A Study on Early-warning of Enterprise Financial Crisis Based on Mixed Multiple Classifier Prediction

Mingrong Deng
Beijing Institute of Science and Technology Information
Beijing, China
dengmr@bjstinfo.com.cn

Anqi Chen
Beijing Institute of Science and Technology Information
Beijing, China

***Abstract*—To help enterprises to know how to spot signs of problem and crisis ahead of time, and get forewarning and intervene in advance so as to safeguard survival and development of the enterprise by reversing the passive and disadvantageous status, we creatively divided them into healthy financial companies, companies with potential financial crisis and the distressed ones, to judge the financial status of healthy companies and whether the healthy companies are on the verge of financial crisis accurately. Based on previous researches, this paper constructs a forewarning model based on combination of multiple classifiers to apply it to company crisis forewarning under situation of multiple classification of company financial status.**

Experiment results show that the combined model has good identification ability. On one hand, it integrates classification information of various basic classifiers and increases classification accuracy. On the other hand, this combined model takes other enterprises in potential crisis besides regular ST enterprises and non-ST enterprises into consideration, reveals the enterprise financial distress situation more clearly and broadens crisis forewarning scope, which has great significance for following studies.

Keywords—Early-warning; Multiple Classifier; Mixed Prediction; Financial situation of multiple classifiers.

I. INTRODUCTION

As global economic integration speeds up, competitions among enterprises evolve from regional and domestic competitions into global competitions, leading to "internationalization of domestic market and domestication of international enterprises". Changes in operation environment brought by fierce competition impose great challenges on enterprise survival and development. If there is any carelessness, enterprise will get mired in crisis. It's crucial for each enterprise to know how to spot signs of problem and crisis ahead of time in a world economy with fierce competitions and changes to get forewarning and intervene in advance so as to safeguard survival and development of the enterprise by reversing the passive and disadvantageous status.

II. LITERATURE REVIEW

Domestic and overseas scholars have done large quantities of researches on enterprise financial crisis forewarning at present and gained various valuable conclusions. Based on previous literature on financial crisis, researches mainly divided enterprise financial status into "normal" and "crisis" types [1]. The author believe that categorizing all the listed companies into two types-company in crisis and healthy company actually simplifies the construction of crisis identification model, rendering it hard to judge the financial status of healthy companies and whether the healthy companies are on the verge of financial crisis. Drawing on previous studies and real situation of China, this paper divides the financial status of listed companies into three categories, namely companies with stable financial status, companies with potential financial crisis and companies in financial crisis.

For construction of crisis forewarning models, the most common crisis forewarning models used in recent literature mainly consist of logistic regression, decision tree, artificial neural network, support vector machine, etc.. Each model has its own strength and weakness in prediction accuracy, stability and interpretation [2]. In recent years, scholars and experts tends to adopt combination of multiple information sources, multiple characteristics withdrawal and multiple identification methods to realize identification system of high performance as well as raise identification rate and confidence [3].

III. MODEL CONSTRUCTION BASED ON MULTIPLE CLASSIFIERS COMBINATION PREDICTION

A. Model Structure Based on Parallel Combination

This research proposes that prediction method based on combination of multiple classifiers can integrate classification information gained by various methods to avoid partiality of single method and increase stability of classification prediction performance. Hence, based on previous researches, this paper constructs a forewarning model based on combination of multiple classifiers (see Fig.1) to apply it to company crisis forewarning under situation of multiple classification of company financial status, expecting to achieve sound prediction effect.

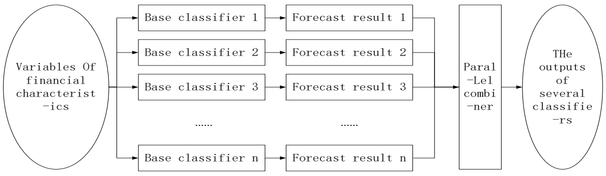

Fig. 1. Forewarning Model for Financial Crisis Based on Parallel Combination

B. Selection of Basic Classifiers

Windeatt and Terry deem that classification patterns of each basic classifier shall have relatively large variation in order to make combined model gain satisfactory classification performance and possess an accuracy rate of more than 50%.

Variation among basic classifiers adopts Q statistics to examine, where Q statistics is defined as [4]:

$$Q_{K1k2} = \frac{ad - bc}{ad + bc}$$

Among which refers to the number of samples that are predicted correctly by classifier f_{k1} and f_{k2} in test set. B refers to the number of samples that are predicted correctly by classifier f_{k1}, but predicted wrongly by f_{k2}. C refers to the number of samples that are predicted wrongly by classifier f_{k1}, but predicted correctly by f_{k2}. D refers to the number of samples that are predicted wrongly by classifier f_{k1} and f_{k2}. Q statistics magnitude is real number between -1 to 1 and decreases as the variation degree increases.

Drawing on the previous literature, this paper adopts three algorithms-logistic regression, C4.5 decision tree and BP neural network as well as classifiers with high classification accuracy as candidate basic classifiers to establish the most accurate basic classifiers by adjusting parameters and structures and calculate variation statistics magnitude Q among various basic classifiers.

In accordance with results listed in table 1, basic classifiers established by logistic regression, C4.5 decision tree and BP neural net have strong positive correlation, but they also have certain deviations.

TABLE I. VARIATION STATISTICS MAGNITUDE AMONG VARIOUS BASIC CLASSIFIERS

Classifier	Logistic Regression	C4.5 Decision Tree	BP Neural Network
Logistic Regression	---	0.874	0.856
C4.5 Decision Tree	---	---	0.724
BP Neural Network	---	---	---
Q Mean Value	0.818		

C. Calculation of Basic Classifiers' Voting Weight

This research adopts classic principle of "Voting Based on Prior Knowledge" [5], first utilizes large quantity of samples to know identification status of each classifier and them forms confusion matrix of identification status of each classifier [6].

$$CM_k = \begin{bmatrix} n_{11}^{(k)} & \cdots & n_{1M}^{(k)} & n_{1(M+1)}^{(k)} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & n_{ij}^{(k)} & \cdots & \cdots \\ n_{M1}^{(K)} & \cdots & n_{MM}^{(k)} & n_{M(M+1)}^{(k)} \end{bmatrix}$$

Among which, k=1,2,⋯⋯,K; $n^{(k)}_{ij}$ represents that classifier f_k identifies samples in C_i category as the number of C_i category. If i=j, that means classifier f_k can correctly identify the sample amount in C_i category; if $i \neq j$, that means classifier has wrongly identified samples in C_i category as sample amount in C_j category.

If the identification result of classifier f_k is C_j, the probability of sample coming from C_j category can be presented as follows by using conditional probability p [7]:

$$p(x \in C_i / f_k(x) = C_j) = \frac{n_{ij}^{(k)}}{n_j^{(k)}} = \frac{n_{ij}^{(k)}}{\sum_{1}^{m} n_{ij}^{(k)}}$$

If the amount of samples of generated confusion matrix CM_k is enough and reflects the distribution of pattern P, then the confusion matrix reflects the identification situation of classifier f_k. CM_k can be used as prior knowledge when classifiers are combined to determine the combination weight formula of various classifiers [8]:

$$W_k(X \in C_i) = P(x \in C_i / f_k(x) = C_j)$$

When classifier f_k identifies sample X as category C_j, the voting weight it gives to C_j category is $W_k(X \in C_i)$, that is the probability of the K classifier predicting sample point X, which actually belongs to C_i, as C_j.

D. Voting Combination plan for Multiple Classifiers

In this study, different basic classifiers are given different weight in voting to fully fulfill advantages of various classifiers and realize complementary advantages of parallel combination. Voting combination plan can be described as follows:

$$F(X_i = C_n, if)$$

$$TW(X_i \in C_n) = \max_{1 \le j \le m} TW(X_i \in C_j)$$

$$TW(X_i \in C_j) = \sum_1^k W_k(X \in C_j)$$

Among which $W_k(X \in C_i)$ represents the voting weight of K classifiers in C_j category. $TW(X_i \in C_j)$ Represents total voting weight of all K classifiers of sample point X_i in C_j.

IV. Forewarning Indicator And Sample Selection

A. Selection of Forewarning Indicators

There is no unified method for selecting financial crisis forewarning variable indicators. Drawing on previous research experience, this paper chooses 19 financial indicators as candidate indicators including profit capacity, paying capacity, growing capacity, and operating capacity and currency flow capacity.

B. Sample Selection

This paper collects 50 companies from Shanghai and Shenzhen stock market that have been in deficit for two consecutive years and been ST in the time period of 2012-2014 as sample companies in financial distress, and then chooses 50 listed companies with normal identification whose net profit of (t-1) year or (t-2) year is negative or whose net asset value per share is lower that par value of stock as sample companies in potential crisis as sample companies in potential crisis. At last, 50 listed companies that have never been ST are rated as sample companies with normal finance in line with matching between industry and asset scale.

The year when the sample companies are in financial distress is regarded as standard year (t-0). The companies' financial indicator data of (t-2) year extracted from Net Ease Financial Information Database is regarded as the original data base. The number of sample data is 150.

V. Experiment Design and Results Analysis

This paper first imports relevant data into SPSS Clementine and establishes the most accurate basic classifiers by adjusting parameters and structures so as to calculate classification performance of each basic classifier and determine the voting weight of parallel combination to finally gain results from the combined model. At last, we evaluate the robustness of models by evaluating their accuracy change rate in testing set and test set.

We import processed data into logistic regression, C4.5 decision tree and BP neural network algorithm in SPSS Clementine and gain prediction results and accuracy of each basic classifier by repeat testing, together with parameter and structure adjustment. See Table II for detailed results.

TABLE II. Prediction Accuracy of Each Basic Classifier

	logistic Regression	C4.5 Decision Tree	BP Neural Network	Mean Value
Total Accuracy Rate	76.7%	72.2%	81.2%	76.70%
False Positive Rate of First Species	23.3%	25.6%	18.6%	22.50%
False Positive Rate of Second Species	26.2%	52.4%	33.3%	37.30%
False Positive Rate of Third Species	20.8%	0.1%	0.1%	7.00%

Note: false positive of the first species refers to the fact that the companies are samples in crisis in real situation, but are identified as normal samples or samples in potential crisis. False positive of the second species refers to the fact that the companies are samples in potential crisis, but are identified as normal samples or samples in crisis. False positive of the third species refers to the fact that the companies are normal samples, but are identified as samples in crisis or potential crisis.

Therefore, we establish confusion matrixes for identification status of each basic classifier in line with their respective prediction results. Based on identification status of each basic classifiers listed in Table III, weights of basic classifiers are determined to calculate prediction accuracy of the combined module.

TABLE III. Confusion Matrixes for Identification Status of Each Basic Classifier

	logistic Regression		
Reality / Identification	**C1**	**C2**	**C3**
C1	73.33%	8.57%	13.21%
C2	6.67%	88.57%	15.09%
C3	20.00%	2.86%	71.70%
	C4.5 Decision Tree		
Reality / Identification	**C1**	**C2**	**C3**
C1	69.57%	24.38%	6.67%
C2	21.74%	73.54%	20.00%
C3	8.70%	2.08%	73.33%
	BP Neural Network		
Reality / Identification	**C1**	**C2**	**C3**
C1	81.40%	14.29%	5.45%
C2	16.28%	80.00%	12.73%
C3	2.33%	5.71%	81.82%

From the prediction accuracy results listed in Table IV, it can be seen that the mean accuracy rate of basic classifiers in training set is 76.7% and the mean accuracy rate in test set is 71.1%, while that of the combined model is at least 5% higher, demonstrating that the combined model has greatly improved the classification accuracy.

TABLE IV. EACH BASIC CLASSIFIER AND COMBINED MODEL'S PREDICTION ACCURACY IN TRAINING SET & TEST SET

	Training Set	Test Set
logistic Regression	76.7%	72.3%
C4.5 Decision Tree	72.2%	65.6%
BP Neural Network	81.2%	75.3%
Mean	76.7%	71.1%
Combined Model	83.9%	76.1%

Comparing the prediction accuracy of the combined model with that of each basic classifier, it shows that the prediction accuracy of BP neural network is most close to that of the combined model, which is in line with classifier's characteristics. The prediction accuracy rate of single classifiers of neural network is the highest and most close to real situation due to its powerful nonlinear training fitting ability, which identification performance of C4.5 decision tree is the lowest, almost 10% lower than that of the combined model.

VI. CONCLUSION

This paper constructs a combined crisis forewarning prediction model based on multiple classification of financial statuses and utilizes real data of listed Chinese companies to undertake empirical studies. Experiment results show that the combined model has good identification ability. On one hand, it integrates classification information of various basic classifiers and increases classification accuracy, which has bright application prospect in crisis forewarning. On the other hand, this combined model takes other enterprises in potential crisis besides regular ST enterprises and non-ST enterprises into consideration, reveals the enterprise financial distress situation more clearly and broadens crisis forewarning scope, which has great significance for following studies. Besides, from the identification accuracy rate of all classifiers in three crisis situations, it can be seen that the positive false rate of normal company samples is the lowest, that of the crisis samples is higher, that of samples in potential crisis is the highest. After discussion, we believe that this may results from insufficient accuracy in defining potential crisis situations. We will further explore this problem in future researches.

ACKNOWLEDGMENT

This research work was supported by Financial Foundation of Beijing under Grant No.PXM2014_178214_000008.

REFERENCES

[1] Jie Sun, Hui Li. Multiple classifiers hybrid combination for predicting companies' financial crisis, vol.2, System Engineering Theory & Practice, 2009, PP.78-86.

[2] Hongling Guo, Xianyi Chen. Selection and integration methods for multiple classifiers, vol.13, Computer Engineering & Application, 2009, PP.186-187.

[3] Jie Sun, Hui Li, Meng Zhang. Multiple classifiers hybrid combination for companies' financial distress prediction, vol.5, Systems Engineering Theory & Practice Online, 2009, PP.292-295.

[4] Bin Luo, Peiji Shao, Jinyao Luo, Duyu Liu, Guoen Xia. Telecommunication customer churn prediction based on the dynamic integration of multiple classifiers, vol 5, System Engineering Journal, 2010, PP.703-711.

[5] Chunying Li, Zhikang Tang, Yuanda Cao. Study on traffic jam prediction model of multiple classifiers combination, vol.23, Computer Engineering and Design, 2010, PP.5088-5091.

[6] Yonglin Liang. Data mining algorithm study and application based on multiple classifier integration, Chongqing University, 2007.

[7] Jie Sun, Hui Li. Financial distress prediction using support vector machines: Ensemble vs. individual, vol.13, Applied Soft Computing Journal, 2012, PP.128-141.

[8] Bo Qian, Khaled Rasheed. Foreign exchange market prediction with multiple classifiers, vol.7, J. Forecast, 2010, PP.293-302.

