Research Article

Research on Agricultural Economic Early Warning Based on Genetic Algorithm and SVM

Xueyong Bian,1 Xuzi Lv,2 and Jie Tian3

1School of Management Science and Engineering, Guizhou University of Finance and Economics, Guiyang 550000, China
2School of Education, Guizhou Normal University, Guiyang 550000, China
3School of Economics, Hebei GEO University, Shijiazhuang 050000, China

Correspondence should be addressed to Xuzi Lv; bxy3729230@mail.gufe.edu.cn

Received 29 March 2022; Accepted 28 April 2022; Published 7 June 2022

Academic Editor: Han Wang

Copyright © 2022 Xueyong Bian et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Agriculture is unique in that the industry is subject to a certain level of uncertainty due to seasonal and other factors, and its risk level is significantly higher than that of other industries. Therefore, it is necessary to establish an appropriate financial early warning model to predict, analyze, and control its financial risks. The article uses a genetic algorithm and support vector machine-based economic forecasting model for agribusinesses to adapt its own pollutant weights in a practical application environment to improve the forecasting accuracy. This model first uses a genetic algorithm to train a feature weight vector of current data so that the weights are adapted to the current prediction problem and then uses this feature weight vector to apply to SVM model training. It is found that 62.79% of the listed agricultural companies are in warning status, and their development is not optimistic. The article provides new ideas for the subsequent research on financial warning methods and also expands the boundaries of theoretical research for the research system of financial warning problems and enriches the experience and evidence of practical research.

1. Introduction

Since China’s social economy is inextricably linked with agricultural development, agricultural development has also received national attention. At the current level of development, the number of listed green agricultural companies is increasing, which is an important help to China’s social and economic development [1]. However, there are still more shortcomings in the development process, especially, many listed green agriculture companies do not clearly carry out investment planning in order to obtain more economic benefits; so, there are more problems in financial management, leading to financial risks [2]. Although opportunities and risks coexist, as long as green agricultural listed companies can establish risk early warning mechanism and use it to do a good job in financial management, they can avoid the impact caused by financial risks and can also effectively solve financial risks when they occur, thus effectively reducing economic losses [3].

In the past 10 years, there are 547 listed companies in Shanghai and Shenzhen A-shares that have been specially handled by SFC, among which there are 17 listed agricultural companies, such as Xiangli, Nongfa Seed, Pingtan Development, Zhonglu B, and ST Jinggu [4]. In addition, in recent years, financial fraud cases of listed agricultural companies have been frequent, such as the well-known “scallop exodus” financial fraud of the Swertia Island Group, the financial fraud case of Wanfushengke (now renamed as Jiawo), the abnormal trading case of Pingtan Development, and the illegal disclosure cases of [3]. All the above cases are staggering, they all show that the actual financial situation of listed agricultural companies is not good, and some companies are even on the verge of bankruptcy [4]. Therefore, it is crucial to analyze the reasons for the formation of financial crises in agricultural listed companies and construct a financial early warning model in line with them, so that the possible financial risks can be predicted and controlled [5].
Foreign scholars have made a great breakthrough in the research methods for the construction of financial early warning models [6]. The earliest research on financial early warning models using statistical methods was conducted by [7], and the study proposed a univariate model. On the basis of their research, [8] proposed a more improved univariate model, which is simple and easy to use, but its discriminative accuracy is not high. Several other scholars applied the logit model and its improved model to financial distress studies, which provided the theoretical basis for subsequent studies [9]. On the other hand, Nagar subramanian et al. [8] and Yu [9] used the research method of AdaBoosted decision tree and combinatorial model to study the determinants affecting financial distress. It is evident from the above literature that although foreign scholars have conducted a large number of constructive studies on financial early warning methods, there are fewer studies that combine these theoretical methods with applications in particular industries [10]. Most of the domestic scholars have improved the existing models based on foreign scholars’ research or compared the existing models and applied them to the financial early warning of Chinese listed companies. Some scholars have used logistic regression models to conduct early warning analysis on the financial situation of listed companies [11]: conducted an early warning analysis of the financial crisis of listed companies by establishing a multiple probability ratio regression model (Probit). [12] used the additional momentum method, conjugate gradient method, and L-M optimization method to construct a prediction model for the financial crisis of listed companies [13]. The agricultural industry has certain special characteristics in terms of operational risk compared with other industries, and most of the existing studies on the establishment and comparison of financial early warning models for listed agricultural companies are limited to traditional methods (e.g., univariate models, multivariate models, logit models, and artificial neural network models) and rarely use more advanced methods combined with computer networks (e.g., unit learning framework models, rough set models, decision tree models, gradient advancement models, and fuzzy OSVR methods) to conduct empirical studies [14]. For example, [15] used univariate analysis and multivariate analysis to study the financial distress of agricultural listed companies, and [16] conducted a financial warning study on agricultural listed companies based on survival analysis method. [17] illustrated how traditional accounting indicators can be applied to financial early warning of agricultural listed companies. [18] constructed a logistic regression analysis model based on EVA theory and also conducted an early warning on the financial crisis of agricultural listed companies.

2. Fundamentals and Prediction Model

2.1. Basic Theory of SVM. Support vector machine is a new type of generalized classifier based on statistical learning theory. It is widely used in various supervised learning scenarios because it minimizes structural risk, has good generalization ability, and can make samples linearly indistinguishable in the low-dimensional input space by mapping them to the higher space after introducing kernel functions. In recent years, many studies have applied SVM to various prediction analysis problems, and its prediction ability is even better than that of BP neural network methods and RBF neural networks when performing small sample data prediction.

Let sample \((x_1, y_1), (x_2, y_2), \ldots, (x_k, y_k) \in \mathbb{R}^n \times R\), where \(x_i\) is the input data, \(y_i\) corresponds to the label, and \(k\) is the total number of data; then, its minimization objective function can be expressed as

\[
R(\omega) = \min_{\omega, b, \xi} \left[ \frac{1}{2} \omega^T \omega + C \sum_{i=1}^{n} \xi_i \right],
\]

s.t. \[
\omega^T \phi(x_i) + b \geq 1 - \xi_i, \\
\xi_i \geq 0, i = 1, \ldots, n,
\]

where \(C\) is a penalty factor balancing the empirical risk of the model with model complexity, \(\xi\) is a nonnegative relaxation variable, and \(\phi(x_i)\) is a function related to the kernel function \(K(x_i, x_j) = \phi(x_i)^T \phi(x_j)\).

The above optimization problem is transformed into a dyadic problem by the Lagrangian method.

\[
\min \left[ \frac{1}{2} \alpha^T Q \alpha - e^T \alpha \right],
\]

s.t. \[
\alpha^T = 0, \\
0 \leq \alpha_i \leq C, i = 1, \ldots, n.
\]

Finally, the decision function can be obtained:

\[
\text{sgn} \left( \sum_{i=1}^{n} y_i \alpha_i K(x_i, x) + \rho \right).
\]

When the model is predicted with the input data \(x\), the corresponding prediction can be calculated by the above equation.

2.2. Genetic Algorithm Optimized SVM Model. In the field of machine learning for data mining, high-dimensional data usually requires feature selection for dimensionality reduction to avoid subjecting the model to dimensional disasters, which makes feature selection an important part of the data preprocessing step of machine learning algorithms.

The genetic algorithm is an iterative optimization algorithm that simulates the evolutionary patterns of species. It optimizes the candidate solution by simulating the mutation and reproduction of biological populations. Since genetic algorithms have the feature of global optimization search, they are widely used in various fields. The flow chart of the genetic algorithm is shown in Figure 1.

In optimizing the SVM model using the genetic algorithm, by selecting a vector \(v = (v_1, v_2, v_3, \ldots, v_m)\) consisting of deflators for each dimension as the population individuals, the training set data is trained with the feature
weight vector \( v \) deflated data \( x' = x \times v^T \) supplying the SVM model, and its F1 score on the validation set is used as the individual fitness [19, 20].

The process of optimizing the feature weight vectors by the genetic algorithm is to adapt the SVM model to the characteristics of the validation set data distribution. When the optimization results are put on the test set for validation, the performance of the model on the test set will degrade compared to the validation set due to overfitting because the validation set data distribution is not exactly the same as the test set data distribution. This degradation itself is due to overfitting and can be suppressed by adjusting the population size, number of iterations, and degree of variation parameters of the genetic algorithm [21–23].

In summary, the complete process of training an agribusiness economic forecasting model based on genetic algorithm and SVM using the original agribusiness economic data \( x \) and national three-level agribusiness economic criteria is shown in Figure 2.

### 3. Experimental Results and Analysis

The experiments are based on more than 1,000 historical data from agricultural enterprises, and the predictions are made separately for the agribusiness economies of interest to the plant and compared with the results predicted by traditional SVM [24].

Since the overfitting problem of the model can be mitigated by increasing the number of data in the validation set and ensuring that the data distribution of the validation set is close to the situation when it is actually applied (test set), the data were divided into training set, validation set, and test set in the ratio of 3:5:2 when segmenting the data [25, 26].

Given a population size of 10 in the genetic network and a variation range of 0 to 0.1, the appropriate number of iterations was determined by testing and observing the fitness on the validation set and the fitness on the test set with the number of iterations. The effect graph obtained from the experiment is shown in Figure 3.

The fitness of the validation set in Figure 3 increases with the number of evolutionary generations in a small oscillation and reaches a stable value around 40 generations, while the fitness of the test set reaches its maximum around the 23rd generation. The optimal number of iterations is conservatively taken to be 20 in order to avoid the effect of randomness of the experiment [27].
Since the agribusiness economy was in line with the agribusiness economy criteria most of the time except for a few cases, the data showed a tendency to be unbalanced, and the labeling balance of this data according to the four prediction objectives is shown in Table 1.

Category imbalanced samples are prone to model overfitting and also prone to higher model correctness and lower recall and accuracy, which has no practical meaning. For example, if there are 100 positive cases and 900 negative cases in the data, the model is trained to always return new samples to predict the results as negative cases; although, the correct rate is as high as 90%, but there is no reference value for practical problems. There are 2 main types of methods to solve the category imbalance problem: undersampling and resampling. Undersampling makes the number of positive and negative cases close to each other by removing the majority of samples, but it is easy to result in too few samples and overfitting; resampling balances the number of positive and negative cases by repeating a small number of samples, but simply repeating a small number of samples will make the information amplified and the model learns too "special" information. In the SVM model parameters in the sklearn library, we set the class_weight parameter to "balanced" and use the weighting method to make the majority class and the minority class in the weighted balance, which can play an equally large role in the loss function of the SVM training process and alleviate the impact of the data. The impact of the class imbalance on the model is mitigated.

In the formal experiments, more than 1,000 agribusiness economic data were predicted according to the agribusiness economy using a genetic algorithm and SVM-based agribusiness economic prediction model and a traditional linear SVM model, respectively. The parameters of the genetic algorithm were referred to the previous tests, and the population size was taken as 10, the variance range was 0-0.1 (uniform distribution), and the number of iterations was 20. The parameters of the SVM model for both are $C = 1.0$, tol = 0.0001, and class_weight = balanced. The correctness and recall of the predictions are shown in Table 2.

As can be seen from Table 2, in terms of agribusiness economic indicators, two agribusiness economic indicators, total phosphorus and BOD5, are better predictors and easier to predict, while agribusiness economic and COD are relatively more difficult to predict. In terms of model comparison, the improved SVM outperforms the traditional SVM in predicting all agribusiness economic indicators, especially in the prediction of agribusiness economic indicators, which has a very great improvement. It shows that the practice of optimizing feature weight vectors by the genetic algorithm to achieve some degree of feature selection can indeed improve the prediction performance of the classifier.

The effects of various economic predictions of different agricultural companies are shown in Figure 4. It can be known that in green agricultural listed companies, the financial risks caused by external factors are often not controllable by human, such as changes in the market environment and the influence of political situation. Each country has policy restrictions for overseas investment, including foreign investment, investment ratio, and approval system, which leads to the macroeconomic environment that has an impact on the company’s financial risk, which in addition to this includes social culture and resource environment.

| Prediction target | Total phosphorus | Total nitrogen | BOD5 | COD |
|-------------------|------------------|----------------|------|-----|
| Proportion of positive cases/% | 22.4 | 64.1 | 19.2 | 51.3 |

| Prediction target | Traditional SVM | Improved SVM | Traditional SVM | Improved SVM |
|-------------------|------------------|--------------|-----------------|--------------|
| Total phosphorus   | 79.2             | 86.5         | 69.5            | 73.6         |
| Total nitrogen     | 68.9             | 79.2         | 66.5            | 76.8         |
| BOD5               | 80.1             | 86.2         | 71.8            | 75.1         |
| COD                | 73.2             | 74.8         | 72.1            | 75.6         |

![Figure 4: Effect of various economic projections for different agricultural companies.](image-url)
These external factors can have a huge impact on the company’s finances, especially in the complex market environment, where supply and demand are changing at all times, and the increase in rivals and products can lead to the company facing sales difficulties and other situations, and despite internal coping strategies, wrong decisions can still drive the company into financial difficulties.

4. Conclusion

It is only important to establish an appropriate financial early warning model for prediction, analysis, and control of agricultural economic risks. In this paper, we use a genetic algorithm and support vector machine-based economic forecasting model for agricultural enterprises to improve the accuracy of forecasting by adapting pollutant weights in a practical application environment. It provides new ideas for the subsequent research of financial early warning methods and also expands the boundaries of theoretical research and enriches the experience and evidence of practical research for the research system of financial early warning problems.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declared that they have no conflicts of interest regarding this work.

References

[1] Y. Zhang, "Food safety risk intelligence early warning based on support vector machine," Journal of Intelligent & Fuzzy Systems, vol. 38, no. 6, pp. 6957–6969, 2020.

[2] J. S. Chou, M. Y. Cheng, Y. W. Wu, and A. D. Pham, "Optimizing parameters of support vector machine using fast messy genetic algorithm for dispute classification," Expert Systems with Applications, vol. 41, no. 8, pp. 3955–3964, 2014.

[3] L. Guo and Y. Zhao, "Research and verification analysis on early warning model through support vector machine algorithm," in 2021 IEEE International Conference on Electronic Technology, Communication and Information (ICETCI), pp. 211–213, Changchun, China, 2021, August.

[4] J. Su, X. Wang, S. Zhao, B. Chen, C. Li, and Z. Yang, "A structurally simplified hybrid model of genetic algorithm and support vector machine for prediction of chlorophyll a in reservoirs," Water, vol. 7, no. 12, pp. 1610–1627, 2015.

[5] X. J. Wang, G. T. Zeng, K. X. Zhang, H. B. Chu, and Z. S. Chen, "Urban real estate market early warning based on support vector machine: a case study of Beijing," International Journal of Computational Intelligence Systems, vol. 13, no. 1, pp. 153–166, 2020.

[6] S. Liu, H. Tai, Q. Ding, D. Li, L. Xu, and Y. Wei, "A hybrid approach of support vector regression with genetic algorithm optimization for aquaculture water quality prediction," Mathematical and Computer Modelling, vol. 58, no. 3-4, pp. 458–465, 2013.

[7] T. Li, J. Yang, and Z. Chen, "The early warning and prediction method of flea beetle based on maximum likelihood algorithm ensembles," in 2010 Sixth International Conference on Natural Computation, vol. 4, pp. 1901–1905, Yantai, 2010, August.

[8] K. Nagasubramanian, S. Jones, S. Sarkar, A. K. Singh, A. Singh, and B. Ganapathysubramanian, "Hyperspectral band selection using genetic algorithm and support vector machines for early identification of charcoal rot disease in soybean stems," Plant Methods, vol. 14, no. 1, pp. 86–113, 2018.

[9] X. Yu, "Disaster prediction model based on support vector machine for regression and improved differential evolution," Natural Hazards, vol. 85, no. 2, pp. 959–976, 2017.

[10] R. Kolachian and B. Saghafian, "Hydrological drought class early warning using support vector machines and rough sets," Environmental Earth Sciences, vol. 80, no. 11, pp. 1–15, 2021.

[11] R. C. Deo, O. Kisi, and V. P. Singh, "Drought forecasting in eastern Australia using multivariate adaptive regression spline, least square support vector machine and M5Tree model," Atmospheric Research, vol. 184, pp. 149–175, 2017.

[12] Z. H. A. N. G. Zhengwan, Z. H. A. N. G. Chunjiom, L. I. Hongbing, and Z. E. Tao, "Multipath transmission selection algorithm based on immune connectivity model," Journal of Computer Applications, vol. 40, no. 12, pp. 3571–3577, 2020.

[13] P. An, Z. Wang, and C. Zhang, "Ensemble unsupervised autoencoders and Gaussian mixture model for cyberattack detection," Information Processing & Management, vol. 59, no. 2, article 102844, 2022.

[14] M. Fallah Shams, H. Jahangirnia, R. Gholami Jamkarani, H. Kordlouie, and R. Derakhshani, "Designing credit risk early-warning system for individual and corporate customers of the bank using multiple logit comparison model and survival function," International Journal of Finance & Managerial Accounting, vol. 7, no. 25, pp. 163–177, 2022.

[15] I. R. Morales, D. R. Cebrián, E. F. Blanco, and A. P. Sierra, "Early warning in egg production curves from commercial hens: a SVM approach," Computers and Electronics in Agriculture, vol. 121, pp. 169–179, 2016.

[16] Y. Wu and X. Long, "Research of groundwater environment early warning based on intelligent algorithm," in 2012 International Conference on Computer Science and Electronics Engineering, vol. 1, pp. 536–540, Hangzhou, China, 2012, March.

[17] L. Sun, Y. Qiandi, D. Peng, S. Subramani, and X. Wang, "Fogmed: a fog-based framework for disease prognosis based medical sensor data streams," Computers, Materials & Continua, vol. 66, no. 1, pp. 603–619, 2021.

[18] Q. Zhiguo, H. Sun, and M. Zheng, "An efficient quantum image steganography protocol based on improved EMD algorithm," Quantum Information Processing, vol. 20, no. 2, pp. 1–29, 2021.

[19] Z. H. A. N. G. Zheng-wan, W. U. Di, and Z. H. A. N. G. Chunjiom, "Study of cellular traffic prediction based on multi-channel sparse LSTM," Computer Science, vol. 48, no. 6, pp. 296–300, 2021.

[20] Q. Zhiguo, S. Chen, and X. Wang, "A secure controlled quantum image steganography algorithm," Quantum Information Processing, vol. 19, no. 10, pp. 1–25, 2020.

[21] J. Su, X. Wang, Y. Liang, and B. Chen, "GA-based support vector machine model for the prediction of monthly reservoir environmental and Computer Modelling, vol. 58, no. 3-4, pp. 458–465, 2013."
storage,” *Journal of Hydrologic Engineering*, vol. 19, no. 7, pp. 1430–1437, 2014.

[22] J. Li, Z. Zhou, J. Wu et al., “Decentralized on-demand energy supply for blockchain in internet of things: a microgrids approach,” *IEEE Transactions on Computational Social Systems*, vol. 6, no. 6, pp. 1395–1406, 2019.

[23] W. Duan, J. Gu, M. Wen, G. Zhang, Y. Ji, and S. Mumtaz, “Emerging technologies for 5G-IoV networks: applications, trends and opportunities,” *IEEE Network*, vol. 34, no. 5, pp. 283–289, 2020.

[24] D. Jiang, F. Wang, Z. Lv et al., “QoE-aware efficient content distribution scheme for satellite-terrestrial networks,” *IEEE Transactions on Mobile Computing*, 2021.

[25] J. Xu, Z. Meng, and L. Xu, “Integrated system of health management-oriented reliability prediction for a spacecraft software system with an adaptive genetic algorithm support vector machine,” *Eksplotacja i Niezawodność*, vol. 16, no. 4, pp. 571–578, 2014.

[26] X. Li, X. Lyu, Y. Tong, S. Li, and D. Liu, “An object-based river extraction method via optimized transductive support vector machine for multi-spectral remote-sensing images,” *IEEE Access*, vol. 7, pp. 46165–46175, 2019.

[27] L. Yi, X. Ke, and S. Junde, “Research on forecasting and early-warning methods,” in *2013 IEEE 9th International Conference on Mobile Ad-hoc and Sensor Networks*, pp. 570–576, Dalian, China, 2013, December.