Management assessment of going concern based on data mining using Adaptive Network Based Fuzzy Inference Systems (ANFIS)

Fezeh Zahedi Fard[1]

Department of Accounting and Management, Neyshabur Branch, Islamic Azad University, Neyshabur, Iran. Email: fzahedi_1362@yahoo.com

Mahdi Salehi
Assistant Professor, Department of Accounting Ferdowsi University of Mashhad, Mashhad, Iran. Email: mahdi_salehi54@yahoo.com

Abstract
Going concern is a fundamental concept for the preparation of financial statements by management. This paper has employed a data mining approach for going concern prediction (GCP) and has applied Adaptive Network Based Fuzzy Inference Systems (ANFIS) based on feature selection method for GCP in Iranian firms, listed in Tehran Stock Exchange (TSE). For this purpose, at the first step, using the stepwise discriminant analysis (SDA) has opted the final variables from among of 42 variables and in the next stage, has applied 10-fold cross-validation to figure out the optimal model for one year ahead. The empirical test signifies that the ANFIS model reached 99.92 and 95.19 percent accuracy rates so as to train and hold-out data.

Keywords
ANFIS; data mining; feature selection; financial ratios; going concern prediction.

1. Introduction

In today's litigious economic atmosphere, the number and the magnitude of non-going concern firms filings have been increasing considerably. On the other hand sometimes the assessment of a firm's going concern status is a tough process. The complexity of the going concern decision has led the development of several models for going concern prediction (GCP). These studies are concentrated on development of GCP models; by using a multiple financial and non-financial variables that might be indicative of going concern decision (Martens et al., 2008). Early studies of going concern prediction were developed by using statistical techniques such as multiple discriminant analysis, Logit, probit and etc. These methods by using historical samples created diagnostic model, even though they cannot inductively learn from new data dynamically, which greatly affects accuracy rate (Sun & Li, 2008). In recent years, data mining, a novel field of intelligent data analysis established , developed and began to appear and grow promptly in the background of abundant data and poor information. It also has developed a new approach for the deep research in finance. Based on this principal by using great database or data warehouse which stores a large number of listed companies' financial data, by utilizing data mining technique extract valuable unknown knowledge dynamically, which can be applied to predict going concern status of companies.

[1] Corresponding author. Tel: +989153116399.

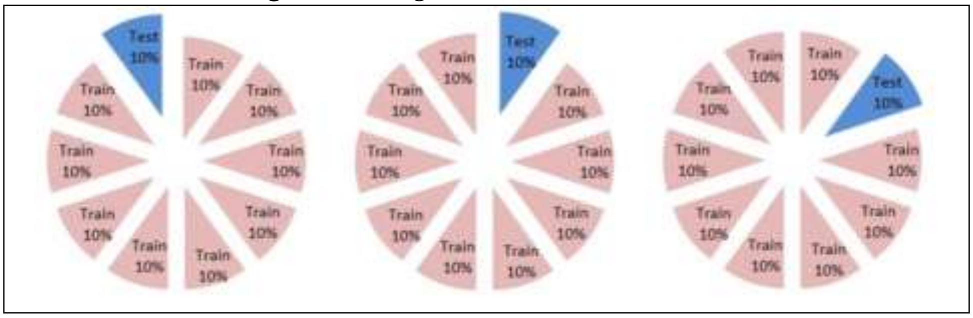
In this paper, Model of GCP will contribute a manager to keep track of firm's performance and will help to identify important trends. The model may not tell the manager what is wrong or right specifically, but it can help him to identify problems and take efficient measure to reduce the coincidence of failure. As well as timely identification of firms' impending failure is indeed desirable (Jones, 1987). In addition, other stakeholders may adopt using predictive model to aid in assessing a firm status. Regulatory organizations are concerned whether a monitored firm is in danger of bankruptcy or not.

2. Research design

2. 1. Cross-Validation

The cross-validation is the standard data mining methodology used to evaluate and compare learning algorithms by splitting the data into two main subdivisions: a training set and test set. Quality of the prediction evaluated on the test set. K-fold cross validation is the primal form of cross-validation. In K-fold cross-validation the data is firstly partitioned into K subsets of approximately or exactly the same size. Then, K iterations of training and test are done such that in each iteration a variant fold of the data is held-out for validation while the rest $K-1$ folds are used for learning and Koutputs from the folds can be averaged and can produce a single estimation (see Fig.1). The advantage of K-fold cross-validation is that all observation are utilized for both training and test sets (Alpaydin, 2010). In data mining and machine learning $K$is typically 10 or 30 that in this study$K = 10$.

Fig.1. Three stage of 10-fold cross-validation



2. 2. The method of Adaptive Network Based Fuzzy Inference Systems

ANFIS is a multi-layer adaptive network-based fuzzy inference system proposed by Jang (1993). This method resembles a fuzzy inference system except that it uses back-propagation for minimizing errors. ANFIS operates in a manner similar to both artificial neural networks and fuzzy logic. In both of them, by the input membership function the input goes through the input layer and by the output membership function the output is display in output layer. Due to the fact that fuzzy logic applies neural networks, a learning algorithm can be applied to alter the parameters until finding an optimal solution. So ANFIS uses either back-propagation or a combination of least squares estimation and back-propagation to appraise the membership function parameters (Jang & Sun, 1997; Chen, 2011).

Assuming that the fuzzy inference system has two inputs (x and y) and one output (f) a common rule set with two fuzzy if–then rules is as follows :

$$\begin{cases} \text{Rule 1: If } x \text{ is } A_1 \text{ and y is } B_1 \text{ , then } f_1 = p_1 x + q_1 y + r_1 \\ \text{Rule 2: If } x \text{ is } A_2 \text{ and y is } B_2 \text{ , then } f_2 = p_2 x + q_2 y + r_2 \end{cases}$$

2. 2. Data collection

The financial data used for this study is obtained from the Tehran Stock Exchange (TSE). The data set used for this study consists of 146 Iranian manufacturing companies in total that has been applied by matched company (73 from bankrupt[2] and 73 from non-bankrupt companies) in duration of 2001- 2011. Because of low number, in terms of industries two groups could not match together.

2. 3. Feature selection

The Proposed variables in this study are based on a combination all variable selection techniques and experiments. Table 1 shows the 42 variables used in this study that all these independent variables are dated t-1. We applied process of future selection by T-test and stepwise discriminant analysis (SDA) at a significant level of 0.05 and selected final variables. The potential advantages of feature selection are: facilitating data visualization and understandable data, reducing the measurement and storage requirements and etc (Guyon & Elisseeff, 2003 ;Ashoori & Mohammadi 2011). The result of SDA process is shown in Table 2. The ratios that are entered in the model are: Total liabilities to total assets (x_9), Retained earnings to total assets (x_{31}), Operational income to sales (x_{36}) and Net income to total assets (x_{34}).

3. Experimental results

The proposed ANFIS model is implemented by using MATLAB ۷.۶. They are the results from the 10 testing data sets (See table 3). This model could classify firms with 99.92 and 95.19 percent overall accuracy rate in the training and hold-out data set, respectively as shown in table 4. In addition, this model could correct classify for going concern firms with 91.71% accuracy rate and 98.75% for non-going concern firms.

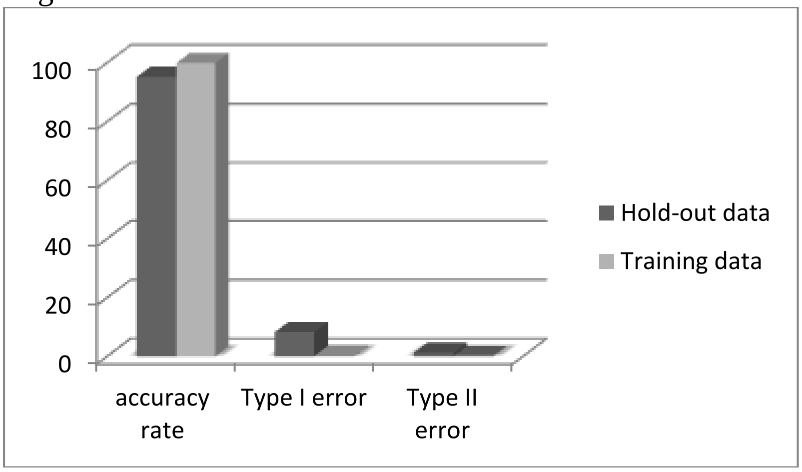


Fig 2. Result obtained by ANFIS for training and hold-out data.

4. Summary & conclusion

In this paper, we considered a set of features that include 42 variables proposed in prior literature dealing with financial status prediction models in Iran. We applied SDA to identify potential variables for predicting model and finally, we selected 4 financial ratios. We constructed ANFIS prediction model based on selected features for t-1. Based on the results, the empirical tests show that ANFIS model achieved 99.92% and 95.19% accuracy rates for training and hold-out data. In summary, obtained results from ANFIS model from 146 companies of Iran indicate that this model has suitable ability to predict going concern status of firms.

[۲] Under paragraph 141 of Iran Trade Law, a firm is bankrupt when its total value of retained earnings is equal or more than 50% of its listed capital.

Table 1
Variables used in the research

No.	Predictor variable name	Financial ratios	Means of group 1	Means of group 2	Sig level
X_1	Earnings before interest & taxes/ Total assets	EBIT/TA	0.18	0.05	0.00
X_2	Long term debt/Shareholders' equity	LTD/SE	0.20	0.56	0.06
X_3	Retained earnings/Stock capital	RE/SC	0.65	0.02	0.00
X_4	Marked value of equity /Total liabilities	MVE/TL	1.40	0.66	0.00
X_5	Marked value of equity /Shareholders' equity	MVE/SE	2.42	2.57	0.22
X_6	Marked value of equity /Total assets	MVE/TA	0.77	0.48	0.00
X_7	Cash /Total assets	Ca/TA	0.05	0.03	0.00
X_8	Log (total assets)	Size	5.25	5.23	0.83
X_9	Total liabilities/Total assets	TL/TA*	0.67	0.80	0.00
X_{10}	Current liabilities/Shareholders' equity	CL/SE	2.27	4.76	0.00
X_{11}	Current liabilities/Total liabilities	CL/TL	0.86	0.85	0.94
X_{12}	(Cash+Short term investments)/Current liabilities	(Ca+STI)/CL	0.11	0.05	0.00
X_{13}	(Receivables+Inventory)/Total assets	(R+Inv)/TA	0.57	0.57	0.88
X_{14}	Receivables/Sales	R/S	0.53	0.40	0.10
X_{15}	Receivables/Inventory	R/Inv	1.18	1.00	0.93
X_{16}	Shareholders' equity/Total liabilities	SE/TL	0.63	0.32	0.00
X_{17}	Shareholders' equity/Total assets	SE/TA	0.35	0.22	0.00
X_{18}	Current assets/Current liabilities	CA/CL	1.31	1.07	0.00
X_{19}	Quick assets/Current liabilities	QA/CL	0.70	0.57	0.00
X_{20}	Quick assets/Current assets	QA/TA	0.37	0.36	0.73
X_{21}	Fixed assets/(Shareholders' equity+Long term debt)	FA/(SE+LTD)	0.60	0.91	0.01
X_{22}	Fixed assets/Total assets	FA/TA	0.22	0.24	0.63
X_{23}	Current assets/Total assets	CA/TA	0.70	0.68	0.66
X_{24}	Cash/ Current liabilities	Ca/CL	0.09	0.04	0.00
X_{25}	Interest expenses/Gross profit	IE/GP	-0.02	-1.21	0.48
X_{26}	Sales/Cash	S/Ca	35.30	44.80	0.11
X_{27}	Sales/Total assets	S/TA	0.93	0.70	0.00
X_{28}	Working capital/Total assets	WC/TA	0.13	0.00	0.00
X_{29}	Paid in capital/Shareholders' equity	PIC/SE	0.53	0.86	0.00
X_{30}	Sales/Working capital	S/WC	2.87	1.73	0.96
X_{31}	Retained earnings/Total assets	RE/TA*	0.08	-0.03	0.00
X_{32}	Net income/Shareholders' equity	NI/SE	0.42	-0.03	0.00
X_{33}	Net income/Sales	NI/S	0.16	-0.02	0.00
X_{34}	Net income/Total assets	NI/TA*	0.13	0.00	0.00
X_{35}	Sales/Current assets	S/CA	1.34	1.07	0.00
X_{36}	Operational income/Sales	OI/S*	0.20	0.06	0.00
X_{37}	Operational income/Total assets	OI/TA	0.17	0.03	0.00
X_{38}	Earnings before interest & taxes/ Interest expenses	EBIT/IE	-5.21	-0.45	0.05
X_{39}	Earnings before interest & taxes/Sales	EBIT/S	0.52	0.10	0.00
X_{40}	Gross profit /Sales	GP/S	0.27	0.15	0.00
X_{41}	Sales/Shareholders' equity	S/SE	3.32	4.68	0.05
X_{42}	Sales/Fixed assets	S/FA	6.29	6.44	0.33

*Final variables selected by SDA.
Group 1: going concern company & Group 2: non-going concern company

Table 2
Selected variables in SDA Analysis

Step		Tolerance	F to Remove	Wilks' Lambda
1	Net income to total assets	1.00	100.77	
2	Net income to total assets	0.94	56.24	0.75
	Total liabilities to total assets	0.94	9.07	0.55
3	Net income to total assets	0.51	8.62	0.52

	Total liabilities to total assets	0.91	11.10	0.53
	Operational income to sales	0.55	6.11	0.51
4	Net income to total assets	0.48	4.75	0.49
	Total liabilities to total assets	0.90	8.55	0.50
	Operational income to sales	0.54	4.57	0.49
	Retained earnings to total assets	0.77	4.37	0.49

Table 3

Pedictive accuracies(%) for training and hold-out data

fold	Training data	Hold-out data
1	100.00	100.00
2	100.00	93.33
3	100.00	100.00
4	100.00	86.67
5	100.00	93.33
6	100.00	92.86
7	100.00	92.86
8	100.00	100.00
9	100.00	100.00
10	99.24	92.86
Min	99.24	86.67
Max	100.00	100.00
Median	100.00	100.00
Variance	0.06	20.93
Mean	99.92	95.19

Table 4

The detailed results obtained by ANFIS via 10-fold cross- validation.

Fold	Accuracy (%)		Type I error(%)		Type II error(%)	
	Hold-out data	Training data	Hold-out data	Training data	Hold-out data	Training data
1	100.00	100.00	0.00	0.00	0.00	0.00
2	93.33	100.00	11.11	0.00	0.00	0.00
3	100.00	100.00	0.00	0.00	0.00	0.00
4	86.67	100.00	25.00	0.00	0.00	0.00
5	93.33	100.00	0.00	0.00	12.50	0.00
6	92.86	100.00	12.50	0.00	0.00	0.00
7	92.86	100.00	20.00	0.00	0.00	0.00
8	100.00	100.00	0.00	0.00	0.00	0.00
9	100.00	100.00	0.00	0.00	0.00	0.00
10	92.86	99.24	14.28	1.54	0.00	0.00
Average	95.19	99.92	8.29	0.15	1.25	0.00

References

- Alpaydin, E. (2010). *Introduction to machine learning,* 2'nd ed.,Cambridge, Mass., MIT Press.
- Ashoori, S., Mohammadi, S. (2011). Compare failure prediction models based on feature selection technique: empirical case from Iran. *Procedia Computer Science*, 3, 568–573.
- Guyon, I ., &Elisseeff,A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research, 3,* 1157–1182.
- Jang, J. S. R. (1993). ANFIS: adaptive-network-based fuzzy inference system, IEEE Transactions on Systems. *Man and Cybernetics, 23 (3),* 665–685

- Jang, J. S. R., & Sun, C.T. (1997). Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence. *Prentice Hall, Englewood Cliffs, NJ.*
- Jones, F. L. (1987). Current techniques in bankruptcy prediction. *Journal of Accounting Literature, 6,* 131–164.
- Martens. D., Bruynseels. L., Willekens. B. M., & Vanthienen, J. (2008). Predicting going concern opinion with data mining. *Decision Support Systems*, 45, 765 – 777.
- Sun, J.,&Li, H. (2008). Data mining method for listed companies' financial distress prediction. *Knowledge-Based Systems, 21,* 1-5