Financial Crisis Prediction Based on GWO-SVM

Sampling from the Chinese Environmental Protection Industry

Jian Ke and Shiqian Yu(✉)

Newhuadu Business School, Minjiang University, Fuzhou 350100, Fujian, China
{kejian,yushiqian_2022mba}@nbs.edu.cn

Abstract. Financial Crisis Prediction (FCP) is an important initiative to prevent the outbreak of financial crisis in enterprises, which is significant to the safe operation and economic stability of enterprises. To improve the prediction accuracy of the occurrence of corporate financial crisis, a financial crisis prediction model based on the optimized support vector machine of the Grey Wolf Optimization Algorithm is proposed. This paper firstly introduces the basic principles of SVM and GWO; then proposes the SVM model based on the penalty parameters C and g optimization; and finally compares the prediction performance of the environmental protection industry by different machine learning methods, taking them as examples. The results show that GWO- SVM can more accurately predict the likelihood of corporate crises. As can be seen the model has high application prospects.

Keywords: GWO-SVM · Machine Learning · Financial Crisis Prediction · The Grey Wolf Optimization Algorithm

1 Introduction

Preventing and resolving major financial risks is an important part of the three major battles. Under the new normal of continued global economic shocks and an increasingly complex business environment, financial risks and business risks for enterprises continue to rise. The environmental protection industry, as one of the seven strategic emerging industries, will bring huge losses to investors and the country in case of a financial crisis. Therefore, it has become a hot research topic to issue early warning before a financial crisis occurs in an enterprise.

Corporate financial crises are highly uncertain, nonlinear, and catastrophic, and research scholars have used various methods to study this problem, such as univariate statistical algorithm [1], conditional probability regression model [2], XGBoost algorithm [3], Decision Tree model [4], Neural Network [5], etc. However, in the face of high-dimensional and nonlinear financial crisis, a single model is more difficult to achieve the desired prediction effect. The results of studies show that a reasonable combination of models can significantly improve the prediction accuracy of the models.

Based on the existing research results, Support Vector Machine (SVM), which has superior generalization ability for classification and high-dimensional problems, was

K. Hemachandran et al. (Eds.): ICAID 2023, AHIS 9, pp. 535–543, 2023.
https://doi.org/10.2991/978-94-6463-222-4_58

selected as a prediction model [6]. When using SVM for enterprise FCP, its prediction performance depends to a large extent on the selection of parameters. There are existing parameter optimization algorithms such as Genetic Algorithm [7], Particle Swarm Optimization method [8] and Harris Hawk's Optimization algorithm [9]. It has been experimentally demonstrated that the Grey Wolf Optimization Algorithm can obtain the global optimal solution in a shorter time compared with the above optimization algorithms, so Grey Wolf Optimization Algorithm is chosen to optimize the SVM model in this paper.

2 Basic Theory

2.1 Principles of Support Vector Machine

SVM is a supervised machine learning method with good robustness proposed by Vapnik et al. in 1995. The essence is to search for an optimal hyperplane that can classify the samples efficiently. For linearly indistinguishable samples, SVM maps the samples to a higher dimensional space through a nonlinear mapping Φ, which is transformed into a linearly divisible problem in the new space, at this point the optimal hyperplane can be found. As we can see this in Fig. 1.

When the sample X belongs to the d-dimensional real space and y is between -1 and 1, the optimal hyperplane problem can be solved as (1). Introducing the slack variable ξ and $\xi i \geq 0$, the objective function is transformed into (2), where C is the penalty factor for classification errors. Finally, Lagrange optimization is introduced to transform the optimal hyperplane problem into a pairwise quadratic programming problem, at which point the final optimal classification surface function is denoted as (3).

$$f(x) = w^T \cdot x + b \tag{1}$$

$$min\frac{1}{2}\|w\|^2 + C\sum\nolimits_{i=1}^{l}\left(\xi + \xi^*\right) \tag{2}$$

$$f(x) = sgn\left\{\sum\nolimits_{i=1}^{n}\alpha_i \cdot y_i(x \cdot x_i) + b\right\} \tag{3}$$

As can be seen, the key to the SVM-based FCP model is to find the penalty parameters C and g, so the internal parameters (C, g) are optimized using the GWO.

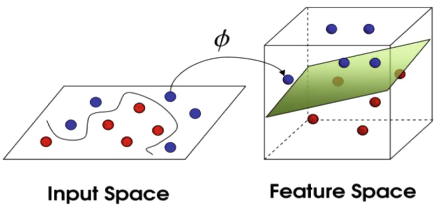

Fig. 1. Linear uncorrelated optimal hyperplane

2.2 Principles of Grey Wolf Optimization

Grey Wolf Optimization (GWO) is a pack intelligence optimization algorithm proposed by Mirjalil et al. to simulate the social hierarchy and hunting process of grey wolves. As we can see in the Fig. 2, the basic idea is as follows: firstly, a pack of grey wolves is randomly generated, and the pack is divided into four ranks according to the social hierarchy, among which α, β and δ wolves are the first, second and third ranks respectively, and ω belongs to the lowest rank; the hunting process is guided by α, β and δ , and ω constantly updates its position according to these three grey wolves; finally, the hunting of prey is realized by gradually shortening the distance from the prey. The optimization process mainly includes three steps for encircling, hunting, and attacking.

2.2.1 Surrounding

The first step of hunting is to surround the prey, and the mathematical model of the process is (4)–(7).

$$D = \left| C \cdot X_p(t) - X(t) \right| \tag{4}$$

$$X(t+1) = X_p(t) - A \cdot D \tag{5}$$

$$A = 2a \cdot r_1 - a \tag{6}$$

$$C = 2 \cdot r_2 \tag{7}$$

where t is the number of iterations; a is the convergence factor, which decreases from 2 to 0 as t increases; r1 and r2 take random values from 0 to 1; D denotes the distance between the wolves and the prey; X(t + 1) is the wolves updating their position according to the prey position and the distance from the prey.

2.2.2 Hunting

The second step of hunting is to hunt the prey. Assuming that α, β and δ is the global optimal solution, ω wolves update their positions according to α, β, δ and keep approaching the prey. The expression of wolf position update in this phase is (8) to (10), X(t + 1) is

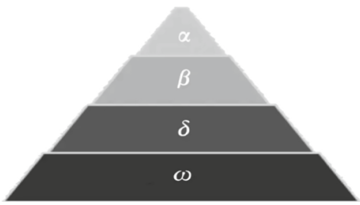

Fig. 2. Grey Wolf Population Level Map

the potential optimal solution after the update.

$$\begin{cases} D_\alpha = |C_1 \cdot X_\alpha - X| \\ D_\beta = |C_2 \cdot X_\beta - X| \\ D_\delta = |C_3 \cdot X_\delta - X| \end{cases} \tag{8}$$

$$\begin{cases} X_1 = X_\alpha - a_1 \cdot D_\alpha \\ X_2 = X_\beta - a_2 \cdot D_\beta \\ X_3 = X_\delta - a_3 \cdot D_\delta \end{cases} \tag{9}$$

$$X_{(t+1)} = \frac{X_1 + X_2 + X_3}{3} \tag{10}$$

2.2.3 Attacking

The last step of hunting is attacking to achieve the prey. This is accomplished by the decreasing convergence factor of a, and as the number of receptions increases a decrease from 2 to 0. From (6), A is determined by a. When $|A| > 1$, the wolves expand the search range to facilitate global search; when $A < 1$, the wolves keep getting closer to the prey in the iterations and finally achieve the capture.

3 Build FCP Model Based on GWO-SVM

3.1 Design Model

After reviewing the existing literature, analyzing the existing FCP models and further improving the parameters, A SVM financial crisis prediction model based on the optimization of the Grey Wolf Optimization Algorithm is proposed. Its design features are:

(1) With the good generalization ability of SVM, we can find the global optimal solution in high-dimensional and nonlinear sample data faster and thus improve the prediction ability of the model.
(2) The Grey Wolf Optimization Algorithm is used to perform a global search for the penalty parameter C and the kernel parameter gamma of the SVM, and the optimal solution is approximated by continuous iterations. Therefore, the GWO-SVM model is used in this model.

3.2 Build FCP Model

The construction process of the FCP model based on GWO-SVM is shown in Fig. 3. First, the collected data are divided into training and test sets and normalized. Secondly, the training set is used for model training and the optimal parameters are found by the Grey Wolf Optimization Algorithm. Finally, the optimized model is used to predict the test set. The model was run using PYTHON 3.9 software.

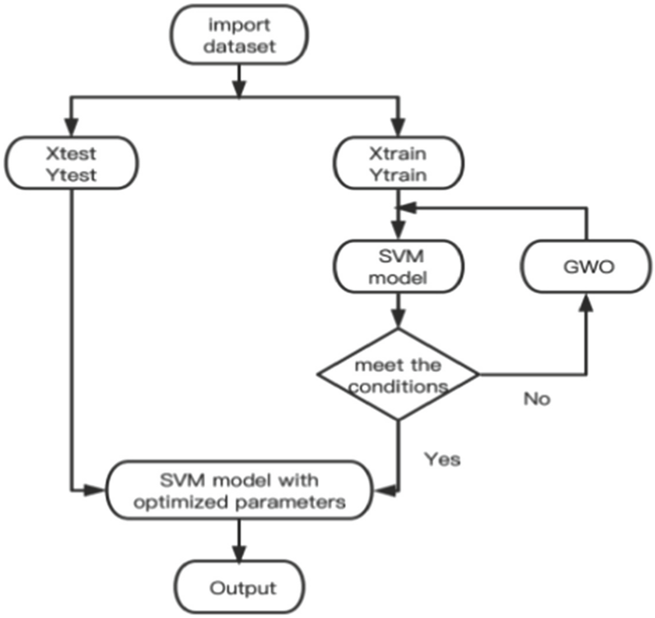

Fig. 3. Workflow of GWO-SVM prediction model

4 Model Selection and Data Processing

4.1 Data Source

The data used in this paper are from the CSMAR database, and the A-share listed companies in the environmental protection industry from 2019 to 2021 are used as the research object, and t-3, t-4, and t-5 going backward are identified as the crisis samples based on whether they are ST or *ST in that year as the discriminant criteria. The control samples were selected according to the principle of comparability, and finally obtained 40 samples for crisis companies and 59 control samples for normal companies.

4.2 Data Preprocessing

(1) When the number of missing values does not exceed 10% of the total number, the mean value is used to fill in the missing values.
(2) Significance tests were performed on alternative indicators using SPSS 27.0. Indicators that are more explanatory and representative were selected.
(3) The data is normalized before the neural network model is run, and the support vector machine data needs to be normalized to improve the efficiency of training.

4.3 Model Selection

To verify the superiority of GWO-SVM, Decision Tree, Random Forest, Super Vector Machine, Neural Network, and GWO-SVM are selected for comparison in this paper.

Table 1. Comparison of FCP Results

model	***Accuracy***	***Precision***	***Recall***	***F1***
DT	*65%*	*66.67%*	*60%*	*63.16%*
NN	*70%*	*75%*	*37.5%*	*50%*
FR	*75%*	*85.71%*	*60%*	*70.59%*
SVM	*85%*	*83%*	*70%*	*82.35%*
GWO-SVM	*90%*	*88.89%*	*88.89%*	*88.89%*

Among them, the DT model combines k-fold cross-validation and normal grid search to get the best parameter ccp_alpha of 0.121 for the model; FR controls the number of trees within 500 and squares the feature variables to improve the accuracy; SVM model selects the optimal rbf Gaussian kernel function by comparing the results of multiple kernel functions, and also uses grid search to find the best parameter; builds a multilayer neural network with early_stopping to prevent overfitting; the population size of the GWO-SVM model is 20, the maximum number of iterations is 20, and the parameters of the SVM range from 0.01 to 10.

5 Results and Analysis

Table 1 shows the comparison of accuracy, precision, recall, and F1 score of each model. As can be seen indicators of the SVM model among the four single models have improved compared to DT (Decision Tree), RF (Random Forest), and NN (Neural Network) model, indicating that SVM (Support Vector Machine) is more effective in crisis prediction in the environmental protection industry. Comparing the optimized SVM with SVM results using the Grey Wolf Optimization Algorithm, As can be seen that GWO- SVM is more effective in prediction. In addition, GWO- SVM do well perform in recall rate, which indicates that GWO- SVM is more capable of identifying crisis companies.

The ROC is used to measure the classification performance of the model, and the closer the ROC curve is to the upper left corner, the better the performance of the model. From Fig. 4, GWO-SVM has the highest classification accuracy relative to the other models. Moreover, a combination of Fig. 5 and 6 shows that the GWO- SVM model significantly reduces the prediction error rate of the model on both the training and validation sets while improving the running speed. The results show that the Grey Wolf Optimization Algorithm can indeed effectively improve the early warning capability for the occurrence of financial crises in the environmental protection industry, thus making the GWO-SVM have a strong fitting capability.

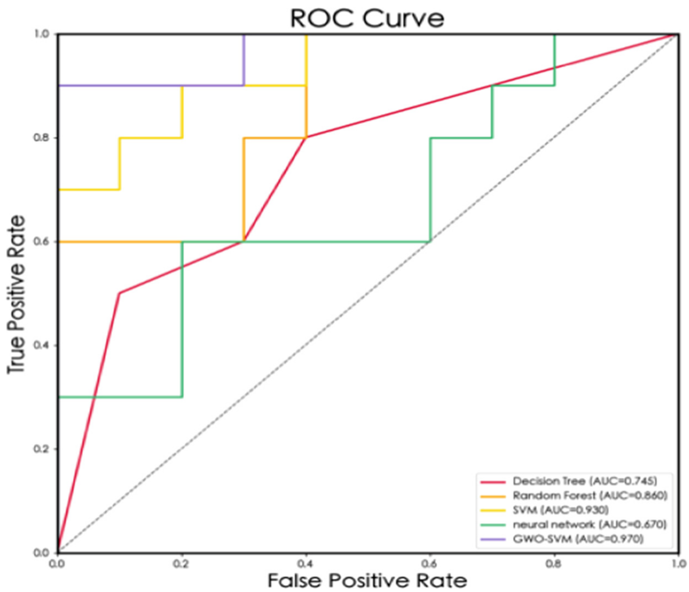

Fig. 4. ROC Curves

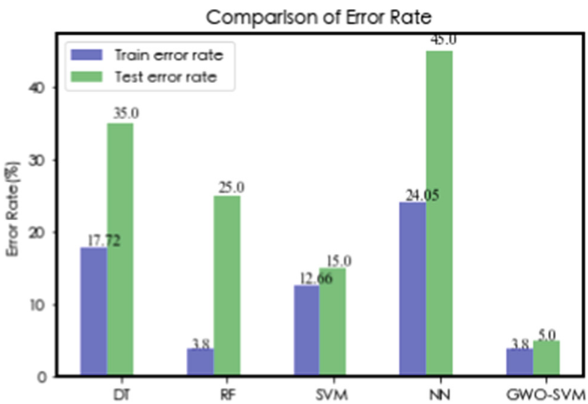

Fig. 5. Comparison of Error Rate

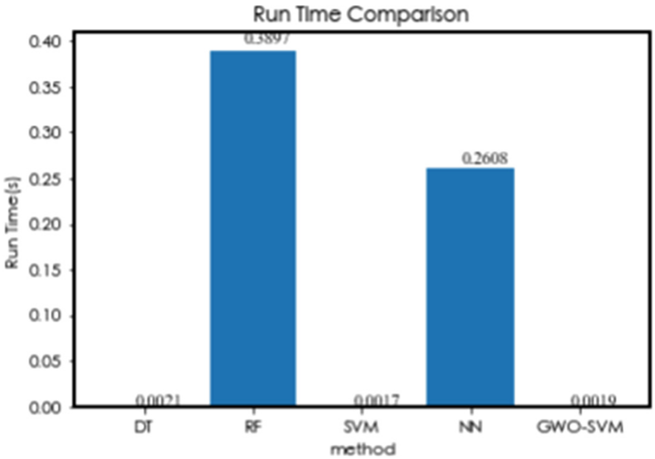

Fig. 6. Run Time Comparisons

6 Conclusions

To improve the accuracy of predicting the financial distress of listed companies in the environmental protection industry, this paper establishes a FCP index system with matching characteristics of the environmental protection industry and applies the Grey Wolf Optimization Algorithm to the optimization of parameters C and gamma of the SVM model. The empirical results show that the GWO-SVM model can predict the financial distress of enterprises more accurately and has good application prospects.

References

1. Beaver, William H. "Financial ratios as predictors of failure." Journal of accounting research (1966): 71-111.
2. Ohlson, James A. "Financial ratios and the probabilistic prediction of bankruptcy." Journal of accounting research (1980): 109-131.
3. Carmona, Pedro, Francisco Climent, and Alexandre Momparler. "Predicting failure in the US banking sector: An extreme gradient boosting approach." International Review of Economics & Finance 61 (2019): 304-323.
4. Chen, Mu-Yen. "Predicting corporate financial distress based on integration of decision tree classification and logistic regression." Expert systems with applications 38.9 (2011): 11261-11272.
5. Aydin, Nezir, et al. "Prediction of financial distress of companies with artificial neural networks and decision trees models." Machine Learning with Applications 10 (2022): 100432.
6. Ahn, Jae Joon, et al. "Usefulness of support vector machine to develop an early warning system for financial crisis." Expert Systems with Applications 38.4 (2011): 2966-2973.
7. Wu, Chih-Hung, et al. "A real-valued genetic algorithm to optimize the parameters of support vector machine for predicting bankruptcy." Expert systems with applications 32.2 (2007): 397-408.

8. Zhang, Guodao, et al. "Optimization of energy consumption of a green building using PSO-SVM algorithm." Sustainable Energy Technologies and Assessments 53 (2022): 102667.
9. Samantaray, Sandeep, Abinash Sahoo, and Deba Prakash Satapathy. "Improving accuracy of SVM for monthly sediment load prediction using Harris hawks optimization." Materials Today: Proceedings 61 (2022): 604-617.