Cost-sensitive back-propagation neural network for financial distress prediction

Wang Hongbao*

College of Applied Foreign Language, Heilongjiang University, No. 74 Xuefu Road, Harbin, Heilongjiang Province, China, 150080

Received 27 January 2014, www.tsi.lv

Abstract

Financial distress prediction (FDP) models, which classify financially distressed companies from healthy ones, prevent market participants from suffering economic loss. In the process of FDP, the misclassification of type I error of the model incurs much higher cost than that of type II error. Most of the previous FPD models do not take the asymmetric costs into consideration. In this paper, cost-sensitive back-propagation neural network (CS-BPNN) FDP model is proposed for minimizing the cost of prediction error such that the loss of users of the model will suffer less. The performance of the model is evaluated by taking 180 Chinese listed companies as sample data and adopting 8 times of sampling to assess different misclassification costs and prediction accuracy. The experimental results suggest that the proposed approach helps to improve the prediction performance in asymmetric cost setup.

Keywords: Financial Distress Prediction (FDP), Cost Sensitive Learning, Back-propagation Neural Network (BPNN), Cost of Prediction Error, Cost-sensitive Back-propagation Neural Network (CS-BPNN)

1 Introduction

As the internal and external environment of a company changes in speed and complexity, business organization with management deficiency and lack of innovation may be very likely to lead to financial distress and even bankruptcy. Financial distress prediction (FDP) has a key influence on the enterprise's development and its stakeholders' decision as well [1]. For a commercial bank, FDP has profound impact on its credit scoring because banks should watch the current and future financial status of their enterprise customers all the time. For shareholders, FDP facilitates to detect the financially distressed condition of a company in advance so that they will withdraw capital before suffering huge economic loss [2]. Therefore, FDP has been a major research area within corporate finance for decades.

The earliest popular techniques in FDP were the statistical models, such as univariate analysis [3], multivariate discriminant analysis (MDA) [4] and logistic regression (Logit) and etc. [5]. Since 1990s, artificial intelligent and data mining techniques took a key role in FDP with the rapid development of computer technology [6]. Neural network (NN) has become one of the most widely used machine learning techniques in FDP due to its strong nonlinear mapping ability. Many researchers compared NN models with MDA and Logit and concluded that the prediction accuracy of NN models was higher than MDA and Logit [7]. Apart from NN, other artificial intelligent techniques were also employed in FDP, such as decision tree [8], genetic algorithm [9], rough sets and etc. [10]. Support vector machine (SVM) is a relatively new machine learning technique and is widely applied in many fields, such as classification, data mining and time series forecasting [11]. SVM is superior to other algorithms for FDP in situations where the variables demonstrate complex nonlinear relationships. However, it still has problems with identifying the relative importance of variables and searching the optimal parameters [12].

These previous classification techniques aim to minimize overall error based on the consumption that the misclassification costs of type I error and type II error are equal. However, this assumption is not valid in FDP, where the cost of misclassifying a distressed company as a healthy one is much higher than the inverse. Therefore, the asymmetric cost information should be taken into consideration in FDP so that different stakeholders could select their favourable models based on their cost preference. Ref [13] incorporated cost information into learning vector quantization (LVQ) approaches in FDP. To our knowledge, the study of cost-sensitive back-propagation neural network (CS-BPNN) is not reported for FDP.

As a result, the main motivation of this paper is to employ CS-BPNN to establish companies' FDP model in order to minimize the cost of prediction error. The main objectives of this paper are to (1) incorporate cost information into BPNN algorithm to make the traditional BPNN cost-sensitive in FDP, (2) exclude the missing and outlier data in the initial data pre-processing stage and use statistical methods to screen financial ratios in order to improve the prediction accuracy of FDP model, (3) compare CS-BPNN approach with BPNN approach in the

* Corresponding author- Tel: +86-139-3650-1851; fax: +86-139-3650-1851;
E-mail: Lodgewang@163.com

aspect of cost of misclassification and prediction accuracy, and (4) expand CS-BPNN approach so that it will provide decision makers with evidence in model selection.

2 Research Background

2.1 NEURAL NETWORK

Among different neural network architectures, BPNN is the most frequently employed architecture due to its simplicity and excellent performance in extracting useful information from samples [14]. P_R denotes the elementary inputs of BP, as shown in Fig. 1. W_i is an appropriate weight of each input. The sum of the weighted inputs and the bias is input to the f function, which transforms the sum of input value into output value of the node.

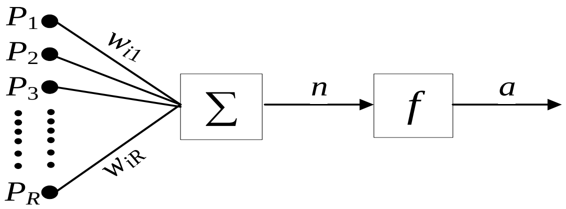

FIGURE 1 Artificial neuron model

2.2 COST-SENSITIVE BPNN

Cost-sensitive learning solves the problem in which different misclassification errors correspond with different costs. The aim of our paper is to make the misclassification cost minimum. If i equals to j, the misclassification cost of sample x classified as j is zero, shown as Equation 1:

$$C(i\mid x)=\sum_{j=1}^{c}C(i,j)P(j\mid x), \tag{1}$$

where c is the misclassification cost, i and j are the two classification class and x is a sample. If i does not equal to j, the misclassification cost of sample x classified as j is c.

In order to make c minimum, the value can be obtained by the following equation:

$$C(i|x)=\min_{i\in c}\arg\sum_{i\in 1}^{c}C(i,j)P(j|x), \tag{2}$$

The purpose of training BPNN is to make mean square deviation least:

$$f(x)=E\left[(t-a)^2\right]=\sum_{i=1}^{N}p_i(t_i-a_i)^2, \tag{3}$$

Cost-sensitive BPNN takes the misclassification minimum as the evaluation indicator of a model performance, as shown in Equation 4:

$$F(x)=E[C(i,j)]=\sum p(i)p(j|i)C(j,i), \tag{4}$$

In FDP, the output layer has only one neuron, and the misclassification cost of the sample x is shown in the Equation 5:

$$\mu(x)=\frac{1}{2}(t_k-a_k)^2C(t_k,|a_k|), \tag{5}$$

3 Empirical Experiment

3.1 EXPERIMENT DATA

3.1.1 Initial data collection and pre-processing

Because the empirical research is carried out on real world information of Chinese listed companies, financially distressed companies are defined as those who have had negative net profit in consecutive two years, or its net capital per share is lower than the face value per share for the reason of one year's substantive loss. Therefore, the financially distressed companies are specially treated (ST) by Chinese Stock Exchange. Healthy companies are defined as those who have never been specially treated.

Since the financial ratios of ST companies have been deteriorating even before ST, the adoption of financial data one year before ST results in overestimation of the prediction performance of the model. In this paper, the financial data from two years to five years before ST is selected, namely $U_{(t-2)}$, $U_{(t-3)}$, $U_{(t-4)}$ and $U_{(t-5)}$. The data used in this research is obtained from RESSET Financial Database. 90 pairs of companies listed in Shenzhen Stock Exchange and Shanghai Stock Exchange are selected as initial data set. 30 financial ratios are selected as initial features, covering debt ability, growth ability, capital structure, activity ability, profitability and indicators per share. In order to eliminate outlier data and missing data, companies with financial ratios deviating from the mean value as much as four times of standard deviation are excluded and companies missing at least one financial ratio are also excluded. The final number of sample data is 170.

3.1.2 Experimental data sets

The empirical experiment aims to validate whether FDP model based on CS-BPNN can minimize the cost of prediction error. 57 pairs of financially distressed companies and healthy companies are selected to form training data set and the rest 28 pairs are used to form testing data set.

3.1.3 Feature selection

In the field of FDP, a large number of financial indicators are usually involved in order to obtain an accurate financial condition of companies. However, some financial indicators cannot precisely identify financially distressed companies from healthy ones. Therefore, the purpose of feature selection addresses the problem by removing irrelevant and redundant features, improving the accuracy of the model, decreasing the computational effort and facilitating the use of the model.

(1) Statistical analysis

In this paper, statistical methods are used to screen financial ratios. The selection procedure is as follows: Kolmogrov-Simironov test is employed to examine whether each financial ratio meets normal distribution. If financial ratios meet normal distribution, T test is employed to validate whether the financial ratios are significant. If financial ratios do not meet normal distribution, Mann-Whitney test is employed to validate whether the financial ratios are significant, as shown in Table 1.

TABLE 1 Significance test of financial ratios in year t-2

Variables	T test		Kolmogrov-Simironov test		Mann-Whitney test	
	T-Statistic	*Prob.*	*KS-Statistic*	*Prob.*	*MW-Statistic*	*Prob.*
Return on Equity			5.365	**0.000	1263.000	**0.000
Return on Assets			2.323	**0.000	2096.000	**0.000
Return on Invested Capital			3.101	**0.002	1744.000	**0.000
Net Profit Margin			6.259	**0.000	1874.000	**0.000
Cost Profit Margin			2.752	**0.000	1797.000	**0.000
Current Ratio			2.420	**0.000	1086.000	**0.000
Quick Ratio			2.862	**0.000	1233.000	**0.000
Equity Ratio			5.055	**0.000	1120.000	**0.000
Debt to Asset Ratio			5.226	**0.000	1210.000	**0.000
Debt to Tangible Asset Ratio			4.602	**0.005	3287.000	0.310
Operating Cash Flow/Total Liability			6.785	**0.002	2146.000	**0.000
Operating Income Growth Rate			5.614	**0.000	2786.000	*0.010
Net Profit Growth Rate			4.696	**0.000	3072.000	0.092
Total Asset Growth Rate			4.209	**0.001	1495.000	**0.000
Turnover Rate of Accounts Receivable			5.532	**0.000	3508.000	0.745
Turnover Rate of Accounts Payable			2.723	**0.000	2435.000	**0.000
Turnover Rate of Current Assets			1.872	**0.002	3392.000	0.492
Turnover Rate of Fixed Assets			5.852	**0.000	2003.000	**0.000
Turnover Rate of Equity			4.419	**0.000	2637.000	**0.002
Turnover Rate of Total Assets			1.559	*0.015	2088.500	**0.000
Earning Per Share			2.708	**0.000	1578.500	**0.000
Net Asset Value Per Share	7.666	**0.000	1.181	0.123		
Operating Revenue Per Share			2.355	**0.000	1114.000	**0.000
Gross Profit Margin			1.529	*0.019	2485.000	**0.000
Net Return on Assets			3.634	**0.000	1868.000	**0.000
Fixed Assets Ratio	-2.760	**0.007	1.075	0.198		
Equity Ratio			5.226	**0.000	1210.000	**0.000
Operating Profit Growth			5.807	**0.000	2898.000	**0.026
Earning Per Share Growth			4.611	**0.000	3114.500	0.121
Every Dividend Profit before Tax			2.604	**0.000	1698.000	**0.000

Note: *Significant at 5%; **Significant at 1%.

(2) Analysis on significance test of financial ratios

As shown in table 1, only net asset value per share and fixed assets ratio pass Kolmogrov-Simironov test. The result is consistent with the previous research conclusion that most financial ratios do not meet normal distribution. Additionally debt to tangible asset ratio, operating cash flow/total liability, net profit growth rate, total asset growth rate, turnover rate of accounts receivable, turnover rate of current assets, turnover rate of equity, operating revenue per share, gross profit margin, net return on assets, fixed assets ratio, equity ratio, operating profit growth, earning per share growth and every dividend profit before tax do not pass the significance test from year *t*-2 to year *t*-5. Therefore, these 15 ratios are discarded, by which healthy companies cannot be distinguished from distressed companies.

3.2 PARAMETER SETTING

3.2.1 Setting of cost of misclassification

When misclassification occurs in FDP model, the costs of misclassification of different stakeholders are different. For example, company shareholders' loss is hugely different from managers' one when FDP model has type I error. Therefore, the user of the FDP model should be identified in the first place. Then cost matrix is used to determine the value of cost.

In the empirical research, the user of FDP model is supposed to be a commercial bank. The bank can use the prediction result of FDP model to make a decision of making loans or not. Therefore, when the model has type

I error, the cost of error of the bank is the full loan. When the model has type II error, the cost is the loan interest loss. Since loan interest rates of a commercial bank to an enterprise in China range from 5% to 30% above personal loan interest rate, 30% above personal loan interest rate is taken into calculation for the sake of unified computing.

3.2.2 Setting of parameters of CS-BPNN

Classification performance of CS-BPNN is affected mainly by the parameter pair, learning rate α and number of training round *g*, whose optimal parameter combination is determined by grid method and leave-one-out cross validation test. The combination, which has the least misclassification cost in the training set data, is selected as the parameters of CS-BPNN, where $\alpha \in \{10^{-3}, 10^{-2}, 10^{-1}, 3\times10^{-1}\}$ and $g \in$ {100, 500, 1000, 3000, 5000, 10000}. As shown in table 2-5, the optimal parameter value in year *t*-2 is $\alpha_{(t-2)} = 0.1$, $g_{(t-2)} = 10000$. The optimal parameter value in year *t*-3 is $\alpha_{(t-3)} = 0.3$, $g_{(t-3)} = 5000$. The optimal parameter value in year *t*-4 is $\alpha_{(t-4)} = 0.3$, $g_{(t-4)} = 5000$. The optimal parameter value in year *t*-5 is $\alpha_{(t-5)} = 0.3$, $g_{(t-5)} = 3000$.

TABLE 2 Leave-one-out cross validation test result of cost-sensitive BP neural network in year t-2

		g=100	g=500	g=1000	g=3000	g=5000	g=10000
	Type I error	0	0	0	0	0	0
a=0.001	Type II error	60	60	60	60	60	59
	Total error	5.4	5.4	5.4	5.4	5.4	4.86
	Type I error	0	0	0	2	3	6
a=0.01	Type II error	60	60	59	47	41	36
	Total error	5.4	5.4	4.86	6.23	6.69	9.24
	Type I error	0	2	3	3	3	2
a=0.1	Type II error	59	43	38	28	24	20
	Total error	4.86	5.87	6.42	5.52	5.16	3.8
	Type I error	2	4	3	3	2	3
a=0.3	Type II error	49	34	34	20	36	22
	Total error	6.41	7.06	6.06	4.8	5.24	4.98

TABLE 3 Leave-one-out cross validation test result of cost-sensitive BP neural network in year t-3

		g=100	g=500	g=1000	g=3000	g=5000	g=10000
	Type I error	0	0	0	0	0	0
a=0.001	Type II error	60	60	60	60	60	60
	Total error	5.4	5.4	5.4	5.4	5.4	5.4
	Type I error	0	0	0	2	3	4
a=0.01	Type II error	60	60	60	50	44	37
	Total error	5.4	5.4	5.4	6.5	6.96	7.33
	Type I error	0	3	4	3	2	4
a=0.1	Type II error	60	44	38	35	34	25
	Total error	5.4	6.96	7.42	6.15	5.06	6.25
	Type I error	2	3	3	2	2	3
a=0.3	Type II error	51	39	36	35	25	25
	Total error	6.59	6.51	6.24	5.15	4.25	5.25

TABLE 4 Leave-one-out cross validation test result of cost-sensitive BP neural network in year t-4

		g=100	g=500	g=1000	g=3000	g=5000	g=10000
	Type I error	0	0	0	0	0	0
a=0.001	Type II error	60	60	60	60	60	60
	Total error	5.4	5.4	5.4	5.4	5.4	5.4
	Type I error	0	0	0	1	2	3
a=0.01	Type II error	60	60	60	55	47	45
	Total error	5.4	5.4	5.4	5.95	6.23	7.05
	Type I error	0	2	3	2	4	4
a=0.1	Type II error	60	47	46	43	38	32
	Total error	5.4	6.23	7.14	5.87	7.42	6.88
	Type I error	1	1	3	2	3	4
a=0.3	Type II error	45	45	43	33	35	32
	Total error	5.05	5.05	6.87	4.97	6.15	6.88

TABLE 5 Leave-one-out cross validation test result of cost-sensitive BP neural network in year t-5

		g=100	**g=500**	**g=1000**	**g=3000**	**g=5000**	**g=10000**
	Type I error	0	0	0	0	0	0
a=0.001	Type II error	60	60	60	60	60	60
	Total error	5.4	5.4	5.4	5.4	5.4	5.4
	Type I error	0	0	0	2	4	5
a=0.01	Type II error	60	60	60	59	57	50
	Total error	5.4	5.4	5.4	7.31	9.13	9.5
	Type I error	0	4	4	3	4	5
a=0.1	Type II error	60	57	53	36	31	28
	Total error	5.4	9.13	8.77	6.24	6.79	7.52
	Type I error	2	3	2	2	5	2
a=0.3	Type II error	59	46	40	33	34	34
	Total error	7.31	7.14	5.6	4.97	8.06	5.06

3.3 EXPERIMENT RESULTS AND ANALYSIS

In order to verify the prediction performance of FDP model based on CS-BPNN, the empirical research makes a comparison with the one based on BPNN. The optimal parameter combination of FDP model based on BPNN is also determined by grid method and leave-one-out cross validation test. In order to get multiple performance statistics, multiple experimental data sets are formed by repetitively and randomly classifying training sample and testing sample. By 8 times of random sampling without replacement, 57 pairs of financial distressed companies and healthy ones are selected as training data set and the rest 28 pairs are selected as testing data set each time, as shown in table 6.

TABLE 6 Experimental results on testing data set in year t-2

Data	**BPNN**				**CS-BPNN**			
	Number of type I error	*Number of type II error*	*Prediction accuracy*	*Total cost of prediction error*	*Number of type I error*	*Number of type II error*	*Prediction accuracy*	*Total cost of prediction error*
set 1	4	5	83.93	4.45	2	8	82.14	2.72
set 2	1	2	94.64	1.18	1	2	94.64	1.18
set 3	6	4	82.14	6.36	2	9	80.36	2.81
set 4	4	4	85.71	4.36	4	9	76.79	4.81
set 5	4	1	91.07	4.09	2	3	91.07	2.27
set 6	2	1	94.64	2.09	1	2	94.64	1.18
set 7	8	4	78.57	8.36	4	9	76.79	4.81
set 8	6	3	83.93	6.27	4	4	85.71	4.36
Average	4.38	3	86.83	4.65	2.5	5.75	85.27	3.02

TABLE 7 Comparison of test results of BPNN and CS-BPNN from year t-2 to year t-5

Year	FDP models	BPNN	CS-BPNN	Year	FDP models	BPNN	CS-BPNN
	Average Number of type I error	4.38	2.5		Average Number of type I error	5.31	2.1
t-2	Average Number of type II error	3	5.75	t-3	Average Number of type II error	2.45	8.6
	Total cost	4.65	3.02		Total cost	5.53	2.87
	Average Number of type I error	7	4.2		Average Number of type I error	6.12	2
t-4	Average Number of type II error	4.33	9.36	t-5	Average Number of type II error	10.65	15.5
	Total cost	7.39	5.04		Total cost	7.08	3.4

As shown in table 6, the average prediction accuracy of FDP model based on BPNN is 86.83, which is slightly higher than that of FDP model based on CS-BPNN, 85.27. It is mainly because FDP model based on CS-BPNN integrates the different costs of prediction errors. Since the cost of type I error is hugely larger than that of type II error, the average cost of prediction error of FDP model based on BPNN is 4.65, which is much higher than that based on CS-BPNN, 3.02.

In table 7, the empirical results suggest that the prediction performance of both BPNN and CS-BPNN become weaker with the selection of earlier training data sets. However, we do not draw the same conclusion in the misclassification cost. For example, the misclassification cost of CS-BPNN in year t-5 is lower than in year t-4. The main reason is that even though the prediction accuracy of CS-BPNN in year t-4 is higher than t-5, the number of type I error in year t-4 is more than the one in year t-5.

The experimental results and analysis suggest that BPNN achieves slightly better prediction accuracy than CS-BPNN. However, CS-BPNN produces a much better result than BPNN in the total misclassification costs. With the selection of earlier training data sets, BPNN and CS-BPNN become weaker in the prediction accuracy, but they do not perform the same way in the misclassification cost.

4 Conclusion

Financial distress prediction is extensively studied in the corporate governance field. Few studies incorporate unequal misclassification costs into FDP model. Cost-sensitive classification models, coping with asymmetric costs of type I error and type II error, are of crucial interest to stakeholders' decisions. This paper verifies how the asymmetric costs of two kinds of errors are integrated into FDP model. This research takes 85 financial healthy companies and matches them with 85 financially distressed companies. 114 companies are selected as training data set and the rest are selected as testing data set. Experimental tests demonstrate that CS-BPNN approach leads to a lower total misclassification cost when compared with the traditional BPNN one.

Acknowledgments

In the process of writing paper, thanks so much for the help and understanding of my wife and my daughter.

References

[1] Zhi X, Xianglei Y, Ying P, Xin D 2012 The prediction for listed companies' financial distress by using multiple prediction methods with rough set and Dempster–Shafer evidence theory *Knowledge-Based Systems* **26** 196-206

[2] Pochang K, Pingchen L 2006 An evolution-based approach with modularized evaluations to forecast financial distress *Knowledge-Based Systems* **19**(1) 84-91

[3] Fitzpartrick P J 1932 A comparison of ratios of successful industrial enterprises with those of failed Finns *Certified Public Accountant* **10**(2) 598-605

[4] Altman E I 1968 Financial ratios discriminant analysis and the prediction of corporate bankruptcy *Journal of Finance* **23**(4) 589–609

[5] Ohlson J A 1980 Financial ratios and the probabilistic prediction of bankruptcy *Journal of Accounting Research* **18**(1) 109-31

[6] Chaudhuri A, De K 2011 Fuzzy support vector machine for bankruptcy prediction *Applied Soft Computing* **11**(2) 2472-86

[7] Odom M, Sharda R 1990 A neural networks model for bankruptcy prediction *IEEE International Conference on Neural Network* (San Diego) **2** 163-8

[8] Frydman H, Altman E I, Kao D 1985 Introducing recursive partitioning for financial classification: the case of financial distress *Journal of Finance* **40**(1) 269-91

[9] Lensberg T, Eilifsen A, McKee T E 2006 Bankruptcy theory development and classification via genetic programming *European Journal of Operational Research* **169**(2) 677-97

[10] McKee T E 2000 Developing a bankruptcy prediction model via rough sets theory *International Journal of Intelligent Systems in Accounting, Finance and Management* **9**(3) 159-73

[11] Haibin. X, Muhua L, Haichao Y, Qian H, Jinhui Z 2013 Study on detection and classification of Tetracycline residue in duck meat using synchronous fluorescence spectra and support vector machine *Journal of Engineering Science and Technology Review* **6**(2) 85-9

[12] Vapnik V N 1955 *The Nature of Statistical Learning Theory* Springer-Verlag: New York

[13] Ning C, Bernardete R, Armando S V, Joao D, Joao C N 2010 Hybrid genetic algorithm and learning vector quantization modelling for cost-sensitive bankruptcy prediction *2nd International Conference on Machine Learning and Computing* (Bangalore) 213-7

[14] Shujiang F, Yi Z, Honghua X, Chuang L 2012 Risk evaluation of rural financial organization operation based on the BP *Journal of Digital Information Management* **10**(5) 341-6

Author

Hongbao Wang, born in April, 1979, Harbin City, Heilongjiang Province, P.R. China

Current position, grades: Lecturer, College of Applied Foreign Language, Heilongjiang University, Harbin, China.
University studies: received his M.Sc. in Investment Analysis from Stirling University in UK. He received his Ph.D. in Corporate Governance from Harbin Institute of Technology in China.
Scientific interest: His research interest fields include company financial distress prediction and data mining.
Publications: more than 10 papers published in various journals.
Experience: He has been teaching for 9 years and has completed three scientific research projects.