Dynamic Early-Warning of Enterprise Financial Distress Based on Gradient Boosting Algorithm

Ying Peng*, Ziyi Chen and Jingyi Wang
Business School, Jianghan University, No.8, Sanjiaohu Road, Wuhan Economic and Technological Development Zone, China

Keywords: Enterprise Financial Distress, Gradient Boosting Algorithm, Dynamic Early-Warning.

Abstract: One of the biggest problems of users of financial statements is whether the enterprise will face financial distress. In this study, an early-warning system model based on gradient boosting algorithm for enterprise dynamic early-warning is presented. Sometimes special treatment (ST) is the warning of abnormal financial or occurring other conditions in China stock exchange. We construct enterprise dynamic early-warning model based on gradient boosting algorithm using the data of ST companies and their matching companies before special treatment 3 years. Our model calculates the relative variable importance (RVI) of each financial distress indicators, and get the average results of models. Through comparing with logit model, the results show that model based on gradient boosting algorithm can get better warning results. Our paper provides a more accurate method for enterprise dynamic early-warning, which can provide reference for users of financial statements improve financial situation, change investment strategy and so on.

1 INTRODUCTION

An enterprise encounter financial distress is a gradual process, not sudden. Before facing financial distress, financial or non-financial indicators of enterprise may appear abnormal. That is to say, we can find indicators and use method to alert for the probability of financial distress. Therefore, the key to early-warning of enterprise financial distress is to establish early-warning indicator system and find out applicability algorithm.

About the early-warning indicator system, it has experienced two stages. The first stage, indicators are instructed based on financial statements; the second stage, indicators are selected based on other information that is also important for enterprises, such as marketing indicators, corporate governance indicators, and so on.

Enterprise financial distress early-warning models can be divided into statistical methods and machine learning methods (Alaka, Oyedele, .et al, 2018). Statistical methods have been introduced into financial early-warning about 60 years ago. They include z-score model, single and multiple discriminant model, logit and probit models, and so on (Altman, 1968; Deakin, 1972; Jones, 1987). Machine learning methods first come into financial distress early-warning in 1990s, and there are some breakthroughs have been made in the financial distress application research area. Such as genetic algorithm, BP neural network, rand forest algorithm, and so on (Brockeet and Cooper, 1995; Sharda and Steiger, 1990; Breiman, 2001; Franco, 2002).

Through machine learning methods can improve accuracy of financial distress early-warning, in spite of the process of early warning seems in a “black box” (Barboza and Kimura, 2017). Therefore, they cannot provide suggestions on how to improve performance of enterprises. However, gradient boosting algorithm as an improved machine learning models, which can overcome the defect of the “black box” problem. It can not only output specific alert results, but also output relative variable importance of indicators, which can help users of financial statements make decisions. In this study, we introduce gradient boosting algorithm into the field of financial early-warning field, which can further expand the application scope of the method.

Peng, Y., Chen, Z. and Wang, J.
Dynamic Early-Warning of Enterprise Financial Distress Based on Gradient Boosting Algorithm.
DOI: 10.5220/0012022800003620
In *Proceedings of the 4th International Conference on Economic Management and Model Engineering (ICEMME 2022)*, pages 17-23
ISBN: 978-989-758-636-1

2 CONSTRUCTION OF FINANCIAL DISTRESS EARLY-WARNING INDICATORS SYSTEM

Gradient boosting algorithm is not affected by collinearity and missing value of indicators, and with the effect of alert is not monotone decreasing. Therefore, we construct financial distress early-warning indicators system including both financial indicators and non-financial indicators.

2.1 Construction of Financial Indicators System

There is no doubt that traditional financial indicators still play an important role in the field of early-warning of enterprise financial distress. They usually include solvency indicators, profitability indicators, operating capacity indicators and development capacity indicators. We can use solvency indicators to reflect mismatch of assets and debt, as the mismatch will cause such as non-effective investment, much bigger operational risk, which influence negatively. The decline of profitability is one of the important manifestations of financial distress. Operating capacity can reflect cash that are occupied by suppliers and customers. If there are too much cash are occupied, that will cause companies lack of cash for their expanded. Development capacity can reflect the growth rate of one company, but the growth rate too fast or too late all influences the happening of financial distress. However, cash flow and the degree of earnings management also occupy a key position in determine whether one enterprise will go bankrupt or not. Therefore, we add cash-flow indicators and earnings management indicators in traditional financial indicators to improve early-warning effect.

Table 1: Financial indicators system.

Classification of Indicators	Financial Indicators
Solvency Indicators	Current ratio (x_{11})
	Quick ratio (x_{12})
	Cash ratio (x_{13})
	Equity to assets ratio (x_{14})
	Working capital to debt (x_{15})
	Long term liabilities to total assets (x_{16})
	Fixed assets to total assets (x_{17})
	Intangible assets to total assets (x_{18})
Profitability Indicators	Net profit to total assets (x_{21})
	Net profit to capital (x_{22})
	Profit before interest and tax to profit before tax + financing expenses (x_{23})
	Gross profit to sales (x_{24})
	Operating profit to sales (x_{25})
Operating Capacity Indicators	Receivables turnover (x_{31})
	Inventory turnover (x_{32})
	Total assets turnover (x_{33})
	Working capital turnover (x_{34})
Development Capacity Indicators	Cumulative capital ratio (x_{41})
	Earnings per share growth rate (x_{42})
	Net profit growth rate (x_{43})
	Self-sustainable growth rate (x_{44})
	Income growth rate (x_{45})
	Total asset growth rate (x_{46})
	Fixed assets growth rate (x_{47})
	Intangible assets growth rate (x_{48})
Cash-flow Indicators	Free cash flow (x_{51})
	Cash flow interest coverage ratio (x_{52})
	Cash meet investment ratio (x_{53})
	Cash to operating income (x_{54})
	Cash to net profit (x_{55})
Earnings Management Indicators	Accrued earnings management degree (x_{61})
	Real earnings management degree (x_{62})
	Absolute value of accrued earnings management (x_{63})
	Absolute value of real earnings management (x_{64})

2.2 Construction of Non-Financial Indicators System

As the traditional financial indicators have many defects, which can only reflect historical information and exist hysteresis (Brochet, Loumioti, 2015; Kraft, Vashishtha,2018). However, non-financial indicators that have good forward-looking and value relevance for users of financial statements. They effectively make up the shortage of traditional financial indicators. Therefore, we construct non-financial indicators system, which include marketing indicators, corporate governance indicators and auditors' behaviour indicators. According to signal transmission theory, if one company suffer financial distress, which transfer negative signal to market, and then it will cause negative market response. Corporate governance is the specific of enterprise's internal stable environment. Good corporate governance will effectively decline agent cost, but bad corporate governance may cause financial distress. Sometimes auditors cannot directly decide one company financial risk, whose behaviour can tell us whether the corporate suffer financial distress, such as raising audit fees, increasing audit delay, issuing nonstandard audit opinion, and so on.

Table 2: Non-Financial indicators system.

Classification of Indicators	Non-Financial Indicators
Marketing Indicators	Price earnings ratio (x_{71})
	Price to sales (x_{72})
	Price to book (x_{73})
	Dividend declared ratio (x_{74})
	Earnings per share (x_{75})
	Net asset per share (x_{76})
Corporate Governance Indicators	Director number (x_{81})
	Institutional investors shareholding ratio (x_{82})
	Equity concentration (x_{83})
Auditors' Behaviour Indicators	Abnormal audit fees (x_{91})
	Audit delay (x_{92})
	Nonstandard audit opinion (x_{93})
	Auditors change (x_{94})

3 ESTABLISHMENT OF EARL-WARNING MODEL AND CALCULATION STEPS

3.1 Establishment of Enterprise Financial Distress Model Based on Gradient Boosting Algorithm

Gradient boosting algorithm is an ensemble learning algorithm that can combine a series of weak classifiers into a strong classifier. Traditional machine learning can only establish one learning model, but ensemble learning can establish a series learning models and can combine all learning models together format a committee-based learning model. Therefore, weaker classifiers can become strong classifier. We can use training data as experience knowledge to establish learning model, which can learn the relationship from input to output. And then we can use the learning relationship on test data.

As we all known, no matter financial indicators or non-financial indicators can alert whether a listed company happens financial distress or not. But sometimes the effective of single indicator to do early-warning is always bad, so we can define the single indicator as a weak classifier. Therefore, we design both financial indicators system and non-financial indicators system, which can seem as a strong classifier. We use the classifier to predict whether a company facing financial distress.

Sometimes we cannot obtain the intuitiveness and accuracy at the same time in one model. Machine learning algorithm can enhance the effective of early learning, but it cannot tell us how to get the results. However, using gradient boosting algorithm to do the dynamic prediction process through calculating indicators relative variable importance (RVI) and the effectiveness indicators of early warning model. That is to say, gradient boosting algorithm can get intuitiveness and accuracy at the same time to some extent. The calculations of RVI can be given in the form:

$$\tau_{\vartheta}^{2}(T)=\sum_{t=1}^{J-1} i_{t}^{2} I[V(T)=\vartheta] \tag{1}$$

Where $\tau_{\vartheta}^{2}(T)$ is the indicators relative variable importance (RVI); J-1 is the number of nodes in the decision tree; $i_{t}^{2} I[V(T)=\vartheta]$ it the classification error of indicators in i node. The bigger RVI, the better early-warning effectivity of indicators.

We use true positive rate (TPR), false positive rate (FPR), and accuracy rate (AR), recall rate (RR) and precision rate (PR) to calculate the effectiveness of early warning model. Accuracy rate can be used to measure accuracy of the model; recall rate can be used to calculate the probability of Type I error; and precision rate can be used to estimate the probability of Type II error. The formulas are shown as follow:

$$TPR=\frac{TP}{N} \tag{2}$$

$$FPR=\frac{FP}{N} \tag{3}$$

$$AR=\frac{TP+TN}{N} \tag{4}$$

$$RR=\frac{TP}{TP+FN} \tag{5}$$

$$PR=\frac{TP}{TP+FP} \tag{6}$$

Where TP is the ST companies' number which is correctly classified by model; FN is the ST companies' number which is wrongly classified by model; FP is the non-ST companies' number which is wrongly classified by model; TN is the non-ST companies' number which is correctly classified by model; N is the number of the whole sample.

3.2 Establishment of Enterprise Financial Distress Calculation Steps

The aim of gradient boosting algorithm is to train a strong classifier, which can improve the effect of early-warning of enterprise financial distress. The strong classifier is combine with many weak classifiers. Therefore, how to train a strong classifier is a problem. The steps are as follow:

Step 1: input training data set;

Step 2: assign the weight of sample point as 1/n;

Step 3: assume there are m financial distress indicators in total:

- ✓ Use the training set that was assigned weight, and get the basic classifier:

$$G_m(x): x \to \{-1,1\} \tag{7}$$

- ✓ Calculate the classification error rate of $G_m(x)$ in the training set:

$$e_m = P(G_m(x_i) \neq y_i) \tag{8}$$

- ✓ Calculate the coefficient of G_m:

$$a_m = \frac{1}{2}\log\frac{1-e_m}{e_m} \tag{9}$$

- ✓ Update the weight distribution of training data set:

$$D_{m+1} = (w_{m+1,1}, w_{m+1,2}, \cdots, w_{m+1,n}) \tag{10}$$

$$w_{m+1,i} = \frac{w_{m,i}}{Z_m}\exp(-a_m y_i G_m(x_i)) \tag{11}$$

$$Z_m = \sum_{i=1}^{n} w_{m,i}\exp(-a_m y_i G_m(x_i)) \tag{12}$$

Where D_{m+1} is the updated weight of training data set.

Step 4: get a linear combination of weak classifiers:

$$f(x) = \sum_{m=1}^{M} a_m G_m(x) \tag{13}$$

Repeat the above process, finally we can get the linear combination f(x) of weaker classifier, which is our strong classifier trained by gradient boosting algorithm. Using this strong classifier to judge whether the listed company will have financial crisis can greatly improve the warning effect.

4 AN ILLUSTRATIVE EXAMPLE

4.1 Data and Sample Selection

Our sample data includes all public listed companies in the Chinese market between 2007 and 2017. 2007 is the start year because it is the year that newly Chinese Accounting Standard issued. The relevant financial data of sample firms are collected from the China Stock Market and Accounting Research (CSMAR) database. In order to test the indicators' differences between ST companies and normal companies, we select out 404 publicly listed companies which are the first time to be ST. We removed financial industry samples and with a large number of missing data samples. Finally, it keeps 266 publicly listed companies that are the first time to be ST.

We paired matching companies of the first time to be ST companies at a ratio of 1:1. The principles for selecting matching samples are as follows: first, matching companies and ST companies are in the same industry; second, the matching companies always be listed throughout the sample range from 2007 to 2017; third, the matching companies have never been ST in the entire sample range; fourth, when it meet the above three conditions, selecting the companies that closest to the total assets of the ST company as the matching samples.

4.2 Descriptive Statistics

In order to analyse how to predict operational risks, we start with the descriptive analysis on showing the distribution characteristics of indicators, and the results are presented in Table 3.

Table 3: Descriptive statistics.

Index	Mean1	Med1	Mean2	Med2	Mean3	Med3
x11	1.06	0.78	1.22	0.93	1.42	1.10
x12	0.74	0.50	0.84	0.59	0.99	0.71
x13	0.26	0.13	0.28	0.14	0.38	0.16
x14	-3.61	1.96	2.31	1.60	1.68	1.30
x15	0.55	-0.37	0.71	-0.15	1.97	0.06
x16	0.08	0.02	0.08	0.03	0.07	0.03
x17	0.34	0.35	0.34	0.33	0.31	0.31
x18	0.06	0.04	0.05	0.03	0.05	0.03
x21	-0.13	-0.09	-0.07	-0.06	0.01	0.01
x22	-0.19	-0.10	-0.07	-0.05	0.03	0.03
x23	-0.66	-0.08	-0.25	-0.01	0.13	0.12
x24	0.08	0.07	0.95	1.00	0.67	0.78
x25	-0.86	-0.20	-0.40	-0.11	-0.02	0.01
x31	332.77	5.82	181.96	5.76	403.51	6.29
x32	103.17	3.73	10.71	3.83	26.37	4.13

x33	0.51	0.43	0.56	0.47	0.63	0.52
x34	0.91	-0.98	-13.76	-0.58	0.48	0.62
x41	-0.35	-0.25	-0.14	-0.12	-0.03	0.02
x42	0.21	0.40	-16.90	-7.31	-0.64	-0.76
x43	-0.12	0.32	-20.11	-6.35	-0.72	-0.65
x44	-0.34	-0.23	-0.16	-0.13	0.03	0.01
x45	-0.07	-0.07	0.13	-0.09	3.35	0.06
x46	-0.07	-0.07	0.00	-0.02	1.22	0.04
x47	-0.03	-0.06	0.26	-0.02	2.75	-0.01
x48	14.77	-0.03	0.74	-0.02	2.29	-0.02

Table 3: Descriptive statistics.

Index	Mean1	Med1	Mean1	Med2	Mean3	Med3
x51	4.1E+07	3.6E+6	1.1E+8	1.8E+7	-9.2E+7	8.3E+6
x52	261.19	0.75	3.57	0.73	-1.25	1.48
x53	0.24	0.25	0.28	0.25	0.68	0.24
x54	1.35	1.03	1.06	1.03	1.05	1.03
x55	-1.49	-0.11	-0.05	-0.09	7.54	1.55
x61	-0.09	-0.08	-0.05	-0.04	0.00	0.00
x62	0.06	0.06	0.08	0.07	0.06	0.05
x63	0.11	0.09	0.07	0.06	0.06	0.04
x64	0.12	0.09	0.13	0.10	0.14	0.10
x71	-60.42	-11.72	-33.73	-14.90	174.14	106.03
x72	19.38	3.02	15.00	2.10	7.88	1.99
x73	1.81	4.25	5.51	2.82	3.32	2.30
x74	0.00	0.00	-0.01	0.00	0.18	0.00
x75	-0.78	-0.50	-0.43	-0.33	0.09	0.04
x76	1.67	1.53	2.56	2.22	3.17	2.62
x81	8.82	9.00	9.13	9.00	9.22	9.00
x82	3.40	1.70	3.46	1.97	3.51	2.09
x83	43.37	41.83	45.22	43.69	47.45	46.29
x91	0.01	0.00	-0.04	-0.07	-0.02	-0.03
x92	98.50	108.00	101.26	107.00	92.02	98.00
x93	0.30	0.00	0.18	0.00	0.08	0.00
x94	0.26	0.00	0.19	0.00	0.19	0.00

Where Mean1 is the mean value before one year of companies' special treated year; Med1 is the median value before one year of companies' special treated year. From Table 3, it can be seen that the mean value is smaller than its median value in major development capacity indicators (x41, x42, …, x48). That is to say, companies went through negative growth. Accrued earnings management indicators are negative, but real earnings management indicators are positive. Because the cost of accrued earnings management is lower than real earnings management. At the beginning of enterprises suffer financial deteriorate, they tend to pay low cost to do earnings management. But they have to do the real earnings management to manipulate the surplus, which shows the two types of earnings management methods have a certain substitution effect.

About the value of all non-financial indicators are go bad from t-3 year to t-1 year. The increase value of abnormal audit fees, audit delay and nonstandard audit opinion with the time goes by, which means that auditors have the demand of reducing risk.

4.3 Analysis of Indicators Relative Variable Importance (RVI)

Relative variable importance (RVI) can provide a reference for executives to improve corporate performance and avoid financial distress. As we all known, financial distress is a gradual process, and different early warning indicators play different roles before companies' special treated 3 years. Therefore, we estimated average score of RVI respectively, and compared the average score of RVI in different years. The results of each classification of indicators are presented in Table 4.

From Table 4, we can get the contribution of each early-warning indicators through calculating average score of RVI. Among all early-warning indicators, profitability indicators contribute most in gradient boosting algorithm, which also play an important role in other financial early-warning models. However, development capacity indicators and marketing indicators have also made great contributions in early-warning indicators system, which usually ignored by a large number of studies.

Table 4: Average score of RVI.

Classification of Indicators	t-1	t-2	t-3
Solvency Indicators	5.03	9.65	15.32
Profitability Indicators	31.26	14.59	32.98
Operating Capacity Indicators	11.50	12.27	18.43
Development Capacity Indicators	23.14	22.22	31.65
Cash-flow Indicators	10.07	10.50	5.99
Earnings Management Indicators	1.51	4.03	8.58
Marketing Indicators	20.34	21.75	37.09
Corporate Governance Indicators	6.30	11.43	14.27
Auditors' Behaviour Indicators	3.05	4.85	4.85

Besides, the importance of indicators changes with the time goes by. The importance of some indicators reduced, such as solvency indicators, operating capacity indicators, earnings management indicators, marketing indicators and corporate governance indicators; however, the importance of other indicators enhanced, such as cash-flow indicators.

4.4 Empirical Results

(1) Average results based on gradient boosting algorithm

We get the results of early-warning from indicators, such as true positive rate, false positive rate, accuracy rate, recall rate and precision rate. From Table 5, we can get the average effectiveness of early warning model, which estimated by the percentage of training samples and test samples as 7:3, 8:2 and 9:1. It can be seen accuracy rate increased with the time near special treatment year.

Each recall rate is higher than accuracy rate and precision rate, that is to say, the probability of Type I error is smaller than Type II error in our model based on gradient boosting algorithm. As we all known, the cost of Type I error is higher than Type II error (Lian, 2017). Therefore, the dynamic early-warning model of enterprise financial distress can better identify enterprises from all sample companies, which can help investors, managers and other enterprise stakeholders to make decisions.

Table 5: Average results based on gradient boosting algorithm.

Indicators	t-1	t-2	t-3	Average
True Positive Rate	0.507	0.510	0.484	0.500
False Positive Rate	0.026	0.049	0.089	0.055
Accuracy Rate	0.900	0.885	0.756	0.847
Recall Rate	0.952	0.913	0.847	0.904
Precision Rate	0.872	0.886	0.758	0.839

(2) Average results based on logit

For further verification early-warning effectivity of gradient boosting algorithm, we construct a comparing model based on logit. From Table 6, we can see all the indicators results in logit model are lower than gradient boosting algorithm. The result shows that effectiveness of gradient boosting model significantly better than logit model. But also the RVI that reported by gradient boosting can provide suggestions for improving management.

Table 6: Average results based on logit.

Indicators	t-1	t-2	t-3	Average
True Positive Rate	0.431	0.427	0.433	0.430
False Positive Rate	0.154	0.176	0.189	0.173
Accuracy Rate	0.788	0.724	0.680	0.731
Recall Rate	0.797	0.741	0.691	0.743
Precision Rate	0.732	0.715	0.659	0.724

5 CONCLUSIONS

This paper constructs enterprise dynamic early-warning model based on gradient boosting algorithm using the data of ST companies and their matching companies before special treatment 3 years. The model calculates the relative variable importance (RVI) of each financial distress indicators, and get the average results of models. Through comparing with logit model, the results show that model based on gradient boosting algorithm can get better warning results. This study provides a more accurate method for enterprise dynamic early-warning, which can provide reference for users of financial statements improve financial situation, change investment strategy and so on.

ACKNOWLEDGEMENTS

Supported by the Discipline Group of "Urban Circle Economy and Industrial Integration Management" of Jianghan University.

REFERENCES

Alaka, H. A., L. O. Oyedele, H. A. Owolabi, V. Kumar, S. O. Ajayi, O. O. Akinade & M. Bilal (2018) Systematic review of bankruptcy prediction models. Expert Systems with Applications: An International Journal.

Altman, E. I. (1968) Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. The journal of finance, 23, 589-609.

Barboza, F., H. Kimura & E. Altman (2017) Machine learning models and bankruptcy prediction. Expert Systems with Applications, 83, 405-417.

Breiman, L. (2001) Random forests. Machine learning, 45, 5-32.

Brochet, F., M. Loumioti & G. Serafeim (2015) Speaking of the short-term: Disclosure horizon and managerial myopia. Review of Accounting Studies, 20, 1122-1163.

Brockett, P. L., A. Charnes, W. W. Cooper, D. Learner & F. Y. Phillips (1995) Information theory as a unifying statistical approach for use in marketing research. European Journal of Operational Research, 84, 310-329.

Deakin, E. B. (1972) A discriminant analysis of predictors of business failure. Journal of accounting research, 167-179.

Desai, H., S. Rajgopal & J. J. Yu (2016) Were information intermediaries sensitive to the financial statement-based leading indicators of bank distress prior to the financial crisis? Contemporary Accounting Research, 33, 576-606.

Kraft, A. G., R. Vashishtha & M. Venkatachalam (2018) Frequent financial reporting and managerial myopia. The Accounting Review, 93, 249-275.

Sharda, R. & D. M. Steiger (1995) Using artificial intelligence to enhance model analysis., 263-279.: Springer.