Financial Crisis Prediction Based on Independent Component Analysis and Neural Network

You Zhou

Jiangsu Vocational Institute of Commerce, Nanjing 211168, China

rtinguu3@163.com

Keywords: crisis analysis, neural network, independent component analysis, dimension reduction

Abstract: Neural network is widely used in financial analysis. However, there are many indicators to describe financial ratio. If these indicators are used as model inputs, the network structure will be too large. Independent component analysis (ICA) is a new signal processing technology developed in recent years, which decomposes multi-channel observation signals into several independent components by optimization algorithm according to the principle of statistical independence. In this paper, we propose to use ICA to reduce the dimension of high-dimensional indexes in the small sample problems such as financial crisis analysis, reduce the network scale and improve the reliability of analysis and prediction on the premise of retaining the vast majority of sample information.

1. Introduction

The prediction of financial crisis of an enterprise is based on the financial statements, business plans and relevant economic data of the enterprise [1,2]. Through the comprehensive analysis and prediction of financial indicators, it reflects the changes of the operation and financial situation of the enterprise, and sends an early warning signal to the operational risks that may occur in each link of the enterprise. At the same time, find out the causes of the financial crisis and the hidden problems in the financial management system of the enterprise to provide decision-making basis for the enterprise management.

Empirical research shows that neural network analysis is more effective than linear model and logistic regression model in determining the accuracy, and it is not affected by the distribution characteristics of variables. It does not need to judge the financial crisis state subjectively and qualitatively, so it can determine the crisis state more reasonably. BP neural network is a kind of neural network which is widely used in financial analysis [3]. However, there are many indicators to describe financial ratios. If these indicators are used as model inputs, the network structure will be too large. For the large network structure, on the one hand, it requires a huge training sample set support, which is difficult to achieve in general research; on the other hand, it will lead to the decline of generalization ability of neural network, and make its prediction performance worse. Therefore, in many researches, only some indexes are selected as model input, which will lead to incomplete analysis information.

Some researchers use principal component analysis (PCA) to reduce the dimension of high-dimensional indexes, but PCA is based on the Gaussian hypothesis. In the case of small samples (in many cases, financial analysis is a small sample), the Gaussian hypothesis of sample distribution is generally not tenable. At this time, PCA's ability to reflect the original index information decreases, which makes the reliability of neural network analysis and prediction results lower.

Independent component analysis (ICA) is a new signal processing technology developed in recent years. According to the principle of statistical independence, the multichannel observation signal is decomposed into several independent components by optimization algorithm. These components are the estimation of the source signal, and the observation signal can be reconstructed by the linear combination of these independent components.

The basic ICA refers to the technology of separating the source signals from the linear mixed signals of multiple source signals. In addition to the known source signal is statistically independent, there is no other prior knowledge. ICA is developed with the blind source problem, so it is also called blind separation.

The basic precondition of ICA is that the source signal is non Gaussian distribution, which is more suitable for dimension reduction and feature extraction of small samples. In this paper, we propose to use ICA to reduce the dimension of high-dimensional indexes in the small sample problems such as financial crisis analysis, reduce the network scale and improve the reliability of analysis and prediction on the premise of retaining the vast majority of sample information.

2. Independent component analysis

Independent component analysis can be regarded as the extension of principal component analysis in the field of data processing [4,5]. The basic idea of both is to obtain the signal characteristic components by decomposing the data, to construct the characteristic space and deal with the corresponding problems in the characteristic space. But because of the difference between the two methods, there are also great differences. PCA uses the second-order statistical law based on data covariance matrix, while ICA algorithm is much more complex, which uses the high-order statistical law of data; in PCA, the data processed is generally assumed to be Gaussian distribution, while in ICA, the data processed is assumed to be non Gaussian distribution.

In many practical problems including financial crisis analysis, most of the data processed are non Gaussian distribution, and the important characteristics of the data are often included in the high-order statistical characteristics. From the above analysis, we can see that ICA is closer to the actual situation and meets more strict mathematical conditions than PCA, so it has greater application potential.

In independent component analysis, let $x = (x_1, x_2,...) , x_n)^T$ is n random observation signals; each observation signal is composed of M unknown source signals $s = (s_1, s_2 , s_n)^T$ (where $n \geq m$) is a linear mixture; each random observation signal and source signal is a zero mean variable. Then the linear model of ICA can be expressed as:

$$x = As \tag{1}$$

$$x_i = \sum_{j=1}^{m} a_{ij} s_j \tag{2}$$

This model describes how the observation signal is mixed by the source.

From formula (2), we can see that each observation signal x_i is obtained by independent components through different a_{ij} linear weighting. The independent source s_j is an implicit variable and cannot be measured directly; $A = [a_1, a_2, ... a_m]$ is an $n \times m$ matrix, which is called a hybrid matrix. The hybrid matrix is also an unknown matrix, and the only known information available is the observed signal X.

Independent component analysis is based on a simple assumption that the source variable s is statistically independent and the source variable is non Gaussian. After the mixed matrix A is estimated, the separation matrix $w = A^{-1}$ can be obtained by finding the inverse of the mixed matrix, and then the estimation u of the independent component s can be obtained

$$u = x = wAs = s \tag{3}$$

Independent component analysis includes two parts: preprocessing and estimation of separation matrix W.

3. Financial index system model

As for the description of financial crisis, Altman and other scholars have integrated four situations to define financial crisis, specifically as follows: first, business failure. Second, insolvency. Third, breach of contract. Fourth, bankruptcy. Most of the Special Treatment (ST) stocks in China's

securities market are due to "two consecutive years of losses or the net asset per share is less than the par value (1 yuan)", that is, the deterioration of financial indicators is the main reason for the special treatment of listed companies.

Due to data disclosure and other issues, domestic scholars generally regard ST listed companies as financial crisis companies. For the same reason, the definition of financial crisis in this paper is based on whether it is St. if the listed company is ST, the company is in financial crisis; otherwise, it is not in financial crisis.

According to the analysis of financial crisis prediction model, there are 30 financial ratios commonly used in financial crisis prediction. If all 30 financial ratio indexes are taken as the input of neural network (i.e. 30 dimensional vector), there are as many as 30 input nodes. On the one hand, too large network structure requires huge training sample set support, on the other hand, it will lead to the decline of generalization ability of neural network; If we simply select some of the 30 indicators as neural network input, the objectivity and reliability of the prediction results will be reduced because of the incomplete input information.

Table 1. Common financial crisis prediction ratio

Serial number	Index	Serial number	Index
1	Cash / current liabilities	16	Quick assets / net sales revenue
2	Total liabilities / equity	17	Current assets / total assets
3	Cash / net sales revenue	18	Quick assets / total assets
4	Net income / total assets	19	Current liabilities / equity
5	Cash / total assets	20	Return on common stock
6	Quick assets / inventory	21	Equity / fixed assets
7	Cash flow / current liabilities	22	Retained earnings / total assets
8	Net sales revenue / total assets	23	Equity / net sales revenue
9	Cash flow / total liabilities	24	total liabilities / total assets
10	Operating income / total assets	25	Inventory / net sales revenue
11	Cash flow / total assets	26	Working capital / net sales revenue
12	EBIT / total interest	27	Long term Liabilities / equity
13	Current assets / current liabilities	28	Working capital / total assets
14	Quick assets / current liabilities	29	Equity market value / liabilities
15	Current assets / net sales revenue	30	Working capital / total assets

4. ICA neural network analysis example

The samples selected in the experiment are divided into two categories: financial crisis (ST) companies and non-financial crisis (non ST) companies. 51 small and medium-sized manufacturing companies are selected from ST company, and then 51 non-financial crisised enterprises are selected according to the same industry and scale. 30 ST and 30 non ST enterprises were selected as the training sample set, and the remaining 21 ST and 21 non ST enterprises were selected as the test sample set. All the data are from the public disclosure on the Internet.

In order to reflect the ability of ICA and PCA to describe information in such small sample problems as financial crisis analysis, this paper adopts the same neural network structure, using ICA and PCA to reduce the original 30 dimensional financial ratio index into 5-Dimensional index, and taking the reduced index as the input of neural network. During the test, different numbers of samples in the training set are used for training, and the test is carried out in the complete test set. See Table 2 for the results.

Table 2. Test error rate of test sample set (42 samples in total)

Training samples	10	20	50	60
ICA dimension reduction	4(9.5%)	4(9.5%)	3（7.1%）	3（7.1%）

ICA dimension reduction	9(21.4%)	7(16.7%)	6（14.3%）	4(9.5%)

From table 2, we can find the following situations:

First, in the case of few training samples, the error rate of ICA dimension reduction is much lower than PCA dimension reduction. Under the condition of 10 and 20 samples, the error rate of ICA is only 9.5%, which is much lower than 21.4% and 16.7% of PCA.

Secondly, with the increase of training samples, the error rates of PCA and ICA are decreasing, but the error rates of ICA are lower and more stable.

Thirdly, it can be predicted that if the number of training samples is increased, the error rate of PCA will be further reduced. However, in practice, it is often difficult to collect enough training samples.

To sum up, as an analysis technology with parallel computing, self-learning and fault tolerance, neural network provides a feasible method for enterprise financial dynamic early warning. In order to avoid too large-scale neural network or lack of information due to simply selecting some indicators, and to adapt to the small sample conditions in the financial crisis analysis, this paper proposes a method that first uses independent component analysis to reduce the dimension of high-dimensional indicators, and then uses neural network to predict financial crisis. According to the prediction of financial crisis of ST and non ST Companies in China, the prediction ability of ICA neural network is significantly higher than PCA neural network under the condition of small sample.

References

[1] Ji Weiwei. The historical evolution of the research on prediction of corporate financial crisis, accounting monthly.2018(4), p.14-22

[2] Altman E I., Financial ratio. Discriminant Analysis and Prediction and Corporate Bankruptcy. Journal of Finance,1968(9), p.67- 69

[3] Zhu Min, Nie Shunjiang, fan Xiaoyan. Research on financial early warning mechanism based on BP neural network. Financial accounting communication, 2007(4), p.45-49

[4] Chen Lei. prediction of corporate financial crisis based on proportional hazard and principal component model, research on financial issues, 2007(7), p.35-38

[5] Jutten C, Herault J. Independent Component Analysis versus Principal Component Analysis. Processing of Europena Signal Processing Conference, Grenoble, France,1988, p.643- 646.