Applied Soft Computing 12 (2012) 2274–2288

Contents lists available at SciVerse ScienceDirect

Applied Soft Computing

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

Visualization and dynamic evaluation model of corporate financial structure with self-organizing map and support vector regression

Mu-Yen Chen

Department of Information Management, National Taichung University of Science and Technology, 129 Sec. 3, San-Min Road, Taichung 40444, Taiwan, ROC

ARTICLE INFO

Article history:
Received 25 February 2010
Received in revised form 11 August 2011
Accepted 18 March 2012
Available online 4 April 2012

Keywords:
Artificial neural networks
Self-organizing map
Support vector regression
Financial bankruptcy

ABSTRACT

Prediction of financial bankruptcy has been a focus of considerable attention among both practitioners and researchers. However, most research in this area has ignored the non-stationary nature of corporate financial structures. Specifically, financial structures do not always present consistent statistical tests at each point of time, resulting in dynamic relationships between financial structures and their predictors. This characteristic of financial bankruptcy presents a significant challenge for any single artificial prediction technique. Therefore, this paper will propose a multi-phased and dynamic evaluation model of the corporate financial structure integrating both the self-organizing map (SOM) and support vector regression (SVR) techniques. In the 1st phase, the inputs to the SOM are financial indicators derived from listed companies' public financial statements adopting the principle component analysis (PCA) to extract useful indicators with a strong influence that each year determines the company's position on the SOM. In addition, we used the SOM to visualize and cluster each corporate in the 2D map. We also investigated each cluster and classified them into healthy and bankrupt-prone ones based on their regions in visualizing the 2D map. In the 2nd phase, we drew the trajectory for the healthy and the bankrupt-prone companies for consecutive years in a 2D map. Therefore, several visualized and dynamic patterns of corporate behavior could be recognized. In the 3rd phase, we used the SVR method to forecast the future trend for corporate financial structure. In addition, this research also compared the hybrid SOM–SVR architecture with single SOM, SVR, and Learning Vector Quantization (LVQ) algorithms. The results showed that the proposed methodology outperformed the other methods in both prediction accuracy and ease of use.

1. Introduction

Leading economists called the financial crisis of 2007–2010 the worst financial crisis since the Great Depression of the 1930s [6]. There were many causes leading up to this major economic shock in the U.S., such as the housing bubble, credit boom, sub-prime lending, predatory lending, and incorrect pricing of risk and collapse of the shadow banking system. All of these contributed to the failure of key businesses, a decline in consumer wealth estimated to be in the trillions of U.S. dollars, substantial financial commitments incurred by the U.S. governments, and a significant decline in economic activity [3]. Initially the companies affected were those directly involved in home construction and mortgage lending such as Northern Rock and Countrywide Financial, as they could no longer obtain financing through the credit markets. Over 100 mortgage lenders went bankrupt during 2007 and 2008. The crisis hit its peak in September and October 2008. Several major institutions either failed, were acquired under duress, or were taken over by the government, including included Lehman Brothers, Merrill Lynch, Fannie Mae, Freddie Mac, Washington Mutual, Wachovia, and AIG [2]. As a result, the U.S. economic crisis affected the global economy and put pressure on all the major sources of external revenue for developing countries, including exports, remittances, foreign direct investment, portfolio equity flows, and aid. In Taiwan, the bankruptcy of the Procomp Corp. and the Cdbank Corp. also caused a major upheaval in Taiwan's financial market, and any related investors incurred heavy losses. Today, financial distress and bankruptcy forecasting has become an increasingly important function with a significant impact on the lending decisions being made and the profitability of financial institutions.

The prediction of failure of financial firms has been an extensively researched area since the late 1960s. A variety of statistical methods have been applied to solve the bankruptcy prediction problem for listed companies. Beaver introduced a univariate technique for the classification of firms into two groups using financial ratios [4]. These financial ratios were classified into six categories, including cash flow, net-income, debt to total-asset, liquid-asset to total-asset, liquid-asset to current debt, and turnover ratios. The data was collected from financial statements. Altman was the first researcher to use the multivariate discriminant analysis (MDA) and *Z*-score to predict the failures of firms in different industries [1]. As the result, the accuracy rate for determining a healthy company is

E-mail address: mychen@nutc.edu.tw

1568-4946/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.asoc.2012.03.046

79% and 93.5% for one that will go bankrupt. Blum used the factor analysis to determine the critical financial ratios to construct the Failing Company Model [5]. Ohlson proposed a logistic regression (logit) model to determine the impact factors for 2058 healthy and 105 bankruptcy-prone companies [23]. The result showed that company size, financial structure, management performance, and asset liquidity influenced bankruptcy probability. More recently, Jones and Hensher compared the mixed logit model with the multinomial logit models for predicting firm distress [16]. Canbas et al. combined discriminant analysis, logistic regression, probit and principal component analysis (PCA) to implement an early warning system for bankruptcy prediction [7]. Doganay et al. integrated multiple regression, discriminant analysis, logit and probit to construct a financial distress prediction model [11]. Recently, many studies have demonstrated that artificial intelligence (AI) such as neural networks can be an alternative method for financial distress prediction [17,24,25].

Current statistical and AI methods used in this field present some drawbacks. Real world applications of statistical methods are limited by strict assumptions such as linearity, normality, independence among predictor variables and pre-existing functional forms relating to the criterion variable and the predictor variable [14]. Researchers using AI fail to explain why they adopt particular financial ratios as inputs for neural networks, and fail to investigate the trajectory of healthy and bankrupt companies, resulting in a lack of visualized or dynamic patterns of corporate behavior for current and future trends. In addition, improving accuracy prediction is still the leading concern in the field [15]. Especially for stock market predictions, even a slight improvement on prediction accuracy could have a positive impact on investment profitability. Finally, a hybrid system for prediction and classification outperforms the traditional system [9,13,22]. Tay and Cao found combining a self-organizing map (SOM) with support vector regression (SVR) can successfully predict financial time series and stock index activity [27]. Therefore, this paper proposes a model of financial distress prediction integrating the SOM and SVR techniques. The proposed model uses statistical methods to process feature selection and draw trajectories for healthy/bankrupt companies. Proper feature selection can reduced data complexity and improve prediction accuracy. The main objectives of this paper are to (1) apply the SOM technique to construct a visualization and dynamic evaluation model for corporate financial behavior, (2) use the SVR technique to improve the accuracy of financial distress predictions, and (3) provide investors and investment monitoring organizations with financial information and investment suggestions.

The rest of this paper is organized as follows. A literature review of related studies is provided in Section 2. In Section 3, we discuss our research methodologies and pre-processing of our experiment materials. In Section 4, we use SOM to construct the static financial bankruptcy prediction model. Then in Section 5 we establish a trajectory using these experimental companies to construct a dynamic model. To prove the performance of the future trend prediction of our approach, we carried out a SVR with several experiments as described in Section 6. In Section 7, we draw our conclusions regarding financial distress forecasting and discuss some future work.

2. Literature review

2.1. Self-organizing map

The SOM algorithm was originally introduced by the Kohonen [18,19]. The SOM (also known as the Kohonen feature map) algorithm is one of the best known artificial neural network algorithms. In contrast to many other neural networks based on supervised learning the SOM is based on unsupervised learning. The SOM is a unique kind of neural network in the sense that it constructs a topology-preserving mapping of the training data where the location of a unit carries semantic information. For this reason, the main application of this algorithm is the clustering of data. With the SOM, a two-dimensional display of the input space is obtained which naturally lends itself to easy visualization. The most important practical applications of the SOM are in exploratory data analysis, pattern recognition, speech analysis, robotics, industrial and medical diagnostics, instrumentation and control, and literally hundreds of other tasks.

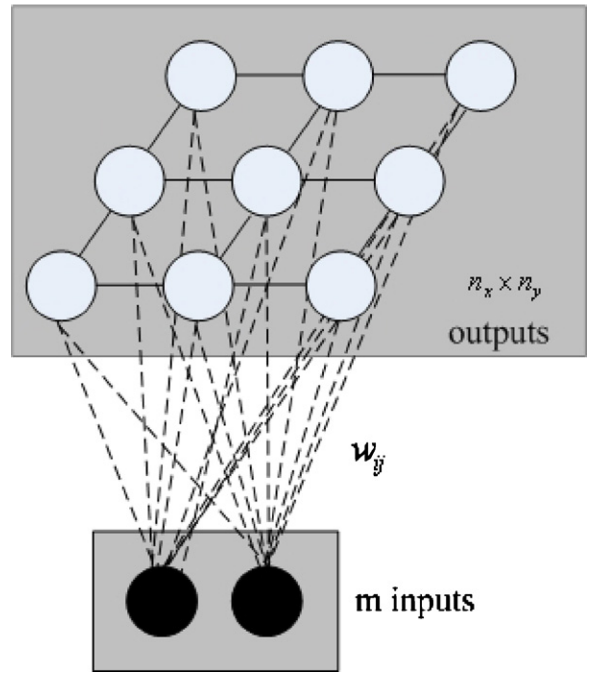

Fig. 1. SOM architecture with m neurons inputs and $n \times n$ neurons outputs.

The SOM model is made up of two neural layers. The input layer has as many neurons as it has variables, and its function is merely to capture the information. Let m be the number of neurons in the input layer, and let $n_x \times n_y$ be the number of neurons in the output layer which are arranged in a rectangular pattern with x rows and y columns, which is called "the map". Each neuron in the input layer is connected to each neuron in the output layer. Thus, each neuron in the output layer has m connections to the input layer. Each one of these connections has a synaptic weight associated with it. Let w_{ij} the weight associated with the connection between input neuron i and output neuron j. Fig. 1 shows a visual representation of this neural arrangement.

Initially, w_{ij} is given random values. These values will be corrected as the algorithm progresses (training). The training is carried out by presenting the input layer with financial ratios, one firm at a time. Let r_{ik} be the value of ratio i for firm k. This ratio will be read by neuron i. Let w_{ijk} be the weighted value w_{ij} of firm k. The algorithm takes each neuron in the output layer one at a time and computes the Euclidean distance as a similarity measure

$$d(j,k) = \sqrt{\sum_i (r_{ik} - w_{ijk})^2} \tag{1}$$

The output neuron for which $d(j,k)$ is smallest is the "winner neuron". Let such neuron be k^*. The algorithm now proceeds to change the synaptic weights w_{ij} in such a way that the distance $d(j,k^*)$ is reduced. A correction takes place, which depends on the number of iterations already performed and on the absolute value