Business Intelligence and Analytics to Prediction of Going Concern using Neuro-Fuzzy Approach

Yee Ming Chen, Yuan Ze University, Taiwan, Republic of China(R.O.C)
Yu-Pu Chiu, Yuan Ze University, Taiwan, Republic of China (R.O.C)

The Asian Conference on Technology, Information & Society
Official Conference Proceedings 2014

Abstract
Rapid advances in technology, enterprise environmental changes and increasing competition has affected the risk of investment. Going concern prediction (GCP) is an important element in investor's evaluation. The evaluation of a enterprise 's going concern status is not an easy task. To assist auditors, going concern prediction models based on statistical methods such as multiple discriminant analysis and multiple linear regression analysis have been explored with some success. Nowadays, the business intelligence and analytics (BI&A) has attracted more and more attentions, which is required to manage immense amounts of data quickly. However, current researches mainly focus on the amount of data. In this paper, the other two properties of BI&A, which include the high dimensionality of data, and the dynamical change of data, are discussed. This study attempts to look at a different and more recent approach—Adaptive Neuro Fuzzy Inference System (ANFIS). ANFIS has effectively solved many large-scale, and dynamical problems. This study explores and compares the usefulness of logistic regression and proposed ANFIS in predicting an enterprise's going concern status. The sample data comprise financial ratios for 165 going concerns and 165 matched non-going concerns. The classification results from view of significance test and predictive accuracy which indicate the potential usefulness of BI&A in a going concern prediction context. These results also indicate that ANFIS shows acceptable performance in terms of accuracy and comprehensibility, and it is an appropriate choice for auditors to assess potential clients and as a means to identify non-going concern enterprises that might require further consideration.

Keywords: Business intelligence and analytics, Big data analytics, Neuro fuzzy approach

Introduction

Business intelligence and analytics (BI&A) and the related field of big data analytics have become increasingly important in both the academic and the business communities over the past few decades. By carefully analyzing the application and data characteristics, researchers and practitioners can then adopt or develop the appropriate analytical techniques to derive the intended impact. Emerging analytics research opportunities can be classified into five critical technical areas—(big) data analytics, text analytics, web analytics, network analytics, and mobile analytics—all of which can contribute to to BI&A. Data analytics refers to the BI&A technologies that are grounded mostly in statistical analysis and data mining. Grounded in statistical theories and models, multivariate statistical analysis covers analytical techniques such as regression, factor analysis, clustering, and discriminant analysis that have been used successfully in various business applications. Since the late 1990s, various data mining algorithms have been developed by researchers from the business intelligence, algorithm, and database communities. Data mining in BI&A area aims to identify valid, novel, potentially useful and understandable correlations and patterns in data (Chung and Gray, 1999). More importantly, data mining techniques such as neural networks and soft computing provide a different approach to predictive modeling. Due to their prediction and classification capabilities, data mining techniques have been employed to facilitate the auditing process, to predict corporate performance, and to facilitate credit risk estimation. The going concern prediction (GCP) has become an important research in finance areas. In general, the objective of GCP is to develop models that can extract novel knowledge from previous observations and appraise corporate status. In this study, we have compared logistic regression and neuro-fuzzy approach in the context of going concern prediction. Both techniques are considered data mining techniques, but only logistic regression can be considered a traditional statistical method. Neuro-fuzzy approaches are developed for pattern recognition and prediction purposes in the area of business intelligence. Neuro-fuzzy approach via utilizing a large number financial data can be extracting, valuable and unknown knowledge dynamically. In this paper, the neuro-fuzzy approach of adaptive neuro-fuzzy inference system (ANFIS) model for prediction have been conducted in GCP area and the findings indicate that this technique is able to predict the going concern status of firms and accounting data are useful in GCP. This study explores ANFIS model and compares the usefulness of logistic regression (a traditional statistical method included in most data mining software as a data mining technique) in predicting a firm's going concern status. The sample data comprise financial ratios for 165 going concerns and 165 matched non-going concerns as a benchmark in auditor judgment. Results from this study will help a manager to keep track of company's performance and to identify significant problems and take efficient measure to reduce the coincidence of failure. The remainder of the paper is organized into the following three sections. Section 2 reviews the going concern literature and the third section ANFIS explores the potential usefulness of and discusses the comparative analysis of logistic regression in the going concern prediction context. Finally, the concluding section highlights the limitations of the study and suggests directions for future research.

Background

The going concern prediction is a significant argument that a company will continue to operate for an indefinite period of time i.e. long enough to meet its objectives and fulfill its commitments. The prediction models of going-concern were discussed extensively in prior studies. Most of previous researchers used financial indicators to predict enterprise financial distress and bankruptcy. The literature on going concern prediction dates back to 1976, shortly after the issuance of Statement of Auditing Standards (SAS) No. 2, which was the first SAS to detail specific considerations for the auditor's assessment of a company's going concern status. The first study about going concern prediction published shortly after the issuance of SAS No.2 in 1974 by McKee (1976). Subsequent research applied statistical methods analysis to test going concern predicting models (e.g. Altman, 1968; Collins & Green, 1972; Balcaen,et al., 2006). Common tools for statistical analysis include linear regression and logistic regression. But, sstatistical models have certain distributional hypothesis that financial statement data do not always fit. Hence some non-parametric techniques have been developed to overcome the constraints of traditional statistical models. Most of them belong to BI&A domain such as decision trees, artificial neural networks (ANN), and support vector machines (SVM). Most researchers use one of the techniques to compare the prediction performance with other techniques for a specific data set (Koh et al., 1999; Shin et al., 2005; Zheng et al., 2007). There are a number of arguments which promote the consideration of the hybrid analysis model for going concern prediction using some BI&A techniques. In addition, these studies have shown that BI&A models outperform traditional statistical models. Kumar and Ravi (2007) provide a detailed review of these models in the domain of going concern prediction.

The going concern prediction studies often use samples of bankrupt firms versus non-bankrupt firms to assess model accuracy. While the ability of firms to continue in business is a concern for firms in any country, most studies have developed going concern prediction models for U.S. firms. In this research, the sample data for constructing the going concern prediction models is select from Predicast's F & S Index of Corporate Changes for the years 1980 to 1992. It comprises 165 going concerns and 165 matched non-going concerns in the US based on a compilation of bankruptcy data from numerous newspapers and periodicals. Studies usually report prediction accuracies separately for non-going concern firms (those which are not likely to survive) and going concern firms. In addition, the going concern prediction literature refers to Type I and Type II errors, which are applicable to going concern prediction models. Type I errors are the misclassification of non-going concern (i.e. bankrupt) firms as going concerns. Type II errors are the reverse – going concern firms misclassified as non-going concern firms. Type I errors are generally considered more costly than Type II errors for several reasons including loss of business (audit clients), damage to a firm's reputation, and potential lawsuits/court costs (see for example Koh [1987]). Therefore, the predictive accuracies discussed here refer to the accuracies obtained for non-going concern firms, unless the results were not presented separately for non-going concern and going concern firms.

Construction of going concern prediction models

In this paper, we have investigated BI&A techniques. In this section, we used the logistic regression and Adaptive Neuro Fuzzy Inference System (ANFIS) models to

predict the going concern and non-going concern firms. The sample data for constructing the two going concern prediction models is taken from Predicast's F & S Index of Corporate Changes for the years 1980 to 1992. Based on the sample data, the objective is to construct a prediction model for the going concern status of firms (a dichotomous dependent variable) based on the six financial ratios (independent variables).The following independent variables were applied to the going-concern companies to select the sample that fits the current study:

(1) Quick assets to current liabilities (QACL).
(2) Market value of equity to total assets (MVTA).
(3) Total liabilities to total assets (TLTA).
(4) Interest payments to earnings before interest and tax (IEBT).
(5) Net income to total assets (NITA).
(6) Retained earnings to total assets (RETA).

To derive the going concern prediction model, the sample data is partitioned into the following two data sets: a construction or training sample (comprising approximately 70 percent of the original samples) and a validation or testing sample (comprising the remaining 30 percent of the original samples).

Logistic regression

This study used logistic regression to build the prediction model of going-concern, logistic regression is an appropriate traditional statistical method to use. The logistic regression results are summarized in Table 1. As can be seen, the logistic regression model indicate a good fit. At a 0.05 level of significance, QACL (p-value= 0.0114), TLTA (p-value = 0.0001) and RETA (p-value= 0.0004) are statistically significant. The numerical sign of the respective coefficient shows that a higher level of liquidity (QACL), a lower of leverage (TLTA) and a higher level of profitability (RETA) are associated with a higher probability of continued going concern status. Based on the results, it can be concluded that the logistic regression going concern prediction model is acceptable.

Table 1. Logistic regression for prediction model of going concern

Variable	Coefficient	P-value
Intercept	13.64	0.0002
QACL	2.74	0.0114
MVTA	0.85	0.2342
TLTA	-23.41	0.0001
IEBT	-0.45	0.4215
NITA	3.74	0.0734
RETA	17.89	0.0004

ANFIS Model

The adaptive network-based fuzzy inference systems (ANFIS) is used to solve problems related to parameter identification. This parameter identification is done through a hybrid learning rule combining the back-propagation gradient descent and a

least-squares method. ANFIS is basically a *graphical* network representation of Sugeno-type fuzzy systems endowed with the neural learning capabilities. The network is comprised of nodes with specific functions collected in layers. ANFIS is able to construct a network realization of IF / THEN rules. Consider a Sugeno type of fuzzy system having the rule base

(I). If x is A_1 and y is B_1, then $f_1 = c_{11}x+c_{12}y+c_{10}$
(II). If x is A_2 and y is B_2, then $f_2 = c_{21}x+c_{22}y+c_{20}$

Let the membership functions of fuzzy sets *Ai, Bi, i*=1,2, be , μ_{Ai} μ_{Bi}.

In evaluating the rules, choose *product* for T-norm (logical *and*).

(1). Evaluating the rule premises results in

$$w_i = \mu_{A_i}(x)\mu_{B_i}(y), \quad i = 1,2 \tag{1}$$

(2) Evaluating the implication and the rule consequences gives

$$f(x,y) = \frac{\sum_i w_i f_i}{\sum_i w_i} \tag{2}$$

This can be separated to phases by first defining

$$\overline{w}_i = \frac{w_i}{w_1 + w_2} \tag{3}$$

(3) Then f can be written as

$$f = \overline{w}_1 f_1 + \overline{w}_2 f_2 \tag{4}$$

All computations can be presented in a diagram form. ANFIS normally has 5 layers of neurons of which neurons in the same layer are of the same function family.

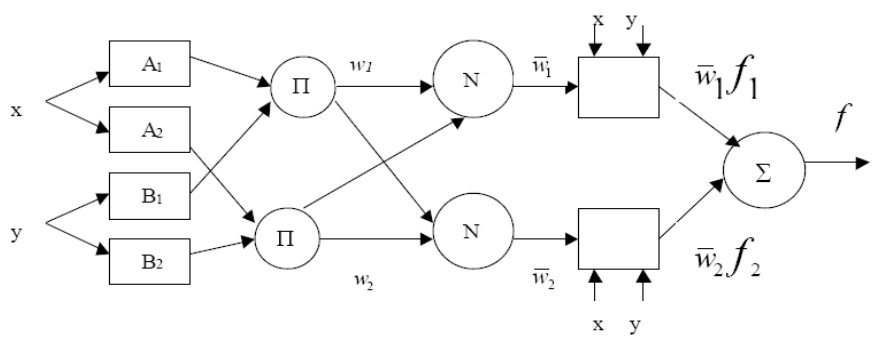


Figure 1 Structure of the ANFIS network.

We get our going concern prediction model on the Predicast's F & S Index of Corporate Changes database. The proposed method for Predicast's F & S Index of Corporate Changes ANFIS model is depicted in Fig. 2 which shows actual and predicted dichotomous value (going concern or otherwise) using MathWorks MATLAB as a software support. The ANFIS model in this study was simulated

using MATLAB (R2012b). ANFIS with two input membership functions, generalized bell membership function and our linear generating function, were trained.

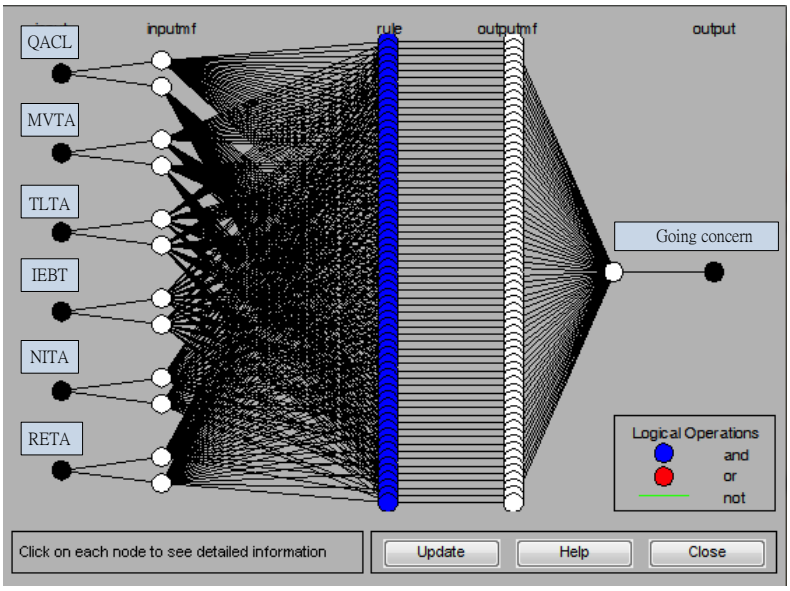


Figure 2 Going concern prediction ANFIS model implemented using Matlab software

The data flow and train process of ANFIS model is shown in Fig. 3.

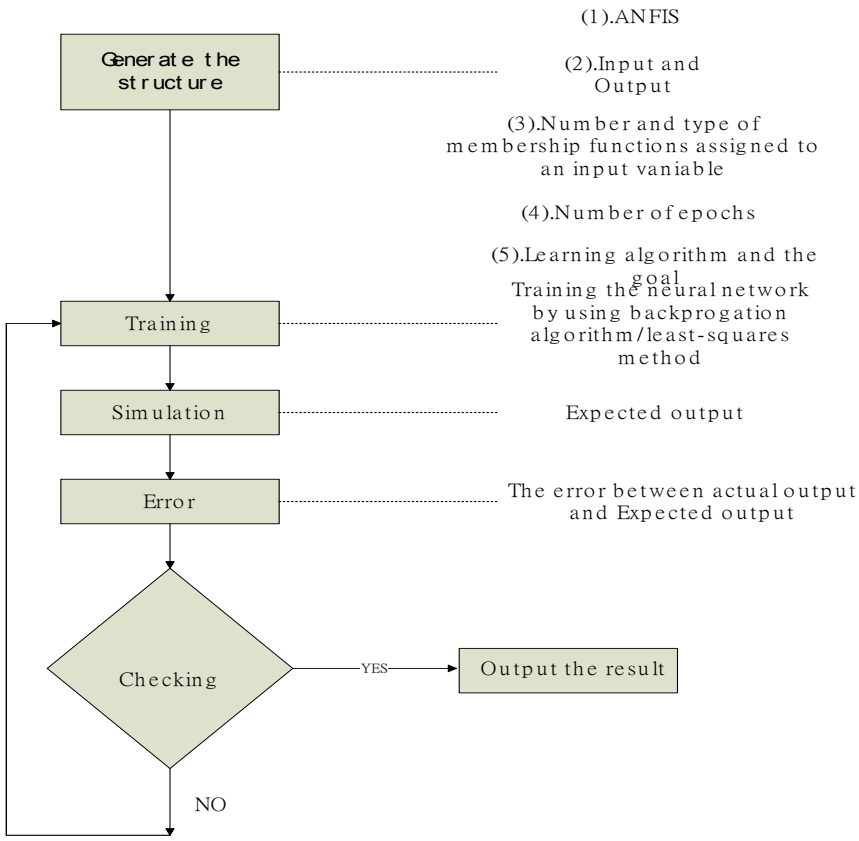


Figure 3 Data flow and train processing in ANFIS model

The data flow and train process for ANFIS model can be divided into two stages as shown in Fig. 3. In the first stage, the generated of the ANFIS structure (depicted in Fig. 2). In the second stage, ANFIS training process starts by obtaining a data set (input-output data pairs from the Predicast's F & S Index of Corporate Changes database) and dividing it into training and checking data sets. Training data constitutes a set of input and output vectors. The data is normalized in order to make it suitable for the training process. The training data set is used to find the initial premise parameters for the membership functions using backpropagation algorithm. The consequent parameters are found using the least-squares method. The train process is terminated when the error becomes less than the threshold value. Then the checking data set is used to compare the model with actual data. Figure 4, for instance, shows the error curves for 100 epochs of ANFIS training. The green curve gives the training errors and the red curve gives the checking errors. The minimal checking error occurs at about epoch 45, which is indicated by a circle. Notice that the checking error curve goes up after 50 epochs, indicating that further training overfits the data and produces worse prediction.

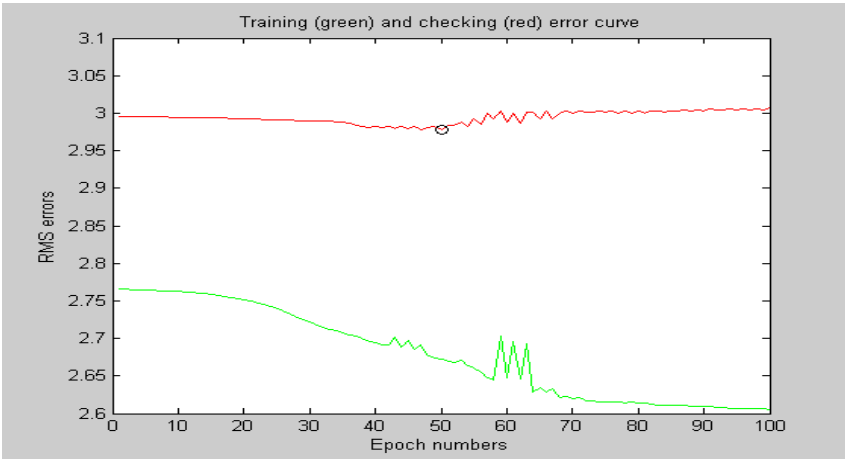


Figure 4 ANFIS training and checking errors

After training association rules in the form of IF-then rules are generated and extracted. an example of initial and final (after rule extraction) decision surfaces are given in Fig. 5. The input-output surface of the ANFIS model is shown in the plot below.

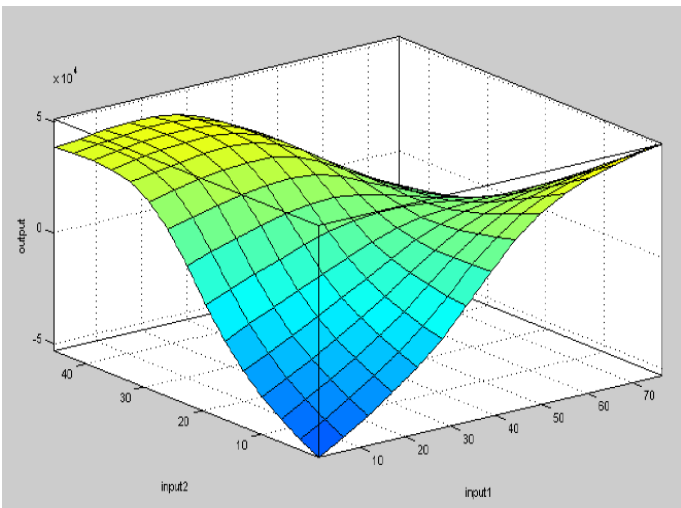


Figure 5 Final decision surface for input 1 and input 2

The predictive capability of using neural network and ANFIS approaches are compared using statistics in the next subsection.

Predictive Results

The two models of logistic regression and ANFIS that were constructed based on BI&A. We use the Predicast's F & S Index of Corporate Changes test to examine whether the predictive performance of the ANFIS method is significantly higher than of traditional statistical method. The predictive ability of a model can also be impacted by whether the results are from tests of a hold-out sample. Hence, to provide an unbiased assessment of the performance of the two prediction models, the validation sample is used 10-fold cross-validation. This method splits the data into two subdivisions: a training set and test set. Quality of the prediction assessed on the test set. In 10-fold cross-validation the data is firstly partitioned into 10. Then, 10 iterations of training and test are done such that in each iteration a different fold of the data is held-out for validating while the rest 9 folds are used for learning and 10 outputs from the folds can be averaged and can produce a single prediction. The validation predictive results of each models listed in Table 2.

Table 2 Predeiictive accuracy(%) of hold-out data

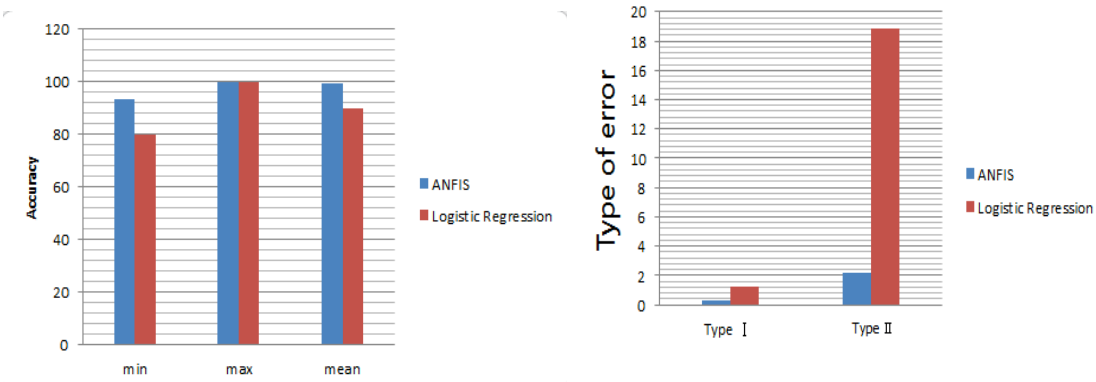
Data sets	ANFIS			Logistic Regression		
	Accuracy	Error Type I	Error Type II	Accuracy	Error Type I	Error Type II
1	100.0	0	0	86.51	0	28.46
2	100.0	0	0	93.15	0	16.51
3	93.35	0	12.52	80.05	0	38.42
4	100.0	0	0	86.55	12.51	14.16
5	100.0	0	0	86.13	0	26.23
6	100.0	0	0	100.0	0	0
7	100.0	0	0	100.0	0	0
8	98.14	0	0	92.57	0	14.32
9	98.14	3.14	9.36	93.15	0	12.53
10	100.0	0	0	80.11	0	38.1
Min	93.35	0	0	80.05	0	0
Max	100.0	3.14	9.36	100.0	12.51	38.42
Mean	99.15	0.31	2.19	89.82	1.25	18.87
Median	100.0	0	0	89.56	0	15.41
Variance	2.018	0.9420	4.432	6.79	3.75	13.03

The validation results also indicate that the ANFIS model (accuracy =99.15%) outperforms the logistic regression model (accuracy=89.82%). Hence, if only one model is to be selected, it will be the ANFIS model as it is expected to perform better on new observations as compared to the traditional models.

Type I error is the probability that a company with non-going concern status to be predicted as a company with going concern status and Type II error is the probability that a company with going concern status to be predicted as a company with non-going concern status. In holdout data type I and II error are also equal to 0.31 and 2.19 percent in ANFIS model and 1.25 and 18.87 percent for obtained model by logistic regression. Determine which of the models that are more applicable than others in GCP, we make the significance test between two models. As shown in Table 5, the result of the Predicast's F & S Index of Corporate Changes test at 5% level indicates that there are significant differences between the two models in GCP.

Table 5 Results of significance test between two models

Models	Logistic Regression
ANFIS Model	$-2.569^{a}(0.011)^{b}$
[a]= *t* statistic;[b]= p-value	



(a) The map of mean accuracy of both models (b) The map of mean of difference types of errors

Figure 6. Going concern prediction experiment al results

From Fig. 6(a), we can be understood that ANFIS outperforms logistic regression on the mean accuracy. Fig. 6(b) shown the related to these two types of errors are very different. The resulting shown incorrectly predicted a company with going concern as a company with non-going concern status (Type II error) is much larger than the Type I error (incorrectly predict a company with non-going concern as a company with going concern status). Type I errors are generally considered more costly than Type II errors for several reasons including loss of business.

Conclusions

In recent years, BI&A data mining has gained widespread attention and increasing popularity in the business world. Successful neuro-fuzzy approach applications have been reported and recent surveys have found that ANFIS has grown in usage and effectiveness. The main purpose of the paper is to predict company's going concern by using the ANFIS model and to show its adequacy experimentally. The results of the experiments show that the suggested model predicts company's situation more

accurately in comparison to logistic regression prediction model, especially evaluating bankrupt companies. It can be stressed that a neuro-fuzzy approach in going concern prediction is a potentially powerful alternative or complement to the more commonly used statistical methods. The development of going concern status analysis models to predict business failures can be thought of as early warning systems, which proved to be very helpful for managers, and relevant authorities who can prevent the occurrence of failures.

References

Altman, E. I. (1968). Financial ratios, discriminant analysis and prediction of corporate bankruptcy. The Journal of Finance, 23, 589 – 609.

Aziz, M., Dar, H. (2006). Predicting corporate bankruptcy- where we stand? Corporate governance journal, 6(1), 18-33.

Balcaen, S., & Ooghe, H. (2006). 35 years of studies on business failure: An overview of the classic statistical methodologies and their related problems. The British Accounting Review, 38(1), 63 – 93.

Collins, R. A., & Green, R. D. (1972). Statistical methods for bankruptcy forecasting. Journal of Economics and Business, 32, 349 – 354.

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

Kumar, P. R., & Ravi, V. (2007). Bankruptcy prediction in banks and firms via statistical and intelligent techniques – A review. European Journal of Operational Research, 180(1), 1 – 28.

McKee, T.(1976). Discriminant prediction of going concern status: A model for auditors. Selected Papers of the AAA Annual Meeting .

Shin, K. S., lee, T. S., Kim, H. J, (2005). An application of support vector machines in bankruptcy prediction model. Expert system Application, 28(1), 127-135.

Zheng, Q., J. Yan-Hui.(2007), Financial Distress Prediction on Decision Tree Models , IEEE, Service Operations and Logistics, Informatics, International Conference.

Corresponding author: chenyeeming@saturn.yzu.edu.tw