ISSN 2616-5902 Vol. 6, Issue 11: 56-60, DOI: 10.25236/AJBM.2024.061109

Research and Analysis of Financial Crisis Prediction Model Based on the Fusion of Financial and Non-Financial Data with CSO Algorithm

Peng Dong

School of Business, Stevens Institute of Technology, Hoboken, New Jersey, 7030, United States
pengdong0128@gmail.com

Abstract: *With the continuous evolution of financial risks faced by enterprises, the traditional financial crisis prediction methods gradually show their limitations. Therefore, this paper proposes an innovative financial crisis prediction model combining competitive particle swarm optimization (CSO) algorithm with financial and non-financial data fusion. In the data fusion stage, the model deeply discusses the interaction between different data types, uses the CSO algorithm to dynamically optimize the model parameters, and realizes efficient feature selection to extract the most representative feature subset for financial crisis prediction. Based on the selected features, the prediction model constructed significantly improves the prediction accuracy and stability. Through systematic comparison with the traditional forecasting model, the experimental results show that the proposed model has significant advantages in accuracy and robustness, demonstrating good practicability, and providing scientific basis and reference for enterprises' financial decision-making.*

Keywords: *Financial Crisis Prediction, CSO Algorithm, Data Fusion, Feature Selection, Financial and Non-Financial Data*

1. Introduction

In the process of enterprise operation, financial crisis, as a major risk, has a wide and far-reaching impact, which is directly related to the survival and long-term development of enterprises. Therefore, it is particularly important to construct an efficient financial crisis prediction model. This model can not only help enterprises identify potential financial risks early, but also provide scientific basis for decision makers, so it is of great significance in theory and practical application. The causes of financial crises are often complex and involve multiple factors, some of which can be quantified while others are difficult to accurately assess.

With the rapid development of information technology, the amount of financial and non-financial data obtained by enterprises continues to surge. The massive increase of data not only makes the nonlinear relationship between data more complex, but also strengthens the internal correlation and deep meaning of data. This complexity is far beyond human processing capacity and significantly increases the difficulty of building high-precision predictive models.

In the research of financial crisis prediction, the financial situation of enterprises is usually simplified into two states: normal and crisis, so this problem can be regarded as a binary classification problem. The current forecasting methods are mainly divided into traditional statistical model and intelligent forecasting model. Traditional statistical methods are usually based on strict assumptions, but their accuracy and objectivity are limited in the face of nonlinear data. In contrast, intelligent prediction model can effectively mine the internal relationship between complex data, has a strong self-learning ability, and can reduce the influence of subjective factors, so it has a significant advantage in prediction accuracy.

Based on the existing research, a new financial crisis prediction model combining particle swarm optimization (CSO) and generalized Extreme Learning Machine (KELM) is proposed in this paper. The model reduces the sensitivity of the model to parameter setting by optimizing the parameters of KELM, and carries out feature selection by CSO algorithm, fully considering the interaction between parameter optimization and feature selection. The design of this model aims to improve the accuracy and efficiency of financial crisis prediction and provide more accurate and reliable support for enterprise management decision.

ISSN 2616-5902 Vol. 6, Issue 11: 56-60, DOI: 10.25236/AJBM.2024.061109

2. Relevant Research

In the context of the current digital era, with the continuous progress of business intelligence technology, the decision-making process of the financial industry increasingly relies on big data analysis and machine learning technology to achieve accurate prediction of financial crisis [1]. By integrating financial and non-financial data, it can not only effectively evaluate the probability of enterprise bankruptcy, but also provide an important reference for the sustainable development of national economy. This approach shows important application prospects in dealing with global economic volatility and promoting stable growth.

In response to this challenge, the financial crisis prediction model [2] proposed by S Hong based on genetic algorithm (GA) makes full use of the advantages of the fusion of financial and non-financial data, and proves the high prediction accuracy of this model through specific case analysis. The research results significantly enhance the ability of enterprises to identify financial risks, and provide strong support for them to take timely countermeasures and risk management. At the same time, the multi-objective squirrel search optimization algorithm (MOSSA) developed by MA Duhayyim combined with the stacked autoencoder (SAE) model [3], through the user financial data processing process in the Internet of Things scenario, including data preprocessing, category imbalance processing, parameter optimization and classification analysis, etc. Effectively improve the efficiency of financial data cloud analysis and decision-making.

In the intelligent application scenario of the Internet of Things, SKS Tyagi developed a financial crisis prediction model [4] based on the quantum artificial butterfly optimization algorithm, which combines long short-term memory network (LSTM) and recurrent neural network (RNN) to classify financial data. After comprehensive testing and verification of the enterprise financial data collected in real time by iot devices, the model is superior to traditional methods in terms of sensitivity, specificity and accuracy, demonstrating its high efficiency in financial crisis prediction.

Z Liu proposed a financial crisis prediction model [5] combining financial indicators and annual report financial texts. This model extracts the features of financial indicators through convolutional neural networks (CNN), and analyzes the potential semantics of financial texts by using long short-term memory networks (LSTM) with attention mechanism. The model significantly improves the forecasting effect by integrating multiple data sources, and emphasizes the importance of considering financial text and financial indicators simultaneously in improving the accuracy of financial crisis detection.

In other studies, the financial risk assessment model [6] proposed by Vaiyapuri combines intelligent feature selection and deep learning technology, carries out feature selection through the water strider optimization algorithm, uses the deep random vector function to connect the network for classification, and uses the improved fruit fly optimization algorithm for hyperparameter optimization, which significantly improves the prediction performance of the model. This method effectively integrates financial and non-financial data and provides a new tool for more accurate financial crisis detection.

C Zhang used K-break random forest algorithm and time series analysis model to develop an enterprise financial crisis early warning system [7]. Through dynamic analysis of historical data, the model reached 89% prediction accuracy, which greatly improved the enterprise's early warning ability. In another study, Chen used logistic regression and BP neural network model [8], and optimized BP neural network by simulated annealing algorithm. The empirical results show that the simulated annealing algorithm can effectively improve the efficiency and prediction accuracy of the model, and provide solid technical support for the financial early warning mechanism of listed enterprises.

3. Theoretical Basis and Related Research of Financial Crisis Prediction Model

3.1 Optimization Research Based on Kernel Extreme Learning Machine and Adaptive Particle Swarm Optimization Algorithm

The algorithm generates a unique optimal solution by randomly initializing the weights of input layer nodes and the bias of hidden layer nodes. Assume that the ELM network contains n input nodes, K hidden nodes and m output nodes. For N different training samples, the input data and class vectors are represented in the form of multidimensional vectors. In the classification process, the output of ELM is a combination of weights and incentive functions of hidden layer nodes and output nodes, with the goal of minimizing the output error and ensuring that the error of any sample is close to zero. To

ISSN 2616-5902 Vol. 6, Issue 11: 56-60, DOI: 10.25236/AJBM.2024.061109

achieve this goal, the system generates a hidden layer feature mapping matrix based on the input data and keeps them unchanged by randomly selecting the input weights and bias, and only adjusts the output weights, thus obtaining a global optimal solution without adjusting all network parameters.

Through Karush-Kuhn-Tucker theory, the training of ELM is transformed into a dual problem, and the analytic solution of the output weights is finally obtained. When the hidden layer feature mapping is unknown, the kernel matrix is used to represent the hidden layer feature mapping, and the classification output only depends on the kernel function, independent of the number of hidden layer nodes. This means that there is no need to set the initial weight and bias of the hidden layer, and the kernel matrix is usually computed using Gaussian kernel functions. The classification performance of KELM is affected by the regularization coefficient C and the kernel function parameter γ, the combination of which directly determines the generalization ability of the model.

In order to optimize the parameters of KELM, adaptive particle swarm optimization algorithm is used to optimize the parameters. The K-fold cross-validation error is often used to estimate the generalization error and is an effective optimization criterion. The existing parameter optimization methods (such as grid search and gradient descent) are more sensitive to initial values, and with the increase of parameter space, the search efficiency decreases and it is easy to fall into local optimal. Particle swarm optimization (PSO), as a meta-heuristic search algorithm, can quickly search parameter combinations and obtain optimal classification performance. PSO realizes the information sharing of individuals in the group space through many iterations, so as to find the global optimal solution.

Using the particle algorithm, particles acquire a multidimensional space in physical form and adjust their position and speed based on parameters such as learning factors. As the repetition rate increases and the adaptability to mass weights is lost, the algorithm becomes more effective as an initial global particle search function and a strong subsequent control of the algorithm. Particle algorithm can coordinate global search and local search, and improve the speed and accuracy of the algorithm.

When selecting a function, a "particle reading" is used to obtain information from the separated surface. The optimal direction of binary particle algorithm is to represent the particle state with binary code and particle velocity. This method ensures the performance and possibility of spatial features and improves the efficiency of the particle selection algorithm.

3.2 Improve the Prediction Model

The corporate financial crisis forecasting model includes categories that are the main tools of the kelm model to define better parameters and functions for more accurate forecasts. Particle adaptation algorithms help to optimize parameters and reduce characteristic synchronization. The selection of partial groups of gamma-variables (c, gamma) is improved by the dissolution of concentration groups and continuous particles above 2%. Two discrete algorithms follow the kelm model variables, but focus on the selection of attributes.

Dissolved and dissolved to develop the best species conservation and use the best subtribute to predict the end of the sequence (c, gamma). The model summarizes the complexity of financial data and ensures the reliability and reliability of projections. To ensure model performance, k-foil sequence results are evaluated to confirm that training results are consistent with test results. The improved vonkel algorithm has significantly improved the accuracy and stability of financial crisis forecasting by correctly choosing identity and improving fundamental variables.

4. Experiment and Analysis of Financial Crisis Prediction Model

4.1 Data Collection and Analysis

This study is based on a dataset of corporate financial information, including records of the financial condition of multiple companies, both bankrupt and operating, over a specific period of time. The selected data sample covers the years leading up to the risk of insolvency, ensuring an accurate assessment of a company's financial health. A variety of financial indicators are involved in the data set, which provides a wide range of information support for in-depth research. Specific data can be found in the relevant tables.

In the data preprocessing stage, the research team pays attention to the cleaning and standardization of the original data to ensure the accuracy and consistency of the data. For the missing values and

ISSN 2616-5902 Vol. 6, Issue 11: 56-60, DOI: 10.25236/AJBM.2024.061109

abnormal data, the appropriate filling method is adopted, and the abnormal samples that deviate from the normal interval are eliminated to reduce the interference to the model analysis. In order to further optimize the prediction accuracy of the model, the research adopted a variety of feature screening methods to extract the most representative features from numerous financial indicators to ensure the efficiency and accuracy of the model.

Through systematic data processing and analysis, this paper constructs a financial crisis prediction model integrating financial and non-financial information. This model provides theoretical and empirical basis for in-depth understanding of corporate financial risk, provides direction for further research in related fields, and provides scientific support for corporate financial management and decision-making.

4.2 Model Experiment and Performance Evaluation

Before the start of the experimental analysis, the research team first preprocessed the data, especially in order to prevent the large differences in eigenvalues from negatively affecting the experimental results, the normalization of the data was prioritized. Through this step, all eigenvalues are adjusted to a uniform range, which not only helps to reduce the numerical differences between each index, but also effectively reduces the complexity of subsequent calculations. In order to evaluate the accuracy and stability of the classifier, the ten-fold cross-validation method is used in the experiment. Operationally, the entire data set is divided into ten equal parts, some of which are used as a test set and the rest as a training set. With ten iterations of the process, each subset can be used as a test set in turn, and finally different validation models are obtained and their classification accuracy is calculated, so that the average performance of the classifier can be synthesized.

In each experiment, the research team conducted random sampling in strict accordance with the proportion of crisis and non-crisis samples in the original data set to ensure the diversity of samples and the representativeness of experimental results. The whole implementation process of the experiment is carried out in the MATLAB environment, using the model of particle swarm optimization algorithm and extreme learning machine, and comparing the performance with other models such as genetic algorithm. The experimental parameters are set to use eight particles, perform 250 iterations, and search for the best values of C and γ within a reasonable exponential interval. At the same time, the speed parameters of PSO have been optimized and adjusted according to the characteristics of the algorithm to ensure the best search performance.

Through multiple rounds of experiments and comprehensive comparison of model performance, the experimental results show that the model combined with PSO algorithm and extreme learning machine is superior to other models in classification accuracy and convergence efficiency. After the feature selection strategy is introduced, the classification performance of the model is further improved, and the number of features to be processed is effectively reduced. Compared with the model without feature selection, the model with feature selection has obvious improvement in classification accuracy, Class I and Class II error rate. Through in-depth comparative analysis of different models, it can be seen that there are significant differences between the improved PSO-KELM model and GA-KELM model in many aspects of performance, especially in the case of fewer features, PSO-KELM can achieve higher classification effect. This shows that PSO algorithm has stronger global search ability in feature selection and optimization, and the anti-noise ability and stability of the model are significantly enhanced.

In order to further verify the superiority of the model, the experiment also conducted several double-tail T-tests to compare the significant differences in classification performance of different models. The results show that there are significant differences between the improved PSO-KELM and the model without feature selection in a number of performance indicators, and although there are no significant differences between the improved PSO-KELM and GA-KELM in some indicators, PSO-KELM performs better in feature selection effect on the whole. In order to visually demonstrate the convergence performance of the model, the research team recorded the changes in the fitness of the model during each iteration. The experimental results show that the fitness curve of the PSO-KELM model rises rapidly at the early stage of training, and becomes stable after reaching the highest classification accuracy, which fully indicates that the model is highly efficient in the process of feature selection and parameter optimization, and can quickly converge to the global optimal solution.

ISSN 2616-5902 Vol. 6, Issue 11: 56-60, DOI: 10.25236/AJBM.2024.061109

5. Conclusion and Prospect

In this paper, a financial crisis prediction model based on improved particle swarm optimization algorithm and extreme learning machine is proposed. By considering the interaction between parameter optimization and feature selection, a new objective function is designed, which significantly improves the classification performance of the model. The experimental results show that the proposed PSO-KELM model is superior to the traditional GA-KELM model and the PSO-KELM model without feature selection in classification accuracy and feature selection efficiency, which verifies its effectiveness in distinguishing enterprises in financial crisis. Future studies could further explore the combination of this model with other optimization algorithms, such as genetic algorithms and ant colony algorithms, to improve overall model performance. In addition, expanding the diversity of the data set and introducing industry characteristics and macroeconomic factors will help to improve the generalization ability of the model, thereby enhancing the prediction effect of financial risks in different contexts. Further, more complex pattern recognition and feature extraction combined with deep learning technology can enhance the decision support ability of the model. This study not only provides a new idea for the prediction of financial crisis, but also brings inspiration to the researchers in related fields, and the future research can be further developed and improved on this basis.

References

[1] Vodithala S, Bhukya R. A Novel Political Optimizer-Based Feature Selection with an Optimal Machine Learning Model for Financial Crisis Prediction [J]. International Journal of Cooperative Information Systems, 2024, 33(03): 56. DOI:10.1142/S021884302350020X.

[2] Hong S, Wu H, Xu X. Analysis of Financial Crisis Prediction Model Based on Genetic Algorithm[C]. The International Conference on Cyber Security Intelligence and Analytics. 2022.

[3] Duhayyim M A, Alsolai H, Fahd N. Optimized Stacked Autoencoder for IoT Enabled Financial Crisis Prediction Model [J]. Computers, materials & continua, 2022, 71(1 Pt.2): 1079-1094.

[4] Tyagi S K S, Boyang Q. An Intelligent Internet of Things aided Financial Crisis Prediction Model in FinTech [J]. IEEE Internet of Things Journal, 2021, (99):1-10.DOI:10.1109/JIOT.2021.3088753.

[5] Liu Z, Liu X, Zhou L. A Hybrid CNN and LSTM based Model for Financial Crisis Prediction [J]. Technical Gazette, 2024, 31(1): 185-192.

[6] Vaiyapuri T, Priyadarshini K, Hemlathadhevi A, et al. Intelligent Feature Selection with Deep Learning Based Financial Risk Assessment Model[J]. Computers, materials, and continuum, 2022, (008): 89. DOI:10.32604/cmc.2022.026204.

[7] Zhang C, Zhong H, Hu A. Research on Early Warning of Financial Crisis of Listed Companies Based on Random Forest and Time Series[J]. Mobile information systems, 2022, 22(Pt.1): 1573966-1573967. DOI:10.1155/2022/1573966.

[8] Chen Y. BP Neural Network Based on Simulated Annealing Algorithm Optimization for Financial Crisis Dynamic Early Warning Model [J]. Hindawi Limited, 2021. DOI:10.1155/2021/4034903.