research Corporation, 2003) that the online retail sector will experience an average annual growth rate of 21% between 2002 and 2007. The report also states that, by 2007, more than 5% of US retail sales will be transacted online. On a similar vein, Forrester Research has also predicted that e-commerce sales, growing at a steady rate of 19% per year, will increase to $229.0 billion in 2008 from $95.7 billion in 2003, with the total number of online shopping households in the US reaching 63 million by 2008 (Rush, 2003). However, the phenomenal growth in Internet related e tailing has suffered a severe setback in the early years of the new millennium. In 2001, Forrester Research had indicated that weak financial strength, increased competition, and investor flight would drive most click-and-mortar companies out of business by 2001 (Grenier, 2003). At the same time, the Gartner Group had predicted that as many as 95% of all click-and-mortar companies would fail by 2002 (Cane, 2000). It is clear that the prediction of the financial pundits has come true and the dotcom bubble has burst. Within a very short time, several corporations that have seen phenomenal growth in their stock prices in the late 1990s have gone out of business. In 2000, when CNNfn.com had asked the market data and research firm Birinyi Associates of Westport, Connecticut, to calculate the market value of the stocks in the Bloomberg US Internet Index, they found that the combined market values of all stocks had fallen to $1.193 trillion from $2.948 trillion at their peak, with a loss of $1.755 trillion in only seven months (Kleinbard, 2000). Several factors responsible for this demise are the dotcoms' inability to improve revenues and earnings, failure to post-profits, attempt to capture a major market share in the smallest possible time, and tendency to operate in limited geographical areas (Sharma, 2001).

The objective of this research is to conduct a rigorous and by far the most comprehensive analysis of the financial statements of

* Corresponding author. Tel.: +91 40 23534981x2042; fax: +91 40 23535157.
E-mail addresses: dkchandra@idrbt.ac.in (D.K. Chandra), rav_padma@yahoo.com (V. Ravi), bose@business.hku.hk (I. Bose).

0957-4174/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2008.05.047

dotcom companies and to discover the factors responsible for their survival or failure. In other words, we build an early warning system for bankruptcy prediction of dotcom companies. The auxiliary objective of the present study is to find the most important financial ratios for predicting the performance of dotcom companies based on historical data. The results of this research may offer important directions to emerging companies that plan to do business on the Internet in the future. It will also provide guidelines about which financial factors to monitor closely for long-term competitive advantage in the market. Further, it will be very useful for the investors planning to invest in such companies.

The rest of the paper is organized as follows. Section 2 reviews the research done in bankruptcy prediction of banks and firms. Section 3 overviews the techniques applied in this paper. Section 4 describes the construction of hybrid intelligent systems via ensembling and boosting. Section 5 presents the results and discussions. Finally, Section 6 concludes the paper.

2. Literature review

The prediction of bankruptcy for financial firms and banks has been the extensively researched area since late 1960s pioneered by (Altman, 1968). Creditors, auditors, stockholders and senior management are all interested in bankruptcy prediction because it adversely affects all of them (Wilson & Sharda, 1994). The most precise way of monitoring banks is by conducting on-site examinations on the premises of a firm or a bank by regulators every 12–18 months, as stipulated by the Federal Deposit Insurance Corporation Improvement Act of 1991. They use a six part rating system to indicate the safety and soundness of the institution. This rating, referred to as the CAMELS rating, evaluates banks and firms in terms of: Capital adequacy, Asset quality, Management expertise, Earnings strength, Liquidity, and Sensitivity to market risk. While CAMELS ratings clearly provide regulators with important information, in Cole and Gunther (1995) it is reported that these CAMELS ratings decay rapidly.

Altman pioneered the work of using financial ratios and multiple discriminant analysis (MDA) to predict financially distressed firms (Altman, 1984). However, the usage of MDA relies on the restrictive assumption on linear separability, multivariate normality and independence of the predictive variables. Unfortunately, many of the common financial ratios violate these assumptions. Bankruptcy prediction problem can also be solved using other types of classifiers. In what follows an overview of such works is given. Ohlson employed the logit model to predict firm failure (Ohlson, 1980). Odom and Sharda's employed BPNN and discriminant analysis for Bankruptcy prediction (Odom & Sharda, 1990). Tam applied backpropagation trained neural network (BPNN) for this problem and compared with methods such as MDA, logistic regression, *k*-nearest neighbor (*k*-NN) method and ID3 and concluded that neural network outperformed other prediction techniques (Tam, 1991). In Salchenberger, Mine, and Lash (1992) it is found that the neural network produced fewer or equal number of total errors, type I errors, and type II errors for each of the forecast periods compared to the logit model.

Later (Serrano-Cinca, 1996) compared the performance of SOM with LDA and BPNN in financial diagnosis. The data set consisted of Altman's variables. He proposed two hybrid neural systems viz., (i) a combination of LDA with SOM, where LDA calculated the Z-score for each firm, which was superimposed onto SOM to obtain isosolvent regions, (ii) a combination of BPNN with SOM. Rahimian, Singh, Thammachote, and Virmani (1996) compared the performance of (i) BPNN, (ii) Athena, an entropy-based neural network and (iii) single layer perceptron on the bankruptcy prediction problem. They compared them with Odom and Sharda (1990) BPNN and discriminant analysis also. Then Olmeda and Fernandez (1997) compared the accuracy of classifiers in stand-alone mode, on bankruptcy prediction problem. They also developed a hybrid system integrating them. Varetto (1998) employed a genetic algorithm (GA) for bankruptcy prediction and compared its performance with that of LDA. Gorzalczany and Pista (1999) presented two different hybrid intelligent decision support systems viz., (i) neuro-fuzzy classifier (N-FC) and (ii) rough classifier (RC) for firm bankruptcy prediction.

McKee (2003) developed a rough set based bankruptcy prediction model. Rough set analysis produced better results when the attribute domains for continuous variables were finite sets of low cardinality. Atiya (2001) surveyed all the prediction techniques applied to the bankruptcy prediction problem and proposed new financial indicators in addition to the traditional ones. Swicegood and Clark (2001) compared DA, BPNN and human judgment in predicting bank failures. The variables were taken from the bank call reports. He used these in the design of a new neural network model. Park and Han (2002) proposed a hybrid of analytical hierarchy process, case based reasoning and *K*-nearest neighbor technique for bankruptcy prediction. Tung, Quek, and Cheng (2004) proposed a new neural fuzzy system, viz., the generic self-organizing fuzzy neural network based on the compositional rule of inference, GenSoFNN-CRI(S), to predict banking failure.

Becerra, Galvao, and Abou-Seads (2005) analyzed the use of linear discriminant models, multi-layer perceptron and wavelet networks for corporate financial distress prediction. They reported that the nonlinear models may be a valid alternative to the linear discriminant models and wavelet networks may have advantages over the multi-layer perceptron. Shin, Lee, and Kim (2005) applied SVM to the problem of corporate bankruptcy prediction. They concluded that SVM outperformed the MLFF-BP in terms of accuracy, as there was reduction in training dataset size. Canbas, Caubak, and Kilic (2005) proposed a methodology for constructing the integrated early warning system that can be used as a decision support tool in bank examinations for detection of banks experiencing serious problems. They applied the methodology to Turkish banks dataset. Ryu and Yue introduced isotonic separation for prediction of firm bankruptcy (Ryu & Yue, 2005).

Recently, a fuzzy rule based classifier for bankruptcy prediction was proposed (Ravikumar & Ravi, 2006). They reported that fuzzy rule based classifier outperformed the well-known technique, MLFF-BP in the case of US banks data. Ravi, Ravi Kumar, Ravi Srinivas, and Kasabov (2007) proposed a semi-online training algorithm for the radial basis function neural networks (SORBF) and applied it to bankruptcy prediction in banks. Semi Online RBFN without linear terms performed better than techniques such as ANFIS, SVM, MLFF-BP, RBF and Orthogonal RBF. Cheng, Chen, and Fu (2006) combined Radial Basis Function Network with Logit Analysis Learning to predict Financial Distress. They compared the proposed technique with logit analysis and a back propagation neural network and found that their method is superior to both the techniques. RaviKumar and Ravi (2006) proposed an ensemble classifier for the bankruptcy prediction problem based on a host of intelligent techniques. The ensemble classifier was developed using simple majority voting scheme and as part of the ensemble they employed seven classifiers such as ANFIS, SVM, RBF, SORBF1, SORBF2, Orthogonal RBF and MLFF-BP. They reported that, models ANFIS, SORBF2, MLP are the most prominent as they appeared in the best ensemble classifier combinations.

In another work, Ravi Kumar and Ravi (2007) conducted a comprehensive review of all the works reported using statistical and intelligent techniques to solve the bankruptcy prediction problem in banks and firms during 1968–2005. It compares the techniques in terms of prediction accuracy, data sources, timeline of each study wherever available. Further, Ravi, Kurniawan, Peter, and