

Improving Transformer for Enterprise Bankruptcy Prediction Based on Swarm Intelligence Optimization Algorithm

Mengli Wang

*School of Economics, Harbin University of Commerce, Harbin, China
3129981928@qq.com*

Abstract: This article explores the use of swarm intelligence optimization algorithms to optimize Transformer models and applies them to the study of enterprise bankruptcy prediction. Through experimental observation, we found that the predictive performance of the model gradually improves with the increase of iteration times. Specifically, the fitness value decreased from the initial 0.0368 to 0.0337 after 15 iterations, indicating the continuous optimization of the model during the training process. In the analysis of key factors in predicting corporate bankruptcy, retained earnings and total accounts receivable were identified as the two factors with the highest correlation with corporate bankruptcy, providing important basis for the selection of input features for the model. In terms of model performance evaluation, the prediction accuracy on the training set reached 100%, demonstrating the model's high fitting ability on known data. However, what is more important is the performance of the model on the test set, where 728 instances were correctly predicted, only 1 instance was incorrectly predicted as 'bankrupt', and 13 instances were incorrectly predicted as 'alive'. These results resulted in a prediction accuracy of 98.11% on the test set, with only a difference of 1.89% compared to the training set. This indicates that the model has good generalization ability and can be effectively applied to predict unknown data. The significance of this study lies in optimizing the Transformer model through swarm intelligence optimization algorithms, which not only improves the accuracy of enterprise bankruptcy prediction, but also enhances the model's generalization ability, providing more reliable risk assessment tools for financial institutions and investors.

Keywords: Swarm Intelligence Optimization Algorithm, Transformer, Enterprise Bankruptcy Prediction

1. Introduction

Enterprise bankruptcy prediction is an important topic in financial risk management, with the core goal of identifying potential bankruptcy risks in advance by analyzing the financial condition and operational performance of the enterprise. This research field originated in the 1930s [1]. With the rapid development of the global economy and intensified market competition, corporate bankruptcies have occurred frequently, causing huge losses to investors, creditors, employees, and the entire economic system [2]. Therefore, academia and industry have begun to focus on how to predict corporate bankruptcy through scientific methods to reduce related risks. Early research mainly relied on financial ratio analysis, such as Altman's Z-score model, but traditional methods have limitations in complexity and accuracy. With the advancement of data science and artificial intelligence

technology, machine learning algorithms have gradually become the mainstream tool for predicting corporate bankruptcy, injecting new vitality into this field [3].

Machine learning algorithms have shown significant advantages in predicting corporate bankruptcy. Firstly, machine learning is capable of processing massive amounts of data, including structured data (such as financial statements) and unstructured data (such as news texts, social media information), in order to capture more potential risk signals [4]. Secondly, machine learning algorithms have strong nonlinear modeling capabilities and can identify complex financial and business models, while traditional methods often rely on linear assumptions. In addition, machine learning models can achieve self optimization by continuously learning new data and adapting to dynamically changing market environments. These characteristics enable machine learning to demonstrate higher accuracy and robustness in predicting corporate bankruptcy [5].

In predicting corporate bankruptcy, commonly used machine learning algorithms include logistic regression, decision trees, random forests, support vector machines (SVM), and neural networks. Logistic regression has been widely used in early research due to its simplicity and interpretability; Decision trees and random forests are capable of processing high-dimensional data and capturing nonlinear relationships, making them suitable for analyzing complex financial features; SVM performs well on small sample data and is suitable for handling imbalanced datasets; Neural networks, especially deep learning models, can automatically extract features from large amounts of data and are suitable for high-dimensional, nonlinear prediction tasks [6]. In recent years, ensemble learning methods such as XGBoost and LightGBM have become popular choices for predicting corporate bankruptcy due to their high accuracy and efficiency. This article is based on swarm intelligence optimization algorithm to optimize transformer for predicting enterprise bankruptcy, and attempts to experimentally observe the predictive performance of the model.

2. Data set sources and data analysis

In terms of dataset selection, we chose an open-source dataset as the input for the experiment. The dataset is sourced from the Kaggle official website and covers the business operations of multiple companies, including their business indicators and final business performance (1 represents survival, 2 represents bankruptcy). The company's operating indicators are current assets, cost of goods sold, depreciation and amortization EBITDA, Inventory, net income, total accounts receivable, market value, net sales, total assets, total long-term debt, pre tax profit, gross profit, total current liabilities, retained earnings, total income, total liabilities, and total operating expenses. The dataset contains a total of 3708 pieces of data.

Select some data for display, and the results are shown in Table 1.

Table 1: Partial data.

X1	X2	X3	X4	X5	X6	X7	X8	status
511.267	833.107	18.373	89.031	336.018	35.163	128.348	372.7519	1
485.856	713.811	18.577	64.367	320.59	18.531	115.187	377.118	1
436.656	526.477	22.496	27.207	286.588	-58.939	77.528	364.5928	1
396.412	496.747	27.172	30.745	259.954	-12.41	66.322	143.3295	1
28.957	79.567	2.024	3.873	10.947	-0.138	15.89	6.1972	1
4424	15482	1092	2248	708	985	1134	9932.415	2
5179	17120	1202	2583	757	813	1303	5958.9688	2
6540	19419	1404	-456	822	-1762	1414	3445.0155	2
4937	18555	1366	-1256	627	-3511	1481	1030.1874	2
4682	14430	1377	582	516	-1228	796	2066.5869	2

Analyze the correlation between various indicators and corporate bankruptcy, as shown in Figure 1. According to the ranking of correlation coefficients, the factor with the highest correlation with corporate bankruptcy is retained earnings, followed by total accounts receivable.

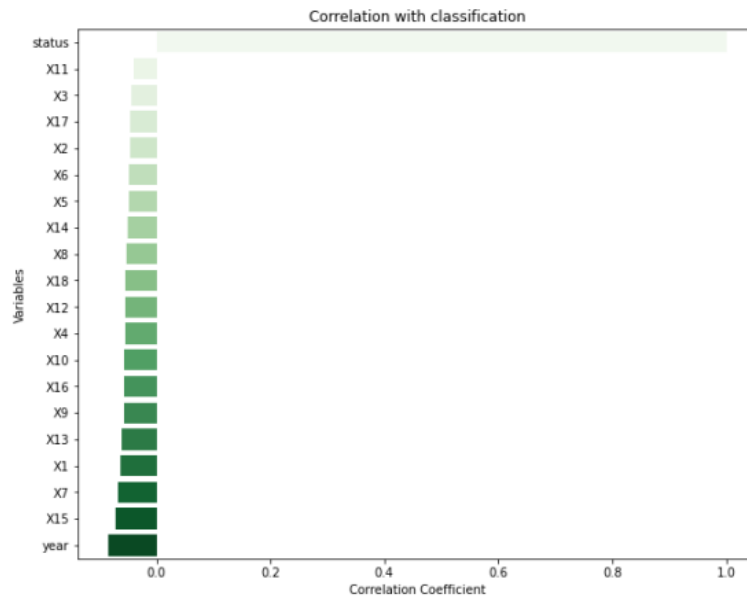


Figure 1: The correlation matrix.

3. Method

3.1. Multi strategy hybrid improved sparrow search algorithm

The Sparrow Search Algorithm (SSA) is a metaheuristic optimization algorithm based on the foraging behavior of sparrow populations. Its core idea is to solve optimization problems by simulating the foraging, alerting, and migration behavior of sparrows. However, traditional SSA may face limitations such as slow convergence speed and susceptibility to local optima when dealing with complex optimization problems [7]. To address these issues, the sparrow search algorithm with multi strategy hybrid improvement has emerged. This algorithm significantly improves its performance and robustness by introducing various strategies, such as dynamically adjusting parameters, blending the advantages of other optimization algorithms, and enhancing global search capabilities. The schematic diagram of the sparrow search algorithm is shown in Figure 2.

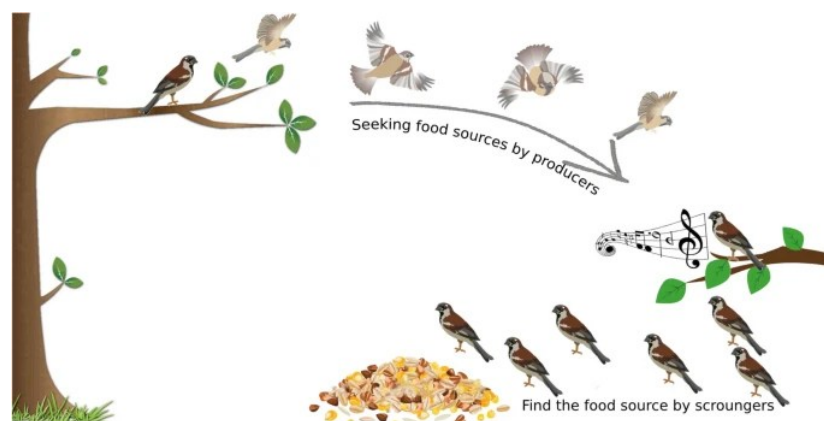


Figure 2: The schematic diagram of the sparrow search algorithm.

The core of the multi strategy hybrid improved sparrow search algorithm lies in combining multiple optimization strategies to balance the algorithm's global search and local development capabilities. For example, by introducing an adaptive parameter adjustment mechanism, the algorithm can dynamically adjust the search range and step size of sparrows according to the iteration process, thereby enhancing global exploration capability in the early stages and improving local development accuracy in the later stages. In addition, algorithms can also combine the advantages of other optimization algorithms, such as the speed update mechanism of particle swarm optimization (PSO) or the crossover mutation operation of genetic algorithm (GA), to enhance population diversity and search efficiency. The mixed use of these strategies enables the algorithm to better adapt to different types of optimization problems [8].

3.2. Transformer

Transformer is a deep learning model based on self attention mechanism, initially proposed for natural language processing tasks such as machine translation. Unlike traditional recurrent neural networks (RNNs) and convolutional neural networks (CNNs), Transformer completely abandons recursive and convolutional structures and instead relies on self attention mechanisms to capture global dependencies in the input sequence. The core idea is to dynamically allocate attention weights by calculating the correlation between each element in the sequence and other elements, in order to better model long-range dependencies. This design enables Transformer to perform well in processing long sequence data, while also having higher parallel computing efficiency [9]. The schematic diagram of Transformer's structure is shown in Figure 3.

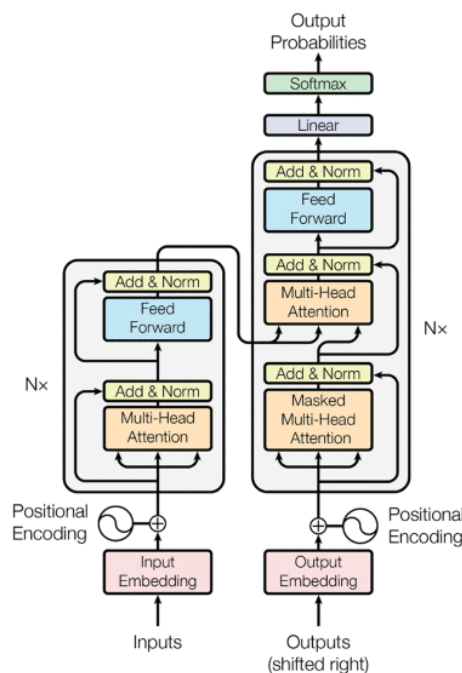


Figure 3: The schematic diagram of Transformer's structure.

The core component of Transformer is self attention mechanism. In the self attention mechanism, each input element (such as a word or pixel) generates a query vector, key vector, and value vector through linear transformation. By calculating the similarity between the query vector and the key vector, the model can determine the degree of attention each element has to other elements, and use these attention weights to weight and sum the value vectors, thereby generating a new representation. This mechanism enables the model to simultaneously consider the information of all elements in the

sequence without being limited by distance. In addition, Transformer adopts an encoder decoder structure: the encoder maps the input sequence to a series of high-dimensional representations, and the decoder generates the output sequence based on these representations. Each encoder and decoder layer includes a multi head self attention mechanism and a feedforward neural network, which enhances the model's expressive power by stacking multiple layers.

3.3. Transformer optimized for sparrow search algorithm

The Sparrow Search Algorithm (SSA) is a metaheuristic optimization algorithm based on the foraging behavior of sparrow populations. Its core idea is to search for the optimal solution by simulating the foraging, vigilance, and migration behavior of sparrows. Transformer is a deep learning model based on self attention mechanism, widely used in fields such as natural language processing and computer vision. The core goal of using sparrow search algorithm to optimize Transformer is to improve its hyperparameter settings or structural design through intelligent optimization techniques, thereby enhancing its performance. This combination utilizes the global search capability of SSA and the powerful modeling capability of Transformer, providing a new solution for complex optimization problems [10].

In the process of optimizing Transformer, sparrow search algorithm is mainly used to adjust the hyperparameters of the model (such as learning rate, layers, attention heads, etc.) or optimize the design of model components (such as the structure of attention mechanism). SSA initializes a set of sparrow individuals (i.e. candidate solutions), simulates the foraging behavior of sparrows, and gradually searches for the optimal combination of hyperparameters. The position of each sparrow represents a set of hyperparameter configurations, and its fitness is evaluated by the performance of the Transformer on the validation set, such as accuracy or loss value. By iteratively updating the position of sparrows, the algorithm can find the hyperparameter combination that maximizes the performance of the Transformer. In addition, the alert behavior of SSA can help avoid falling into local optima, thereby finding global optimal solutions in complex hyperparameter spaces.

4. Result

When conducting classification experiments using Transformer optimized based on sparrow search algorithm, the first step is to set the parameters. The Transformer model has 6 layers, 8 attention heads, 512 hidden layer dimensions, 2048 feedforward neural network dimensions, a learning rate of 0.0001, a batch size of 32, and 50 training rounds. The sparrow search algorithm has a sparrow population size of 20, a maximum iteration count of 100, a discoverer ratio of 0.2, an alert ratio of 0.1, and an initial step size factor of 1.0 that linearly decreases to 0.1 with each iteration. The fitness function uses the classification accuracy on the validation set and optimizes hyperparameters such as learning rate, number of layers, and attention heads of the Transformer through SSA to achieve optimal performance. In terms of hardware, this experiment uses a 3060 graphics card with 32GB of memory, and in terms of software, we use Matlab R2022a.

Introduce the model for training, first output the fitness change curve of the model, as shown in Figure 4. The fitness curve reflects the changes in model fitness as the training process progresses. According to the fitness curve, the fitness increased from the initial 0.0368 to 0.0337 after 15 iterations, and the predictive performance of the model gradually improved.

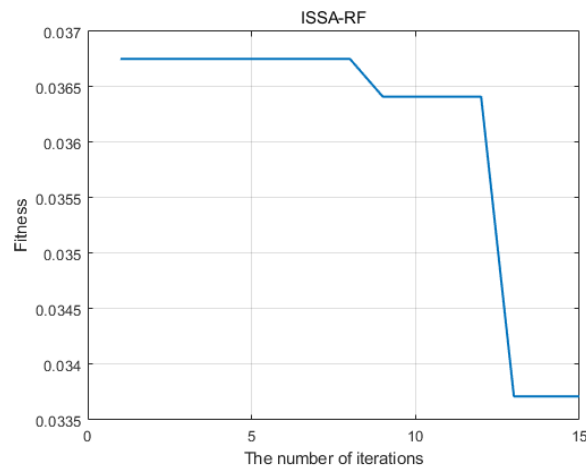


Figure 4: The fitness change curve of the model.

In terms of results, the output model predicts the bankruptcy of the enterprise, and the prediction results are classified into the prediction results of the training set and the test set. The confusion matrix predicted by the training set is shown in Figure 5, and the confusion matrix predicted by the test set is shown in Figure 6.

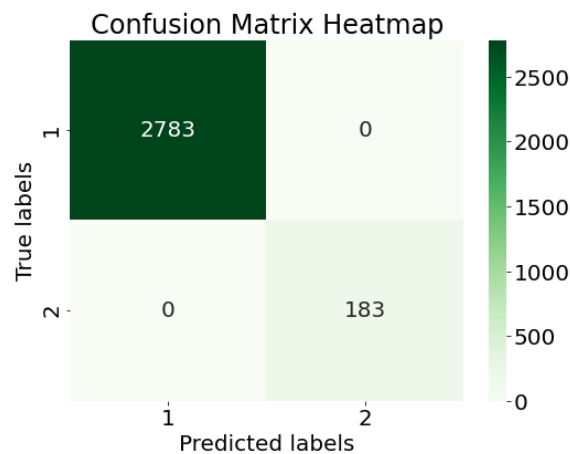


Figure 5: The confusion matrix.

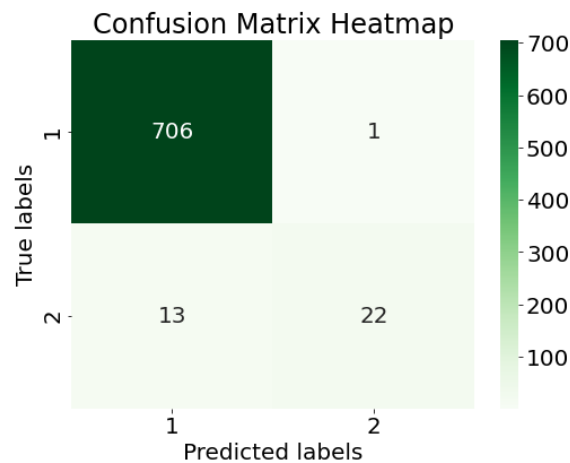


Figure 6: The confusion matrix.

According to the prediction results of the training set, the prediction accuracy is 100%. According to the prediction results of the test set, 728 instances were predicted correctly, 1 enterprise that should have been predicted as "survival" was predicted as "bankruptcy", and 13 enterprises that should have been predicted as "bankruptcy" were predicted as "survival". The prediction accuracy of the model in the test set is 98.11%, which is 1.89% lower than that in the training set. The model has good generalization ability.

5. Conclusion

This study uses swarm intelligence optimization algorithms to optimize the Transformer model and applies it to the task of predicting corporate bankruptcy. Through experimental observation, the predictive performance of the model has significantly improved. According to the correlation coefficient analysis, the factor with the highest correlation with corporate bankruptcy is retained earnings, followed by total accounts receivable. The identification of these key features provides important basis for the optimization of the model. During the model training process, the fitness curve showed that the model's fitness decreased from an initial 0.0368 to 0.0367 after 15 iterations, indicating a gradual improvement in the model's predictive performance.

On the training set, the model achieved a prediction accuracy of 100%, demonstrating extremely high fitting ability. However, the high accuracy of the training set is not sufficient to fully demonstrate the practicality of the model, so we further validated it on the test set. The results of the test set showed that the model correctly predicted 728 instances, with only 1 enterprise that should have been predicted as 'surviving' 'being misjudged as' bankrupt', and 13 enterprises that should have been predicted as 'bankrupt' 'being misjudged as' surviving'. However, the prediction accuracy of the test set still reached 98.11%, which is only 1.89% lower than the accuracy of the training set. This indicates that the model has strong generalization ability and can handle unknown data well.

The significance of this study lies in optimizing the Transformer model through swarm intelligence optimization algorithms, which not only improves the accuracy of enterprise bankruptcy prediction, but also provides new technical means for financial risk warning. The experimental results show that the optimized model performs well on both the training and testing sets, especially with high accuracy on the testing set, further verifying the reliability of the model in practical applications. In the future, this model is expected to be widely applied in the financial field, helping companies and financial institutions identify potential risks earlier and take effective measures to avoid losses.

References

- [1] Ghazalat, Anas, and Said AlHallaq. "Predicting and assessing bankruptcy risk: the role of accounting conservatism and business strategies." *Journal of Financial Reporting and Accounting* (2024).
- [2] Xu Z, Gong Y, Zhou Y, et al. Enhancing kubernetes automated scheduling with deep learning and reinforcement techniques for large-scale cloud computing optimization[J]. *arXiv preprint arXiv:2403.07905*, 2024.
- [3] Yu K, Bao Q, Xu H, et al. An extreme learning machine stock price prediction algorithm based on the optimisation of the Crown Porcupine Optimisation Algorithm with an adaptive bandwidth kernel function density estimation algorithm[C]//*Proceedings of the International Conference on Digital Economy, Blockchain and Artificial Intelligence*. 2024: 116-121.
- [4] Wang Z, Chen Y, Wang F, et al. Improved Unet model for brain tumor image segmentation based on ASPP-coordinate attention mechanism[C]//*2024 5th International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE)*. IEEE, 2024: 393-397.
- [5] Liu H, Xu Y, Huang J, et al. Surface Defect Detection Based on Machine Learning and Data Augmentation[C]//*2024 IEEE 6th International Conference on Civil Aviation Safety and Information Technology (ICCASIT)*. IEEE, 2024: 1098-1103.
- [6] Rashid, Umra, et al. "Unravelling trends, patterns and intellectual structure of research on bankruptcy in SMEs: A bibliometric assessment and visualisation." *Heliyon* (2024).

- [7] Turunen, Riina. "Bankrupt widows. Gendered features of merchant business and bankruptcy in late pre-industrial society." *Women's History Review* (2024): 1-23.
- [8] Liu, Jianmin, et al. "Digital economy development and corporate bankruptcy risk: Based on the perspective of institutional isomorphism." *Emerging Markets Finance and Trade* 60.4 (2024): 793-807.
- [9] Islam, Jahirul, et al. "Cognitive Modelling of Bankruptcy Risk: A Comparative Analysis of Machine Learning Models to Predict the Bankruptcy." *2024 12th International Symposium on Digital Forensics and Security (ISDFS)*. IEEE, 2024.
- [10] Gangwani, Divya, and **ngquan Zhu. "Modeling and prediction of business success: a survey." *Artificial Intelligence Review* 57.2 (2024): 44.