Research on the Risk Crisis Prediction of Enterprise Finance by Genetic Algorithm

Tingting Ye

Abstract—With the development of the global economy, the competition between enterprises is getting fiercer and enterprises are facing increasing crises and challenges As a result, the prediction of risk crisis of enterprises becomes very important. The growing numbers of data within enterprises have caused great inconvenience to the risk crisis forecasts. In this paper, the genetic algorithm is analyzed. With this algorithm, a large number of high-latitude data are reduced in dimension. The genetic algorithm is also used to optimize the neural network. A genetic algorithm model is established by MATLAB. The experiment proves that this algorithm can effectively optimize the BP neural network, and has an obvious early warning effect on the financial risk crisis with high accuracy, which provides a reference for the application of the algorithm in the prediction of the financial risk crisis.

Keywords—BP neural network, crisis prediction, financial risk, genetic algorithm

I. Introduction

Enterprise is a kind of artificial system and organization [1]. With the rapid development of the market economy, market competition is becoming increasingly fierce. With the development of information technology, enterprises have begun to have more and more huge data, which brought inconvenience to data processing. In the context of international economic integration, the prediction of the risk of corporate finance crisis is particularly important, so we need to study an effective and scientific method to predict the risk of corporate financial crisis. Li *et al.* [2] proposed a financial crisis prediction model, which combined genetic algorithm with support vector machine to use the financial information spread on the network with high accuracy. Experimental results showed that the proposed model could be used in financial and corporate governance and realize more accurate financial crisis prediction. Hu *et al.* [3] established a new risk rating method based on default distance and order statistics, and then studied the financial crisis forecast based on feature-weighted SVM model, and proved that the model had good performance in predicting the financial crisis of listed companies. Wang *et al.* [5] developed a FOA based Logistic financial risk warning model and proved its effectiveness through experiments. Fan [7] proposed a weighted logistic regression based financial risk warning model and found that it had favorable nonlinearity. Huang *et al.* [8] established a SVM based financial warning system for listed companies and believed that it could be extensively applied. In this study, genetic algorithm was analyzed and combined with BP neural network as an evaluation function for predicting accuracy. Moreover a genetic algorithm model was established using MATLAB for predicting the financial risk crisis. The validity of this model was verified through experiments.

T. T. Ye is with School of Business Administration, University of Science and Technology, Liaoning, Anshan, 114051, China (e-mail: tingtye@sina.com).

II. Enterprise financial risk related factors

Factors affecting the financial risks of enterprises can be divided into internal factors and external factors.

The external factors are uncontrollable. Under the environment of market competition, enterprises as a part of the market economy will inevitably be affected by the whole market. Enterprises cannot determine the policies released by the government. In addition, the increasingly fierce market competition makes enterprise development unstable, and an enterprise may go up or down in a blink. The financial condition of an enterprise may be affected once it become inferior in the competition, which can induce financial risks, cause financial crisis and even lead to bankruptcy.

The internal factors are also an important part of the financial risk. In many cases, the wrong decision of the operator will cause irreparable loss. Business operation is far-reaching; a small mistake will cause huge reaction. Therefore enterprise operators can hardly absolve themselves from the blame in the problem of financial risks. Moreover many enterprises have unreasonable capital structure, bad management and weak profitability, which can greatly increase the probability and severity of financial risks.

III. Enterprise financial crisis

The enterprise financial crisis is manifested in the following aspects:

(1)Management: deficit caused by low sales income; profit decline caused by low efficiency; capital flow influenced by decreased survival turnover; insufficient cash flow.

(2)Investment: Diversified investment should be implemented in limited financial circumstances, considering coordination between corporate asset structure and capital structure as well as between profitability and liquidity. Mere consideration of profitability would put the company in technical difficulties, and the mere consideration of liquidity would leave the company idle most of its properties and lose profitability.

(3)Financing: It is an important factor for the development of enterprises to make profits through the use of other people's funds. In this way, enterprises can raise the funds needed for development and improve the rate of return on capital. While the capital return rate is improved, financial risks are also increasing. When the enterprise is in debt, it is easy to lose the ability to repay due to poor management, which causes the enterprise to go bankrupt.

(4)Other aspect: In the process of operation, enterprises are prone to large lawsuits and other emergencies, from which they are likely to suffer huge losses and even go bankrupt.

Genetic algorithm can reduce the dimension of the financial data in enterprises and optimize the BP neural network prediction model, which plays an important role in the prediction of corporate financial crisis.

IV. Genetic algorithm features

Genetic algorithm is a new global optimization algorithm based on biological genetics [9]. It can improve the individual fitness of the population through continuous evolution, thus achieving the purpose of finding the optimal goal.

Compared with other algorithms, it has the following features:

(1)Self-organization and adaptability features.

(2)It can encode the individuals and process with the encoded characters.

(3)Extensive feature.

(4)Parallelism feature.

(5)Expandability [10].

With these features, it can be applied to function optimization design, combination optimization design, production scheduling optimization design and image signal processing [11]. Genetic algorithm is a probabilistic search algorithm which can use encoding technology to encode individuals to form chromosomes. In this process, the evolutionary process of the individuals composed of character strings is simulated [12].

The main steps of the algorithm are:

(1)Code according to actual problems.

(2)Generate primary population: randomly generate N individuals to generate primary population.

(3)Evaluation of fitness value: Evaluation of fitness of individuals of primary population based on the target of optimization.

(4)The genetic operator operations of selection, crossover and mutation are carried out.

(5)Form an evolving population.

(6)If the conditions are met, it goes to the next step; if not, it goes back to step 3.

(7)Output the results.

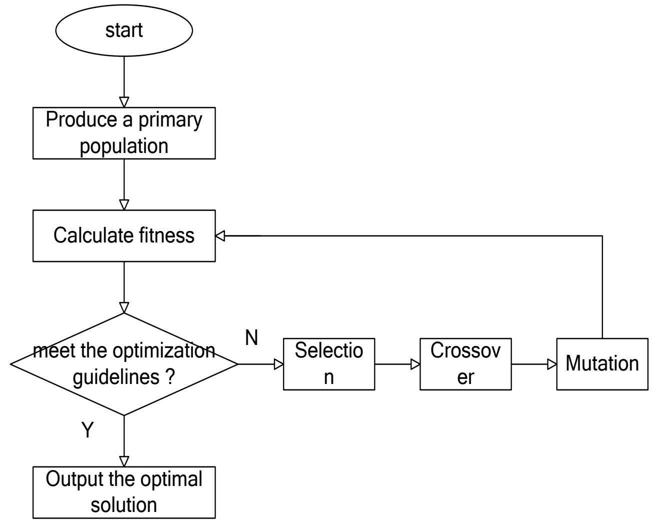

Fig. 1 Genetic algorithm flow

A. *Feature model based on genetic algorithm*

In this paper, genetic algorithm is used in feature selection, and the advantages of Filter model [13] and Wrapper model [14] are combined to form a hybrid selection model which can realize feature selection of information through filtering as well as feature selection of data by sealing method so that data can be processed quickly while its quality is ensured.

B. *Feature selection based on hybrid selection mode*

The main characteristic of the filtering method is that the evaluation criterion is independent of the classification algorithm and the relationship between the two characteristics is not considered [15]. Therefore, the selection speed of the filtering algorithm is fast, but its accuracy is not high. In this paper, it is used for the screening of the original data. The main feature of the encapsulated method is that the evaluation criterion is independent of the learning algorithm of the classifier, which can encapsulate the evaluation algorithm in the learning algorithm [16]. Therefore, the algorithm can select a subset of features with high performance, which is applied in this paper for the fine screening of the features after initial screening to ensure the performance of the feature subset.

C. *Double criteria evaluation model*

This paper applies the double criteria evaluation model and obtains a union set of two results, which not only rapidly screens targets but also prevents the randomness of a single criterion. The double criterion used in this paper is information gain and correlation coefficient. Information gain is information measurement, and correlation coefficient is dependency measurement.

(1) Information gain

Entropy is the average amount of information after the elimination of redundant information, which is generally referred to as information entropy [17]. Information entropy mainly uses the numerical form to measure the uncertainty of the value of a random variable.

Suppose N to be a random variable, $q(n)$ to be the probability when $N=n$, and the information entropy of N can be expressed as $H(N)$, then the equation of $H(N)$ is as follows:

$$H(n)=-\sum_{i} q(n_i)\log(q(n_i)) \tag{1}$$

The information entropy $H(N)$ is not relevant to the value of N but its probability distribution. With the increase of the uncertainty degree of the value of N, its probability distribution increases, which leads to the increase of its information entropy $H(N)$, requiring more information. When the value of N reaches the highest degree of uncertainty, its value of information entropy will also reach the maximum. In this case, the probability of the random variable N is the same, so that each value has the same chance to appear. At this moment, we cannot determine the specific value of the random variable N but only choose one random value. But when X has only one value of x, the probability distribution of this value is 1, i.e., $q(n)=1$, where the value of X is definite, and its information entropy is minimized.

Conditional entropy refers to the dependence of one variable on another variable, and when a variable is known, conditional entropy is the uncertainty of another variable. If the random variable M is the known variable, then the conditional entropy of the random variable N on M is:

$$H(N|M)=-\sum_{i}\sum_{j} q\left(m_j\right)q\left(n_i|m_j\right)\log(q(n_i|m_j)) \tag{2}$$

Information gain refers to the information obtained after the elimination of uncertainty, which is the difference of information entropy [18]. When information gain is taken as the evaluation criteria, it mainly considers the information amount its feature provides for the classification. The more information a feature provides to the classification, the larger the information gain. If a feature is deleted, then the change in the overall amount of information before and after the feature is the information amount provided by the classification and the changed part is the information entropy of the feature.
Information gain is defined as follows:

$$IG\left(N,M\right)=H(N)-H(N|M) \tag{3}$$

Information gain can most intuitively show the amount of information provided by the feature for the overall classification and can be widely applied to feature selection.
(2) In this paper, Spearman correlation coefficient [19] is used to collect data with uncertain distribution. The correlation between two features is tested by the uniformity between them. When both variables increase or decrease simultaneously, these two variables are Spearman-related. The data in the data set is ranked according to their size. Average grade calculation is performed on repeated data and whether the rank of two variables is relevant is checked. Then, the Spearman correlation coefficient can be expressed as:

$$a_l=1-\frac{6}{x(x^2-1)}\sum_{i=1}^{x}(W_i-P_i)^2 \tag{4}$$

Where the rank correlation coefficient a_l refers to the magnitude of the rank correlation, when it is positive, it refers to positive correlation, when it is negative, it refers to negative correlation, when it is 0, it refers to zero correlation (its value is between -1 to 1). W_i and P_i refer to the rank of n_i and m_i .

V. BP NEURAL NETWORK MODEL BASED ON GENETIC ALGORITHM

A network structure with 14 input layer nodes, 16 hidden layer nodes and 1 output layer node is adopted in this design.

A. BP network weight based on genetic algorithm

(1) Chromosome expression and primary population
Coding is an important problem in genetic algorithms. With the increase of the complexity and the dimension of data, the shortcoming of too long coding in binary coding is exposed. To avoid this problem, we directly apply floating-point encoding to the performance of solution to improve the efficiency and accuracy of solving problems.

Randomly generate a population of X individuals, $N=\left\{N_1,N_2,\cdots N_N\right\}$, each individual $N_1=\left\{n_1,n_2,\cdots n_x\right\}$ represents initial weights and threshold distribution of a neural network; each gene value represents a link weight or threshold for a neural network, then the length of the individual is the sum of the number of weights and thresholds of the neural network, as follows:

$$L=S*L_1+L_1*L_2+L_1+L_2 \tag{5}$$

where S refers to the number of input layer, L_1 refers to the number of hidden layer, L_2 refers to the number of output layer.
(2) Objective function and fitness function
The error sum of squares of the network output value and expected output value is an important performance of BP network. The smaller the square sum, the better the network performance.
Mark the expected output of the output layer neuron of the ith group to be R^*, the actual output to be R; A refers to the total sample number and its error sum of squares $E\left(\alpha\right)$ is:

$$E\left(\alpha\right)=\frac{1}{2}\sum_{i=1}^{A}(R^*-R)^2 \tag{6}$$

Then, genetic algorithm can be applied to complete the search of network weights. Substituting each chromosome's weight and threshold into the BP network to minimize the error function $E(\alpha)$ in the neural network and the target function is defined as follows:

$$\min E(\alpha)=\frac{1}{2}\sum_{i=1}^{A}(R^{*}-R)^{2} \tag{7}$$

The fitness function $D(\alpha)$ of genetic algorithm is:

$$D(\alpha)=1/E(\alpha) \tag{8}$$

(3) Genetic operation

Each chromosome is calculated and proportionally assigned. Then, the possibility of descendants left is determined based on the probabilities of various fitness degree. If there is an individual α , and its fitness degree is $F(\alpha)$, then the probability of being selected is:

$$Q_i=F(\alpha_i)/\sum_{i=1}^{b} f(\alpha_1) \tag{9}$$

VI. Simulation results

The model established above is applied for the simulation experiment under Matlab environment and compared with normal neural network model. The collected data were input into the normal neural network model and the results include misjudgment, accuracy and modeling time.

The data of 75 listed companies during 2013-2015 were respectively input into the traditional neural network and the optimized BP neural network, among which, the data of 60 listed companies were taken as training and learning and the rest was taken as predicting. The predicted results were compared with the actual enterprise crisis situation. The results are shown in Table 1.

Table 1 Results of the optimized neural network prediction model

		Non-ST companies	ST companies	Average accuracy	Modeling time
2013	Sample number	9	1	55%	3.12s
	Misjudgment number	1	4		
	Accuracy rate	90%	20%		
2014	Sample number	6	7	88.90%	0.97s
	Misjudgment number	0	2		
	Accuracy rate	100%	77.78%		
2015	Sample number	11	3	95.84%	1.85s
	Misjudgment number	1	0		
	Accuracy rate	91.67%	100%		

Table 2 Results of the traditional neural network prediction model

		Non-ST companies	ST companies	Average accuracy	Modeling time
2013	Sample number	5	2	47.43%	163.2s
	Misjudgment number	6	2		
	Accuracy rate	45.45%	50%		
2014	Sample number	4	5	61.12%	5.11s
	Misjudgment number	2	4		
	Accuracy rate	66.67%	55.56%		
2015	Sample number	9	3	78.41%	7.16s
	Misjudgment number	2	1		
	Accuracy rate	81.81%	75%		

As shown in Table 1 and 2, the accuracy of the prediction model increased each year and the average accuracy of the optimized model was increased greatly. Each year's sample data individual differences made the accuracy of each year different. However, after the optimization of the genetic algorithm, BP neural network improved the accuracy of financial risk prediction and reached the expected target.

The results showed that the prediction accuracy was the lowest in 2014 and the accuracy was only 55% after optimization. In the pre-ST data of listed companies, the data for 2016 did not have much research value and using the data of 2013 to predict financial risks in three years did not have obvious advantages. Therefore, an effective prediction period started from 2014. It was also found that the accuracy of 2015 was the highest, reaching 95.84%.

Through the data of the three years, it was found that the modeling speed of the BP neural network after optimization improved a lot. With the characteristics of survival of the fittest in biology, genetic algorithm does not depend on the traditional gradient statistics and directly establishes the winning population, which greatly improves the modeling speed. In practice, it makes the prediction more convenient and fast.

Corporate finance can profoundly affect the life and death of a company, and the sooner a company can discover the risks, the sooner the business decision makers can make responding measures. Through experiments, it can be concluded that the model proposed in this paper can effectively analyze the enterprise's financial risk and avoid the unknown risk of the enterprise due to the financial situation.

VII. Discussion

With the development of economic globalization, enterprises are facing increasingly larger challenges and risks, which increase the difficulty in decision making. Financial crisis is a progressive process. A small mistake may bring irreparable loss. If there is no good warning system to help operators quickly and make appropriate response, the enterprise is likely to fall into the financial crisis. Many enterprises have been destroyed by large risks which are previously small. Once financial crisis appears, financial risks will flow from internal to external if there is no effective measure to control the risks, which can not only damage the interests of enterprises but also greatly affect the economic condition of enterprise stakeholders and the whole region [20]. In this case, an effective financial risk early warning model is necessary.

An effective financial early warning system is highly sensitive to the omen of financial crisis. It can warn operators timely, and operators can take measures early and nip financial crisis in the bud. An effective financial early warning system can also analyze the causes of financial problems and help managers find a specific way to avoid it. More importantly, financial early-warning system can try to avoid the same mistakes through a series of analysis.

Many effective risk warning models have been verified in the actual operation. Tian *et al.* [21] applied a warning model which combined quantitative methods with qualitative methods to warn real estate enterprises and got favorable feedback. Li *et al.* [22] developed an early warning model of financial risks in core enterprises under supply chain environment based on a large number of theories and tested it through examples.

In this study, a genetic algorithm based BP neural network model was established for warning financial risks of enterprises. In the simulation experiment, it was found that the accuracy of the model was higher than that of the traditional BP neural network.

To make the financial risk early warning system work better, enterprises should combine qualitative and quantitative methods to make it more accurate in the actual operation process. Timely, effective and accurate financial information is also very important for the financial early warning system; therefore enterprises should strengthen financial management and internal control and establish financial information system to improve the prediction accuracy and reduce loss.

VIII. Conclusion

With the development of the global economy, enterprises are facing more and more crises, so it is becoming more and more important to establish the crisis prediction model. This paper introduced genetic algorithm and established feature model. Combining genetic algorithm model with BP neural network, the enterprise financial risk crisis prediction model was established and the model was tested. Experiments showed that the accuracy and speed of the prediction model were greatly improved after optimized by the genetic algorithm, which provides a reference for the application of genetic algorithm in the enterprise financial risk prediction.

References

[1] M. Mazen, “The implications of the acquisition of a new nationality for the right of return of palestinian refugees,” Asian J. Int. Law, vol. 5, no. 2, pp. 356-386, 2015.

[2] C. K. Li, D. Liang, F. Y. Lin, and K. L. Chen, “The application of corporate governance indicators with XBRL technology to financial crisis prediction,” *Emerg. Markets Finan. Trade,* vol. 51, pp. S58-S72, 2015.

[3] W. H. Hu, F. Gao and C Huang, “Financial crisis prediction based on distance to default and feature weighted support vector machine,” *Int. Confer. Nat. Comput.*, pp. 58-63, 2016.

[4] Q. Lu, Y. J. Jin and Z. X. Yu, “The Financial Early Warning Model of Listed Companies Based on Lasso Method and Logistic Regression,” *Adv. Appl. Mathem.*, vol. 6, no. 4, pp. 572-582, 2017.

[5] Y. Wang, Q. Y. Ji and J. Chai, “An Research of Financial Early Warning About Logistic Model Based on FOA,” *J. Syst. Sci. Mathem. Sci.*, vol. 37, no. 2, pp. 573-586, 2017.

[6] Y. Liang, “Empirical study on China's financial risk early warning under the new normal,” *International Conference on Industrial Economics System and Industrial Security Engineering,* pp. 1-7, 2017.

[7] X. Y. Fan, “Financial Risk Early Warning Constraint Method Based on Weighted Logistic Regression,” *Boletín Técnico,* vol. 55, no. 2, pp. 272-279, 2017.

[8] H. Huang, W. Jiang and S. Wang, “Research of Financial Early-warning for Listed Companies Based on SVM,” *International Conference on Computational Science and Engineering,* pp. 278-281, 2015.

[9] H. S. Madraswala, “Modified genetic algorithm solution to unit commitment problem,” *Int. Confer. Nascent Technol. Engin.*, pp. 1-6, 2017.

[10] A. L. Spedding, G. D. Fatta and M. Cannataro, "A Genetic Algorithm for the selection of structural MRI features for classification of Mild Cognitive Impairment and Alzheimer's Disease," *IEEE Int. Confer. Bioinform. Biomed.*, pp. 1566-1571, 2015.

[11] E. Jehamy, M. Salaün, M. Ney, and G Landrac, "Application of genetic algorithm for constrained metal plate lens profile optimization," *Int. Symp. Antenn. Technol. Appl. Electromagn.*, pp. 1-4, 2007.

[12] M. M. Smith and Y. S. Chen, "A novel evolutionary algorithm for the homogeneous Probabilistic Traveling Salesman Problem," *IEEE/ACIS, Int. Confer. Comp. Informat. Sci.*, pp. 1-6, 2016.

[13] D. Meissner, S. Reuter, E. Strigel, and K. Dietmayer, "Intersection-Based Road User Tracking Using a Classifying Multiple-Model PHD Filter," *Intell. Transp. Syst. Magaz.*, vol. 6, no. 2, pp. 21-33, 2016.

[14] X. Hou, B. R. Hodges, S. Negusse and C. Barker, "A multi-model Python wrapper for operational oil spill transport forecasts," *Comput. Sci. Discov.*, vol. 8, no. 1, pp. 014004, 2015.

[15] V. Aguiarpulido, M. Gestal, M. Cruzmonteagudo, J. R. Rabuñal, J. Dorado and C. R. Munteanu, "Evolutionary computation and QSAR research," *Curr. Comp. - Aided Drug Design*, vol. 9, no. 2, pp. 23-25, 2013.

[16] J. Bergstra, N. Pinto, and D. D. Cox, "SkData: data sets and algorithm evaluation protocols in Python," *Comput. Sci. Discov.*, vol. 8, no. 1, 014007, 2015.

[17] V. Zelevinsky, M. Horoi and B. A. Brown, "Information entropy, chaos and complexity of the shell model eigenvector," *Phys. Lett. B*, vol. 350, no. 2, pp. 141-146, 2016.

[18] C. Yin and J. Xi, "Maximum entropy model for mobile text classification in cloud computing using improved information gain algorithm," *Multimed. Tools Appl.*, pp. 1-17, 2016.

[19] M. T. Puth, M. Neuhäuser, G. D. Ruxton, "Effective use of Spearman's and Kendall's correlation coefficients forassociation between two measured traits," Anim. Behav., vol. 102, pp. 77-84, 2015.

[20] M. X. Hong, "A study on financial risk control of group enterprise," J. *Chem. Pharm. Res.*, vol. 7, no. 4, pp. 326-333, 2015.

[21] X. R. Tian, H. D. Jiang and J. L .Liu, "The Research of Financial Risk Warning Model about Real Estate Enterprises in China," *Adv. Mater. Res.*, vol. 989-994, no. 3, pp. 2625-2628, 2014.

[22] L. Li, L. Wang and L. Yu, "Research of financial risk warning model of listed companies in the Supply Chain environment," *Control Decision Conference*, pp. 3441-3445, 2012.